Cryptobenthic Coral Reef Fishes: Resolving Critical Gaps in Species Delimitation, Sampling Techniques, and Phylogenetic Relationships

Marta Cristina Gómez-Buckley

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Luke Tornabene, Chair

Lorenz Hauser

Jacqueline Padilla-Gamiño

Program Authorized to Offer Degree:

Aquatic and Fishery Sciences

University of Washington

Abstract

Cryptobenthic Coral Reef Fishes: Resolving Critical Gaps in Species Delimitation, Sampling Techniques, and Phylogenetic Relationships

Marta Cristina Gómez-Buckley

Chair of the Supervisory Committee:

Luke Tornabene

Aquatic and Fishery Sciences

Cryptobenthic reef fishes (CRF) are the smallest vertebrates on coral reefs but represent about 40% of the fish species and about 50% of fish abundance in coral reef ecosystems. Their diversity can be explained by their extremely limited dispersal abilities and short generation times (promoting allopatric speciation) coupled with their ability to partition microhabitats at a very fine scale. One of their main contributions to coral reef ecosystems is as food for larger piscivorous fishes and invertebrates. Despite their importance, CRF are often overlooked, as their cryptic nature and very small size make them difficult to assess using visual methods. The

prevailing method for quantifying communities of CRF is to use enclosures and anesthetics to collect fishes from within the reef. Environmental DNA (eDNA) sampling of water around reefs may be an alternative or supplemental way to quantify CRF communities. In Chapter 1, I compared the ability to characterize CRF communities using eDNA sequences from water samples collected from within the interstices of the reef, with corresponding anesthetic stations from the same microhabitat. A total of 676 CRF from 33 species were collected from anesthetic stations, whereas only four species of CRF were detected from the eDNA samples taken at the same locations. Main factors contributing to these results may have been: (1) low standing biomass of CRF, thus low abundance of detectable DNA in the eDNA water samples; (2) large number of non-targeted metazoan reads that "drowned" the detections of specific CRF sequences; and (3) lack of sequences available in public databases that represent the high level of endemism present in CRF and a need for more complete taxonomic inventory.

For some groups of CRF, their small size, cryptic nature, and conserved morphology have resulted in many undetected cryptic species, which may require a genome-wide species delimitation approach to discern how many species are present. One of the most species-rich groups of CRF, the genus *Eviota*, has 132 species described to date, is widely distributed from the Red Sea to Hawaii and French Polynesia, and is known to comprise numerous cryptic species. In Chapter 2, I focused on the *Eviota sigillata* complex which is represented by two nominal species described by morphological characters, yet preliminary genetic data suggest the presence of multiple cryptic lineages. I use molecular data from mitochondrial DNA and genome-wide SNP data generated via double digest restriction site associated sequencing (ddRADseq), in combination with morphological data to infer the number of species in the *E.*

sigillata complex. Specifically, I constructed phylogenetic trees and conducted several types of single-locus and multilocus species delimitation analyses and compared these to groupings based on morphology, as well as their geographic distribution. Overall, I found evidence for the presence of 9-13 lineages within the *E. sigillata* species complex, with genetic lineages corresponding well with the biogeography of the group. I further confirmed that the original morphological diagnostic characters used for the separation of the two nominal species were not useful for distinguishing each of the nine clades in the complex but may be helpful in diagnosing groups of species. Overall, Chapter 2 shed light onto the patterns of speciation within CRF and provided a glimpse of the tremendous hidden diversity that remains in coral reef fishes.

The goby genera *Eviota* and *Sueviota* (family Gobiidae) are commonly known as dwarfgobies, and collectively, the two genera are among the most abundant and diverse groups of fishes on coral reefs. Despite the diversity, abundance, and ecological importance of this group, and the large and growing number of species described to date (132 *Eviota*, 9 *Sueviota)*, there is a lack of understanding of the phylogenetic relationships among the major clades of *Eviota*, poor knowledge of the relationships between dwarfgobies and other Gobiidae species, and no information on the placement of *Sueviota.* In addition, as is the case with most small reef fishes, a clear understanding of taxonomically informative phenotypic characters is also lacking. In Chapter 3, to resolve the evolutionary history of dwarfgobies, I inferred a time-calibrated phylogeny of the group using genome-wide data from 440 ddRADseq loci captured across 98 *Eviota* and *Sueviota* specimens, plus 66 specimens of other related gobies. I also assessed the distribution of 14 external and osteological morphological characters across the tree to assess which may be useful for diagnosing clades. Results from my Chapter 3 robustly established the

non-monophyly of the genus *Eviota*, which was resolved into two separate clades, both of which were resolved within a lineage of other coral associated genera (*Gobiodon*, *Paragobiodon*, *Pleurosicya*, *Minisicya* and *Bryaninops*). One of the two clades is herein elevated to its own genus, *Eviotops*, a name which was previously considered synonymous with *Eviota*. Additionally, I established that the genus *Sueviota* is deeply nested within one of the *Eviota* clades and is herein synonymized with *Eviota.* Furthermore, I found strong phylogenetic signals for 12 out of the 14 phenotypic traits examined, providing strong complementary support for the two recovered clades and establishing the validity of phenotypic traits that strongly correspond with genetic groupings that should aid in future taxonomic studies for this group.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

DEDICATION

To my loving husband and our dearest daughter Adriana. Without their continuous support and encouragement, I would not have accomplished my PhD dream. To my husband specially for his excitement about my research and for being my SCUBA diving buddy during my field work in The Kingdom of Tonga.

To my advisor Luke Tornabene for his ardent support and patience in teaching his "*first PhD pancake*" all the intricacies of the taxonomic field, and the complexity and fascinating genomic tools to detect species among one of the most challenging group of fishes.

Finally, to my family in Spain, my sister Rocío, and my brothers Álvaro and Alfonso. They were routing for me all the way during the high and the low moments of these past seven years. I know that *nuestros Madre y Padre* would have been proudly celebrating my degree.

ACKNOWLEDGMENTS

I want to start by recognizing my beloved family, my husband Ray Buckley and our daughter Adriana Gómez-Buckley. They are my most ardent admirers. They have been the pillars for my completing my PhD. Adriana, being a fellow UW student and getting her undergraduate UW degree in astrophysics and getting admitted to the University of Arizona's PhD program in Astrobiology and Microbiology has been inspirational for me. 'Perdón hija', but I got my PhD before you! Ray has been there cheering me on and supporting me in so many ways throughout these past seven years...from bringing me food and drinks while I worked at the computer, to helping me in the field collecting samples and cleaning gear. His love and devotion to me and my work kept me going in the good moments, and in the low moments when the pandemic stopped our ability to get to the field to complete my last chapter. He has always believed in me and has been my strong shining and reliable lighthouse.

My deepest gratitude goes to the people of Vava'u, Kingdom of Tonga, for allowing me to do my field work in such incredibly beautiful setting. I especially would like to thank Richard Wylie who invited us to join him in his little parcel of paradise in Euakafa Island in 2016. He introduced us to Karen Stone and her work as Director of the Vava'u Environmental Protection Association who, along with Dr. Tuikolongahau Halafihi (aka Hau), were fundamental in getting us all the permits required for cryptobenthic reef fish collections and transport outside of their country. Karen has been an inspirational figure for me throughout my PhD journey. Her power to accomplish so much with very limited resources makes her a superhero in my eyes, especially through the pandemic years. I am also grateful to Richard Ball (Rikki), who provided his boat, expertise about sites for sampling, and help with diving gear. Many heartwarming thanks to Adriana and Kjell for their Mystic Sands accommodations, their interest to support science, for allowing us to setup our 'lab' in our Mystic Sands' bungalow, and for having the great food and love from Mossi, her daughter Peta and their family at the Utungake Village.

Thank you to my committee members, Jackie Padilla-Gamiño, Lorenz Hauser, and Ryan Kelly for your encouragement, support, and always pushing me to stretch my thinking on how to relate my research to their areas of expertise and beyond. I am especially thankful for all the guidance, support, and patience that my main advisor, Luke Tornabene, gave me through my PhD. I am proud to be his first 'PhD pancake'! and to have been so lucky to work with him in the lab, in the field, and at the computer for countless hours. He taught me all I know about fish taxonomy and fish phylogenetics. It is hard to believe how much I have learned thanks to him about a field I had so very little background in before I started my degree.

Thanks to my collaborators Ramón Gallego and Vanesa Arranz who helped me 'untangle' the eDNA data for my Chapter 1. Thanks for transferring all those bioinformatic skills that made it possible to see through that chapter and its ensuing publication, and that permeated into my next two chapters. Gracias to Carmen Pedraza-Marrón for teaching me all about processing the data for species delimitation analyses, 'me salvaste la vida, mi amiga'!! I could not have done any of my chapters without the monumental work of so many taxonomists that have worked through the past many decades on the CRF, namely David Greenfield, Richard Winterbottom, Mark Erdmann, and many other that I only had the pleasure to know via their published work. Thank you to Adam Leaché and to Hyden Davis for their invaluable help with the species delimitation

analyses in my Chapter 2. Thanks to the Esteven Roberts lab for allowing me to join their UW high performance computer node and for helping troubleshoot my data processing. Sam White was fundamental in me not ‘throwing the towel’ when encountering roadblocks with ‘the supercomputer’. More supercomputer help came from the MER lab crew. Anita Wray and Bryan Briones, both lending me a hand many times navigating the supercomputer.

Thank you to my lab mates for all the times we interacted in lab meetings, conferences and social gatherings. Being with you and feeling your excitement (even if via zoom) for learning and sharing all nerdy fish things helped me move forward through the harder pandemic times. Thank you to Katherine Maslenikov, our lab manager who was also inspirational in so many ways. Her skillfulness guiding us through the permit process before we traveled to Tonga and when coming back with the samples through US Customs, made what could have been a stressful process so much easier. All the undergrad students that helped me with DNA extractions and with the archival of over a thousand CRF specimens that we collected from Tonga; you are much appreciated.

I want to thank all the SAFS community in general, but especially the graduate advisors, Samantha Scherer and earlier Amy Fox, who were always quick to respond to all my queries. Thanks to all the IT staff who rescued me on multiple occasions. Thanks to my ‘FAB’ writing group colleagues Kristin Privitera-Johnson, Lee Qi, Chris Mantegna, and Madi Shipley. I accomplished so much during our writing meetings, but they also served with the very important part of the equation, connectedness! Thank you to my emeritus(a) SAFS faculty friends Don Gunderson, and Frieda Taub for our fun Sunday meetings, and to Frieda for letting us crash at her house when I traveled from Oly during my last PhD quarters. Thank you to all undergraduate students that I taught, I learned a lot from you!

Thank you to Lauren Updyke and the *Whole U* program, who kept me and so many in our UW community sane and healthy through the pandemic and beyond. I’m stronger now than I’ve ever been thanks to the Tuesday and Thursday Zoom weight training sessions. I have ‘guns’ thanks to you Lauren!

My PhD was in part funded by multiple TA and RA positions at SAFS, Oceanography, and Biology departments, and multiple quarter scholarships from SAFS and Friday Harbor Labs donors. Additionally, I received the following scholarships: The Hall of Conservation Genetics Research Fund (UW College of the Environment); the Lerner-Gray Memorial Fund of the American Museum of Natural History; the American Academy of Underwater Sciences Foundation; The Seattle Rho Chapter of Washington Women Educators (Thanks you Loveday Conquest and Kate Griebshaber for five years of funding!); The William W. and Dorothy T. Gilbert Ichthyology Research Fund; and NSF Award #1827013.

Chapter 1

Comparing anesthetic stations and environmental DNA sampling to determine community composition of cryptobenthic coral reef fishes of Vava'u, Kingdom of Tonga

Published in: *Coral Reefs*, 10 May 2023. Reproduced with permission from Springer Nature.

Marta C. Gómez-Buckley[1], Ramón Gallego[2], Vanessa Arranz[3], Tuikolongahau Halafihi[4], Karen Stone[5], Mark Erdmann[6], and Luke M. Tornabene[1]

[1] School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA

[2] Universidad Autónoma de Madrid, Madrid, Spain

[3] Massey University, Auckland, New Zealand

[4] Ministry of Fisheries, Nuku'alofa, Tongatapu, Tonga

[5] Vava'u Environmental Protection Association NGO, Vava'u, Tonga

[6] Conservation International Aotearoa, Auckland, New Zealand

ABSTRACT

Cryptobenthic reef fishes (CRF) are an important component of coral reef communities that are often overlooked, as their cryptic nature and very small size make them difficult to assess using visual methods. The prevailing method for quantifying communities of CRF is to use enclosures and anesthetics to collect fishes from within the reef. Environmental DNA (eDNA) sampling of water around reefs may be an alternative or supplemental way to quantify CRF communities. Here we compare the ability to characterize CRF communities using eDNA sequences from water samples collected from within the interstices of the reef, with corresponding anesthetic stations from the same microhabitat. A total of 676 CRF from 33 species were collected from anesthetic stations, whereas only four species of CRF were detected from the eDNA samples taken at the same locations. Main factors contributing to these results may have been: (1) low standing biomass of CRF, thus low abundance of detectable DNA in the eDNA water samples; (2) large number of non-targeted metazoan reads that "drowned" the detections of specific CRF sequences; and (3) lack of sequences available in public databases that represent the high level of endemism present in CRF and a need for more complete taxonomic inventory.

INTRODUCTION

Coral reef ecosystems are biodiversity hotspots that house an estimated one quarter to one third of all marine species (Plaisance et al. 2011). Despite coral reef fishes being a large proportion of this tremendous biodiversity (~ 6,000 species; Brandl et al. 2018; Eschemeyer et al. 2010), most reef biodiversity surveys overlook nearly half of this community by ignoring small and cryptic species that are important components of reef trophodynamics (Bierwagen et al. 2018; Bohmann et al. 2014; Brandl et al. 2018, 2019). The need to obtain a more comprehensive census of life in coral reef ecosystems has become urgent due to the imperiled status of the shallow reef building corals (McClenachan 2015) that house most of these 'less visible' reef fauna (Plaisance et al. 2009).

The smallest vertebrates on coral reefs are cryptobenthic reef fishes (CRF). Species of CRF belong to at least 17 different families of fishes, are typically less than 5 cm long, and live very close to the coral substrate, often hiding within the crevices of the reef (Depczynski and Bellwood 2004; Brandl et al. 2018). Their minute sizes allow them to exploit nearly all available microhabitats found in coral reef ecosystems. CRF can use these microhabitats in very specific ways, which may help explain the high species richness observed within this group (Brandl et al. 2019; Greenfield 2017, 2021; Winterbottom et al. 2014). CRF represent about 40% of the fish species present in coral reef ecosystems, and about 50% of fish abundance (Depzynski and Bellwood 2003, 2005; Brandl et al. 2018). Because of their high turnover rates (e.g., the goby *Eviota sigillata* has the shortest known vertebrate lifespan at only 59 days; Depczynski and Bellwood 2005), they transfer substantial energy as a constantly replenishing source of prey for larger reef inhabitants (Brandl et al. 2019). Consequently, there is a clear need to include CRF when quantifying community composition on reefs.

The most effective method for the study of CRF is to collect specimens from the reef using chemicals and euthanizing them so that they can be properly identified. These collections are key for contributing to studies on systematics as well as those looking at community structure, trophic ecology, and other ecological questions (Ahmadia et al. 2018; Atta et al. 2019; Brandl et al. 2020a,b; Troyer et al. 2018). Collecting cryptobenthic fishes requires the use of ichthyocides or anesthetics, often in combination with mesh and or plastic enclosures (Ackerman and Bellwood 2000; Robertson and Smith-Vaniz 2008, 2010; Depczynski and Bellwood, 2004). Different chemicals may have varying levels of effectiveness on different species. Some of the more secretive species that are deep within the reef crevices or those less sensitive to the chemical solutions may still go undetected (Ackerman 2002). Moreover, identification after capture often requires high quality color photographs of fresh specimens, examination under a microscope, and a moderate to high degree of taxonomic training coupled with available taxonomic literature; even then, molecular analysis may be required for some groups.

Sampling environmental DNA (eDNA) is an emerging technique for surveying biodiversity (Thomsen and Willerslev 2015). Environmental DNA can be defined broadly as the pool of DNA isolated from the environment (Pawlowski et al. 2020). Despite the perceived versatility of eDNA sampling for biodiversity studies, there are issues that affect the success of this type of sampling. The persistence of eDNA in the marine environment varies depending on the surrounding conditions (Collins et al. 2018). Higher water temperatures are correlated with higher degradation rates of fish-originated eDNA (Tsuji et al. 2017). Shorter fragments of shed DNA may be recovered more readily in the hotter tropical environments and thus it is often more fruitful to use a combination of genetic markers and primers that could capture these shorter, but more abundant eDNA fragments (Collins et al. 2018). Several PCR primers have been designed to specifically recover fish DNA, including those targeting mitochondrial 12S or 16S (Miya et al. 2015; Valentini et al. 2016; DiBattista et al. 2017; see Zhang et al. 2020, for a review of fish primers). Despite the versatility and popularity of the 12S and 16S for fish eDNA studies, these markers usually do not contain sufficient variation to identify sequences to species level or differentiate among rapidly diversifying and closely related taxa, which is the case for many groups within the CRF (Tornabene et al. 2015; Mabuchi et al. 2014; Lin and Hastings 2013). Moreover, there are fewer sequences of 12S and 16S in public databases when compared to loci like mitochondrial cytochrome c oxidase subunit I (COI), which is the most abundant locus in GenBank (Clark et al. 2016) and BOLD (Ratnasingham and Hebert 2007) due to widespread use in DNA barcoding studies (Hebert et al. 2003). The number of COI sequences for bony fishes (Class Actinopterygii) in GenBank is four to five times larger than what is available for 12S or 16S (as of August 2021). Because of its high mutation rate, it is recommended for taxonomic assignments at the species level, with the added advantage of allowing for discrimination among haplotypes within species populations (Andújar et al. 2018). These factors make it particularly useful for CRF eDNA experiments. A standard barcoding fragment from the larger whole COI gene (658 bp out of a total of ~ 1,500 bp) targeted by “Folmer primers” (Folmer et al. 1994) has been specifically helpful in taxonomic studies of many CRF groups including gobies (Winterbottom et al. 2014), blennies (Lin and Hastings 2013), clingfishes (Conway et al. 2020), basslets (Baldwin et al. 2016), and cardinalfishes (Mabuchi et al. 2014).

There have been several recent studies using eDNA to survey reef fishes (Table 1). DiBattista et al. (2017) collected eDNA water samples from the surface above reefs in the Red Sea and used fish-specific mitochondrial 16S primers. They found that only about half of the species present in their combined NCBI, custom 16S genetic database, or observed using underwater surveys, were detected in their eDNA samples. Only one CRF was detected at the species level, which was recovered from NCBI. Nguyen et al. (2020) conducted visual fish surveys and compared the results with the fish detected in eDNA samples collected 10-20 cm above the substrate. They used primers targeting COI from all metazoans (Leray et al. 2013). Despite detecting >200 metazoan taxa previously recorded for the local databases, only 0.7% of the reads were assigned to bony fishes, with only two CRF species detected in their samples. Most recent studies targeting fishes (Table 1) detected a range of 0-59 CRF at the species level, but only Bessey et al. (2020, 2021) detected two species of the genus *Eviota* (denoted as *E*. sp.1 and sp. 2, as species-level matches were not found), which is problematic as the genus *Eviota* is the most specious genus of CRF (123 species; Greenfield 2021) and is frequently the most abundant group of CRF (Ahmadia et al. 2018). There is a clear need to explore new approaches to tackle CRF detection deficiencies in eDNA studies.

Given the low standing biomass of CRF (Brandl et al. 2019) coupled with their behavioral propensity for remaining closely associated with the benthos, it may be easier to detect their DNA closer to the substrate. No eDNA studies published to date have focused on detection of CRF by collecting *in situ* underwater eDNA samples from within the interstices of the reef substrate where shed DNA fragments may be more concentrated (Turner et al. 2015; Gelis et al.

2021), nor have they collected the fishes from the same microhabitats from where the eDNA samples were taken. Here we focused on collecting eDNA samples from within the interstices in coral heads and rubble substrates, followed by collection of CRF from within the same sampled microhabitats. This approach should theoretically make it easier to detect CRF in eDNA samples, since defining traits for this group include their mostly continuous physical contact with the substrate and their preference for hiding in interstices of the reef (Depczynski and Bellwood 2004). The objective of this study was to determine if *in situ* eDNA water samples would be a feasible method to survey CRF as an alternative or supplement to anesthetic stations. We specifically addressed how eDNA sampling compares to anesthetic stations when used as a tool to quantify CRF communities.

METHODS

Sampling Collection

All samples were collected during December 2018-January 2019 at Vava'u, the northernmost group of islands in the Tongan Archipelago (Figure 1). See supplementary material for site description. At each site we sampled three different microhabitat types: live coral bommies (LC), dead coral rock (DC), and coral rubble (CR) (see supplementary material for descriptions). The number of combined samples across sites for the live coral, coral rubble and dead coral microhabitats was eight, seven, and six, respectively (Table 2).

Environmental DNA and CRF samples were collected underwater using standard SCUBA gear. When a suitable microhabitat was selected, eDNA samples were either collected at the substrate surface (when interstices were absent), or within the interstices of the coral rubble or the coral heads, using 200 ml modified syringes (Figure S1), allowing for the rapid collection of eDNA samples, thus reducing the potential for including water from the surrounding environment (See supplementary material for detailed methods). Two to three replicates were taken from each station, with combined volumes ranging from 200-1,200 ml per station. We restricted the amount of water that we sampled from each station because our objective was to capture the volume of water within the confines of the crevices of the microhabitats, or as close as possible to the substrate were the CRF are found. Water samples were filtered, and filter discs were preserved in Longmire's solution (Longmire et al. 1997). Immediately after collecting the eDNA samples, we collected CRF from that same microhabitat using enclosed anesthetic stations following protocols modified from Ahmadia et al. (2018), photographed specimens in the field, preserved them in 95% ethanol and identified them back in the lab (see Supplementary Material for details).

eDNA lab processing

DNA from each whole filter disc corresponding to individual sample replicates was extracted and kept as a separate sample for the subsequent amplification and sequencing processing. See supplementary material for details on DNA extraction. Extracted raw DNA from all the eDNA samples was sent to Jonah Ventures Lab (Boulder Colorado, USA) for sample amplification. A 313 bp section of the mitochondrial COI gene was selected and amplified via PCR using the forward UniCO1F and reverse UniCO1R primers (Leray et al. 2013). This smaller COI section was selected based on the following four premises: (i) it is located within the 658 bp 'Folmer' region, which is the mostly widely recognized universal marker for species identification (Herbert et al. 2003), making it the most well-represented marker in public molecular databases like GenBank (NCBI) and BOLD; (ii) when dealing with eDNA samples in warm tropical waters, a shorter segment of the COI would be more abundant than the complete 658 bp fragment used in DNA barcoding studies; (iii) COI (vs. 12S or 16S) has been used successfully to identify CRF species at the species level, and it has proven particularly useful for resolving taxonomic delineation of the genus *Eviota* and *Trimma* (dwarfgobies and pygmygobies respectively), two of the most abundant and diverse groups of CRF; (iv) the availability of an extensive inhouse *Eviota* COI sequence database spanning much of the Indo-Pacific region (Tornabene, et al. 2013a,b, 2015, 2016, 2021; Tornabene unpublished data). A 2-step PCR process was used to amplify target amplicons and identify unique samples, using standard lab procedures. Full details and concentrations can be found in Supplemental materials. The final library was run using a V2-500 cartridge on an Illumina MiSeq.

Bioinformatics and taxonomic assignments

The bioinformatic pipeline used for the cleaning and filtering of sequenced reads can be found in the GitHub repository - https://github.com/ramongallego/Nextera_Dada2. We chose the well-known sequence similarity method, BLAST (Altschul et al. 1997), which has been demonstrated to perform well as a taxonomic assignment tool (Hleap et al. 2021 – See supplementary material for details on the bioinformatic pipeline and the parameters used for taxonomic assignments).

To increase classification sensitivity for taxonomic assignments, a custom reference sequences database was prepared using MARES pipeline (Arranz et al. 2020). MARES includes all the COI sequences from GenBank and BOLD for marine eukaryote taxa. Three important features of MARES are: (1) it restricts the number of sequences to just taxa from eukaryote families that are represented in the marine environment, making the querying process more efficient; (2) allows to add custom unpublished sequences; (3) a taxonomic normalization step synonymize information from BOLD, GenBank and custom sequences. Marine eukaryote sequences were retrieved from both GenBank and BOLD on July 17, 2021. For our study, and to supplement the reference database produced by MARES, we merged our "in-house" *Eviota* COI sequences database, which contains 460 sequences, with most (~75%) not yet uploaded to either GenBank or BOLD, each with a photographed voucher specimen. This *Eviota* database contains COI fragments (each ~658 bp) from an extensive and ongoing collecting effort from the entire Indo-Pacific region, with most specimens sampled from the Coral Triangle region, and lesser representation (2-20 specimens for each location) from Fiji, the Maldives, Moorea, Pohnpei, American Samoa, Solomon Islands, northeast Australia, and the Red Sea. This database also included 68 new COI sequences from the Vava'u Island group, Tonga collected in 2018. Most of the sequenced specimens from Vava'u were part of the *Eviota* genus, yielding 16 species, 10 of them undescribed.

A final identification of each ASV was produced with a custom R script which used the LCA function from the R package taxonomizr to calculate the last common ancestor of all matching sequences in the database with a similarity of > 95% over at least 250 bp (See supplement for more detail). All statistical analyses and data visualization were done with R version 4.0.3.

RESULTS

eDNA samples

The primers used in this study successfully aligned within the fish COI sequences available in the used reference database (Figure S2). A total volume of 13,900 ml was filtered from the 21 samples spread across three microhabitats (Table 2). The total number of raw sequences captured from all 21 eDNA samples after sequencing was 1,686,370. The read depth values for all the eukaryotes detected ranged between 9,186 and 91,613 per sample (combined replicates) with an average of 46,866 per sample (Table S1). After bioinformatic filtering using our combined MARES and *Eviota* database for taxonomic matching, a total of 313,184 marine eukaryote reads remained (Table 3). Among these reads, only 1,812 were assigned to CRF, belonging to four species from a single family, Gobiidae (Table 4). All four species belong to the genus *Eviota*, and three of the four species are undescribed: *Eviota* Tonga sp. 1 (1,392 reads), *E.* Fiji sp. 1 (404 reads), *E. cf. punctulata* (8 reads), and *E.* Papua New Guinea sp. 1 (8 reads). Other non-CRF coral-reef fishes (19 species, 8 families) were detected for a total of 2,039 reads. A total of 1,105 of these reads belonged to *Nasso brevirostris* and came from a single sample. A total of 1,964 reads from 19 ASVs could only be assigned to class Actinopterygii, and 4 reads from 1 ASV could only be assigned to class Chondrichthyes.

Anesthetic station samples

A total of 689 specimens belonging to 15 families (46 species), were collected from the 21 anesthetic stations where eDNA samples were also taken (Table 4,Figure S3). The number of specimens collected in each anesthetic station ranged from 2 to 99. Seven of the 17 core CRF families, as defined by Brandl et al. (2018), were represented by the 676 CRF specimens collected. The family Gobiidae made up the bulk of all fishes and CRF with 664 specimens. Only 10 specimens belonging to eight species distributed among six other CRF families were found (Table 4): Apogonidae (2), Blenniidae (2), Plesiopidae (1), Oxudercidae (1), Pseudochromidae (1), and Syngnathidae (1). Within the Gobiidae, the genus *Eviota* was about half of all CRF species, with 42% of all CRF belonging to one single species, *E. Tonga* sp. 1, which had 291 specimens collected from the combined three microhabitat types. This species along with *E. Tonga* sp. 2, and *Eviota cf. punctulata*, both also present in the eDNA samples, belong to the *Eviota sigillata* and to the *Eviota punctulata* species complexes respectively, and are currently undergoing taxonomic revision. The same is the case for the *E.* Fiji sp. 1, belonging to the *E. fasciola* species complex, and *E.* Papua New Guinea sp. 1, belonging to the *E. melasma* species complex.

Anesthetic stations and eDNA comparisons

In total, 33 CRF species in 16 genera and 9 families were captured using eDNA and anesthetic stations. Eight other families of non-CRF were also recorded mainly from eDNA samples, with the family Pomacentridae representing 50% of all non-CRF species. The only CRF species that was found in both eDNA and anesthetic stations was *E. Tonga* sp. 1. It made up 76% of all the CRF reads detected (1,812) in eDNA samples, with more than half of them recovered from the coral rubble microhabitat. In the anesthetic stations, *E. Tonga* sp. 1 was also found in larger numbers in the coral rubble (46% of all specimens), as compared to the other two microhabitats combined. The other three species of CRF, *E. Fiji* sp. 1, *E. cf. punctulata, and E. Papua New Guinea* sp. 1, were not collected from anesthetic stations. The combined total species richness of CRF from coral rubble microhabitats was low (7) as compared to live coral (17), or dead coral microhabitats (19). In contrast, for eDNA samples detected more species from coral rubble microhabitats (11) than from live coral (3) and dead coral (7), as compared to the cospatial anesthetic station, with 10 species in coral rubble, 20 in live coral, 20 in dead coral (Figure 2 and Figure S4).

There was no relationship between the number of fish reads detected in eDNA samples and the abundance of fishes collected from anesthetic station (Figure 3). This was true for both CRF and non-CRF species. For eight out of the 21 stations, zero fish reads were detected in the eDNA samples, while the number of fishes collected in corresponding anesthetic stations ranged from 10-62. The highest number of fish collected from a single anesthetic station was 99 (95 CRF), from a dead coral head, while 35 CRF reads were detected in the corresponding eDNA sample. In another anesthetic station (also a dead coral head microhabitat), a total of 9 fishes were collected (8 of them CRF).

DISCUSSION

Analysis of eDNA from water collected from within the matrix of the reef was not an effective technique to survey the CRF community. Only 0.6% of all reads recovered from our eDNA samples were assigned to a total of 4 species of CRFs, compared to the 33 species found in the anesthetic stations. Except for two eDNA studies (Marwayana et al. 2021; Juhel et al. 2020) that documented ~60 CRF species from the Coral Triangle (Indonesia), most eDNA studies studying fishes report zero to 18 CRFs (Table 1). Our study is the first to focus on CRFs by collecting both eDNA samples and specimens from within reef microhabitats, whereas most other eDNA studies focus on reef fishes more broadly. Nevertheless, the many factors complicating the use of eDNA to detect CRF are shared across eDNA studies.

Two obvious factors that contribute to the detectability of target organisms in the environment are the size and abundance of the organism itself, and the rate at which the organisms are shedding DNA into the environment (Barnes and Turner 2016). Despite their very high abundances relative to the total coral reef fish community (50% of all individuals, Brandl et al. 2018), CRFs have body sizes ranging from less than 1 cm to 5 cm total length, and only represent a very small slice (<5%) of the total standing stock fish biomass at any given moment on reefs (Brandl et al. 2019). Thus, shedding of DNA material from CRF must be minimal. This would include indirect contributions to the system from other coral dwellers that ingested CRF and dispersed their DNA in feces.

The number of reads in eDNA samples may not directly correlate with the abundance of organisms present in the sampled environment, due mostly to uneven amplification efficiency for different taxa during PCR process (Collins et al. 2019; Shelton et al. 2016; Kelly et al. 2017, 2019), and this holds true for our samples overall (Figure 3). However, the most abundant species in our anesthetic stations (*E.* Tonga sp. 1, 291 specimens) also had the most reads in the eDNA samples (1,392). On the other hand, no specimens of *E.* Fiji sp.1 were captured in anesthetic stations, while 404 reads from this species were detected in eDNA samples. A possible explanation for this mismatch could be the known primer bias amplification, which is common when using COI (Deagle et al. 2014), but it could also be attributed to the patchiness and random distribution of DNA particles in the environment (Barnes and Turner 2016). The low amounts of CRF DNA detected overall and the high variability in the number of reads present in our samples (0-860 reads from CRFs per sample) speaks to the patchiness of target DNA present in the environment, and the need to optimize lab processing of the samples so a more reliable and homogeneous picture of all the CRF present in the surveyed sites is accomplished.

It is likely that eDNA from CRF is indeed incredibly rare around reefs and therefore requires copious amounts of water for reliable detection, especially in warm tropical waters (Marwayana et al. 2021). Marwayana et al. (2021) recommended that as much as 300 x 1-liter samples per region may be needed to adequately document fish diversity in species-rich tropical ecosystems like the Coral Triangle but recognized that this may not be feasible and recommended 30 x 10-liter samples as an alternative. This amount is still challenging when sampling in remote locations with limited resources for processing and filtering of samples in a timely manner to prevent sample deterioration. Here collected only 1,200 ml per anesthetic station so that we could more closely reflect the volume of the interstices where the CRF would be found and avoid diluting the sample with water from outside the reef matrix, and it’s possible that restricting total volume may have prevented detecting more CRF. Nevertheless, other eDNA studies that have detected CRF using the COI locus have also had poor results when collecting larger volume samples either near the substrate or from the water column. When compared to the 4 CRFs detected from 13,900 ml of water in our study, Nguyen et al. (2020) detected 13 species of CRF in the Caribbean, after filtering 134,000 ml of water sampled just above the benthos (10-30 cm), while in Indonesia, Gelis et al. (2021) detected just one CRF species after filtering 112,000 ml of water collected from the water column and from substrate samples. Water volume seems to be an important factor affecting fish detections from eDNA samples and achieving a leveling of accumulation curves for species detected remains an elusive goal even for studies that filter copious volumes of water (Marwayana et al. 2021; Bessey et al. 2020).

Another important factor that may have limited our detection of CRF is the use of primers that target all eukaryotes (Leray 2013). This fact, combined with the collection of our samples from the substrate, possibly contributed to having excess amounts of marine metazoans other than CRF, which may have overwhelmed the low levels of eDNA from CRF. Nguyen et al. (2020) targeted fishes and invertebrates using the COI primers from Leray et al. (2013). They recognized that the co-amplification of non-targeted metazoans, as well as non-metazoan taxa, which is common when using this loci/primer combination, may have lowered detection of rare taxa of fishes and invertebrates in their samples. We chose this locus and primer combination despite this, rather than using fish-specific primers for loci like 12S or 16S, since the latter genes have comparatively fewer sequences for CRF in reference databases and lack the sensitivity required to differentiate CRF at the species level.

One of the main obstacles echoed in many metabarcoding studies is the incomplete coverage of taxa in available public molecular databases, along with cases where sequences are assigned to the wrong species due to contamination or public sequences linked to misidentified voucher specimens. For example, our samples contained 2,176 reads that according to GenBank were assigned to the ray-finned fish *Alburnus scoranza* (99-100% identity), a freshwater species only found in a European lake. When running that specific sequence as a separate search in BOLD, the result was *Homo sapiens* at a 100% match. Using the MARES+*Eviota* database for taxonomic matching after BLASTing the curated sequenced data, revealed that except for *Eviota cf. punctulata* specimen sequences from Tonga that were recently uploaded to BOLD and GenBank from a recent publication (Greenfield et al. 2018), none of our eDNA sequences for

CRF matched sequences from BOLD and GenBank. This confirms the underrepresentation of CRF in general in public genetic databases and may also be an indication of the underrepresentation of fish sequences in general from under sampled tropical regions (Nguyen et al. 2020).

The fact that so few CRF species are detected in eDNA studies may be related to their high levels of micro-allopatry and endemism (Brandl et al. 2018). Many groups of CRF, including some of the most diverse and abundant genera such as *Eviota*, speciate across extremely fine geographic scales (Tornabene et al. 2015; Taylor and Hellberg 2005; Victor 2015), which is likely the result of limited gene flow, local larval retention, and rapid evolutionary rates (Brandl et al. 2018, 2019). This results in species complexes containing morphologically similar species that are restricted to their own island or island group and differ genetically by as little as 0.5% divergence in COI in recently diverged species (Greenfield and Tornabene 2014), to >5% divergence in older species others (Tornabene et al. 2016). Many of these micro-allopatric species have yet to be described (Winterbottom et al. 2014, 2020, Greenfield 2017; Greenfield et al. 2018). This further emphasizes the need for comprehensive taxonomic studies on CRFs that span geographic regions, as well as DNA barcoding efforts that sample densely within geographic regions (Weigt et al. 2012; Victor 2015; Winterbottom et al. 2020).

Even though sampling of CRF has been done now in many regions (see Brandl et al. 2018 for a review), it was not until the early 2000s that most surveys included photographic records, tissue samples or cataloging of voucher specimens in collections. Using underwater visual surveys and from museum records of specimens collected in multiple expeditions, Randall et al. (2003)

published an exhaustive checklist of fishes from the Tongan Archipelago from, reporting as many as 296 CRF species. Recent underwater visual surveys from a rapid biological survey conducted in the Vava'u Archipelago (Atherton et al. 2014) focused on fish species > 5cm. Despite CRFs not being a targeted group, their checklist included two Gobiidae and two Blenniidae species. Nevertheless, our study, even if restricted to the Vava'u island group, is the first that we know in which the collection of CRF includes photographs, tissues, and voucher specimens from Tonga.

CONCLUSION

Future eDNA or metabarcoding efforts for documenting the community assemblages of CRF should go hand in hand with taxonomic efforts to robustly describe faunas and contribute DNA sequences to public databases. DNA sequences should be linked to specimens cataloged in collections available to the scientific community (de Santana et al. 2021). This is especially relevant for CRFs, which are an integral component of reef fish assemblages, but have mostly been ignored using traditional underwater survey methods. Although our study concludes that AS are the most reliable technique to study CRF community assemblages, and that eDNA sampling may not be suitable at this point to reliably document CRFs, we foresee a continued effort to refine eDNA outcomes, including improvements during collection of samples through the final steps in the data processing. These efforts should go hand in hand with collections of CRF specimens in under sampled regions, including rigorous taxonomic identifications and molecular analysis that will ultimately help document CRF biodiversity.

ACKNOWLEDGMENTS

We thank the people of Tonga and their government agencies, and Vava'u Environmental Association for allowing us to collaborate in their efforts to document reef fish biodiversity and providing us with their knowledge, assistance, and permits to access the sites and for transporting the CRF specimens to the UW fish collection. Special thanks go to: Ray Buckley who provided his expertise as SCUBA diver and assistance with the fish sampler; to Richard Ball (Ricky), who provided his boat, expertise about sites for sampling, and help with gear. Many heartwarming thanks to Adriana and Kjell for their Mystic Sands accommodations, and their interest to support science, for allowing us to setup our base of operations, and for having the great food and love from Mossi and her family at the Utungeke Village. We thank Emily McFarland and Sam Ghods for their help with DNA extractions. Many thanks to Katherine Maslenikov for helping with importing and cataloging specimens. This project was in part funded by the Hall of Conservation Genetics Research Fund (University of Washington), 2018 Lerner-Gray Memorial Fund of the American Museum of Natural History, the American Academy of Underwater Sciences Foundation, the Seattle Rho Chapter of Washington Women Educators, the William W. and Dorothy T. Gilbert Ichthyology Research Fund and NSF Award #1827013.

REFERENCES

Ackerman J. 2002. Comparative efficiency of clove oil and rotenone for sampling tropical reef fish assemblages. *Journal of Fish Biology*, *60*(4), 893–901. https://doi.org/10.1006/jfbi.2002.1891

Ackerman J L and Bellwood D R. 2000). Reef fish assemblages: A re-evaluation using enclosed rotenone stations. *Marine Ecology Progress Series*, *206*(1954), 227–237. https://doi.org/10.3354/meps206227

Ahmadia G N, Tornabene L, Smith D J and Pezold F L. 2018. The relative importance of regional, local, and evolutionary factors structuring cryptobenthic coral-reef assemblages. *Coral Reefs, 37*(1), 279–293. https://doi.org/10.1007/s00338-018-1657-2

Altschul S, Madden T L, Schäffer AA, Zheng J, Zhang Z, Miller W and Lipman D J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. https://doi.org/10.1093/genetics/156.4.1997

Andújar C, Arribas P, Yu D W, Vogler A P and Emerson B C. 2018. Why the COI barcode should be the community DNA metabarcode for the metazoa. *Molecular Ecology*, *27*(20), 3968–3975. https://doi.org/10.1111/mec.14844

Arranz V, Pearman W S, Aguirre J D and LigginsL. 2020. MARES, a replicable pipeline and curated reference database for marine eukaryote metabarcoding. *Scientific Data*, *7*(1), 1–8. https://doi.org/10.1038/s41597-020-0549-9

Atherton J N, McKenna S A and Wheatley A. 2014. Rapid biodiversity assessment of the Vava'u archipelago, Kingdom of Tonga. SPREP. Apia, Samoa. 312 pp.

Atta C J, Coker D J, Sinclair-Taylor T H, DiBattista J D, Kattan A, Monroe A A and Berumen M L. 2019. Conspicuous and cryptic reef fishes from a unique and economically important region in the northern Red Sea. *PLoS ONE*, *14*(10), 1–19. https://doi.org/10.1371/journal.pone.0223365

Baldwin C C, Robertson D R, Nonaka A and Tornabene L. 2016. Two new deep-reef basslets (Teleostei, Grammatidae, Lipogramma), with comments on the eco-evolutionary relationships of the genus. *ZooKeys*, 2016(638), 45–82. https://doi.org/10.3897/zookeys.638.10455

Barnes M A and Turner C R. 2016. The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics* DOI 10.1007/s10592-015-0775-4

Bessey C, Jarman S N, Berry O, Olsen Y S, Bunce M, Simpson T Power, M, McLaughlin J, Edgar G J and Keesing J. 2020. Maximizing fish detection with eDNA metabarcoding. *Environmental DNA*, *January* 1–12. https://doi.org/10.1002/edn3.74

Bessey C, Jarman S N, Simpson T, Miller H, Stewart T, Keesing J, Berry O. 2021. Passive eDNA collection enhances aquatic biodiversity analysis. *Communications Biology*. 4:236 https://doi.org/10.1038/s42003-021-01760-8 | www.nature.com/commsbio

Bierwagen S L, Heupel M R, Chin A and Simpfendorfer C A. 2018. Trophodynamics as a tool for understanding coral reef ecosystems. *Frontiers in Marine Science*, *5*(FEB), 1–13. https://doi.org/10.3389/fmars.2018.00024

Bohmann K, Evans A, Gilbert M T P, Carvalho G R, Creer S, Knapp M, Yu D W and de Bruyn M. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. In *Trends in Ecology and Evolution* (Vol. 29, Issue 6, pp. 358–367). Elsevier Ltd. https://doi.org/10.1016/j.tree.2014.04.003

Brandl S J, Casey J M and Meyer C P. 2020a. Dietary and habitat niche partitioning in congeneric cryptobenthic reef fish species. *Coral Reefs*. https://doi.org/10.1007/s00338-020-

01892-z

Brandl S J, Goatley C H R, Bellwood D R and Tornabene L. 2018. The hidden half: ecology and evolution of cryptobenthic fishes on coral reefs. *Biological Reviews*, *93*(4), 1846–1873. https://doi.org/10.1111/brv.12423

Brandl S J, Johansen J L, Casey J M, Tornabene L, Morais R A and Burt J A. 2020b. Extreme environmental conditions reduce coral reef fish biodiversity and productivity. *Nature Communications*. https://doi.org/10.1038/s41467-020-17731-2

Brandl S J, Tornabene L, Goatley C H R, Casey J M, Morais R A, Côté I M, Baldwin C C, Parravicini V, Schiettekatte N M D and Bellwood D R. 2019. Demographic dynamics of the smallest marine vertebrates fuel coral reef ecosystem functioning. *Science*, *364*(6446), 1189–1192. https://doi.org/10.1126/science.aav3384

Clark K, Karsch-Mizrachi I, Lipman D J, Ostell J and Sayers E W. 2016. GenBank. Nucleic Acids Research, Vol. 44, Database issue D67–D72 doi: 10.1093/nar/gkv1276

Collins R A, Bakker J, Wangensteen O S, Soto A Z, Corrigan L, Sims D W, Genner M J and Mariani S. 2019. Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution*, *10*(11), 1985–2001. https://doi.org/10.1111/2041-210X.13276

Collins R A, Wangensteen O S, O'Gorman E J, Mariani S, Sims D W and Genner M J. 2018. Persistence of environmental DNA in marine systems. *Communications Biology*, *1*(1), 1–11. https://doi.org/10.1038/s42003-018-0192-6

Conway K W, King C D, Summers A P, Kim D, Hastings P A, Moore G I, Iglésias S P, Erdmann M V, Baldwin C C, Short G, Fujiwara K, Trnski T, Voelker G and Rüber L. 2020. Molecular Phylogenetics of the Clingfishes (Teleostei: Gobiesocidae)-Implications for Classification. *Copeia*, *108*(4), 886–906. https://doi.org/10.1643/CI2020054

Deagle B E, Jarman, S N, Coissac E, Pompanon F, Taberlet P. 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biol Lett*. 10(9):20140562. doi: 10.1098/rsbl.2014.0562. PMID: 25209199; PMCID: PMC4190964.

de Santana C D, Parenti L R, Dillman C B, Coddington J A, Bastos D A, Baldwin C C, Zuanon J, Torrente-Vilara G, Covain R, Menezes N A, Datovo A, Sado T, Miya M. 2021. The critical role of natural history museums in advancing eDNA for biodiversity studies: a case study with Amazonian fishes. *Sci Rep*. 13;11(1):18159. doi: 10.1038/s41598-021-97128-3. PMID: 34518574; PMCID: PMC8438044.

Depczynski M and Bellwood D R. 2004. Microhabitat utilization patterns in cryptobenthic coral reef fish communities. *Marine Biology*, *145*(3), 455–463. https://doi.org/10.1007/s00227-004-1342-6

Depczynski M and Bellwood D R. 2005. Shortest recorded vertebrate lifespan found in a coral reef fish. *Current Biology*, *15*(8), 288–289.

DiBattista J D, Coker D J, Sinclair-Taylor T H, Stat M, Berumen M L and Bunce M. 2017. Assessing the utility of eDNA as a tool to survey reef-fish communities in the Red Sea. In *Coral Reefs* (Vol. 36, Issue 4, pp. 1245–1252). Springer Verlag. https://doi.org/10.1007/s00338-017-1618-1

Eschemeyer W N, Fricke R, Fong J, D and Polack D A. 2010. Marine fish diversity: history of knowledge and discovery (Pices). Zootaxa. 2525: 19-50

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol, 3: 294-299.

Gelis E R E, Kamal M M, Subhan B, Bachtiar I, Sani L M I, Madduppa H. 2021. Environmental biomonitoring of reef fish community structure with eDNA metabarcoding in the Coral Triangle. *Environ Biol Fish* 104:887-903. https:// doi.org/ 10. 1007/ s10641-21-1118-3

Greenfield D. 2017. An overview of the dwarfgobies, the second most speciose coral-reef fish genus (Teleostei: Gobiidae: *Eviota*). *Journal of the Ocean Science Foundation*, *29*, 32–54. https://doi.org/10.5281/zenodo.1115683

Greenfield D. 2021. Addendum to the 2016 key to the dwarfgobies (Teleostei: Gobiidae: *Eviota*). *Journal of the Ocean Science Foundation*, *38*, 1–12. https://doi.org/10.5281/zenodo.4458248

Greenfield D, Tornabene L, Gómez-Buckley M and Erdmann M. 2018. *Eviota maculosa*, a new dwarfgoby from the Western Pacific Ocean (Teleostei: Gobiidae). *Journal of the Ocean Science Foundation*. https://doi.org/10.5281/zenodo.1445810

Greenfield DW and Tornabene L. 2014. *Eviota brahmi* n. sp. from Papua New Guinea, with a redescription of Eviota nigriventris (Teleostei: Gobiidae). *Zootaxa*, 3793 (1):133-146.

Hebert P D N, Cywinska A, Ball S L and DeWaard J R. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, 270(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Hleap J S, Littlefair J E, Steinke D, Hebert P D N and Cristescu M E. 2021. Assessment of current taxonomic assignment strategies for metabarcoding eukaryotes. *Molecular Ecology Resources*, 0–40. https://doi.org/10.1111/1755-0998.13407

Juhel J B, Rizkie S U, Marques V, Vimono I B, Sugeha H, Kadarusman Y, Pouyaud L, Dejean T, Mouillot D and Hocdé R. 2020. Accumulation curves of environmental DNA sequences predict coastal fish diversity in the Coral Triangle. *Proc. R. Soc. B* 287: 20200248.http://dx.doi.org/10.1098/rspb.2020.0248

Kelly R P, Closek C J, O'Donnell J L, Kralj J E, Shelton A O and Samhouri J F. 2017. Genetic and manual survey methods yield different and complementary views of an ecosystem. *Frontiers in Marine Science*, *3*(JAN), 1–11. https://doi.org/10.3389/FMARS.2016.00283

Kelly R, Shelton A O and Gallego R. 2019. Understanding PCR processes to draw meaningful conclusions from environmental DNA studies. *Scientific Reports*. https://doi.org/10.1101/660530

Leray M, Yang J Y, Meyer C P, Mills S C, Agudelo N, Ranwez V, Boehm J T and Machida R J. 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, *10*(1). https://doi.org/10.1186/1742-9994-10-34

Lin H C and Hastings P A. 2013. Phylogeny and biogeography of a shallow water fish clade (Teleostei: Blenniiformes). *BMC Evolutionary Biology*, *13*(1). https://doi.org/10.1186/1471-2148-13-210

Longmire J, Malbi M and Baker R. 1997. Use of "lysis buffer" in DNA isolation and its implications for museum collections. *Museum of Texas Tech University*, 1–4.

Marwayana O N, Gold Z, Meyer C P and Barber P H. 2021. Environmental DNA in a global biodiversity hotspot: lessons from coral reef fish diversity across the Indonesian archipelago. *Environmental DNA*. 00:1-17

Mabuchi K, Fraser T H, Song H, Azuma Y and Nishida M. 2014. Revision of the systematics of the cardinalfishes (Percomorpha: Apogonidae) based on molecular analyses and comparative reevaluation of morphological characters. *Zootaxa* (Vol. 3846, Issue 2). https://doi.org/10.11646/zootaxa.3846.2.1

McClenachan L. 2015. Extinction risk in reef fishes. In C. Mora (Ed.), *Ecology of Fishes on Coral Reefs* (pp. 199–207). Cambridge University Press. https://doi.org/10.1017/CBO9781316105412.022

Miya M, Sato Y, Fukunaga T, Sado T, Poulsen J Y, Sato K, Minamoto T, Yamamoto S, Yamanaka H, Araki H, Kondoh M and Iwasaki W. 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of more than 230 subtropical marine species. *Royal Society Open Science*, *2*(7). https://doi.org/10.1098/rsos.150088

Nguyen B N, Shen E W, Seemann J, Correa A M S, O'Donnell J L, Altieri A H, Knowlton N, Crandall K A, Egan S P, McMillan W O and Leray M. 2020. Environmental DNA survey captures patterns of fish and invertebrate diversity across a tropical seascape. *Scientific Reports*, *10*(1), 1–14. https://doi.org/10.1038/s41598-020-63565-9

Pawlowski J, Apothéloz-Perret-Gentil L and Altermatt F. 2020. Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. *Molecular Ecology*, 29, 4258-4264.

Plaisance L, Knowlton N, Paulay G and Meyer C. 2009. Reef-associated crustacean fauna: Biodiversity estimates using semi-quantitative sampling and DNA barcoding. *Coral Reefs*, *28*(4), 977–986. https://doi.org/10.1007/s00338-009-0543-3

Plaisance L, Caley M J, Brainard R E and Knowlton N. 2011. The diversity of coral reefs: What are we missing? *PLoS ONE*, *6*(10). https://doi.org/10.1371/journal.pone.0025026

Randall J E, Williams J T, Smith D G, Kulbicki M, Tham G M, Labrosse P, Kroenen M, Clua E and Mann B S. 2003. Checklist of the shore and epipelagic fishes of Tonga. *Atoll Research Bulletin* (502): 1-35. https://doi.org/10.5479/si.00775630.502.1

Ratnasingham S and Hebert P D. 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes, 7*:355–364. doi: 10.1111/j.1471-8286.2006.01678.x

Robertson D R and Smith-Vaniz W F. 2010. Use of clove oil in collecting coral reef fishes for research. *Marine Ecology Progress Series*, *401*(February), 295–302. https://doi.org/10.3354/meps08374

Robertson D R and Smith-Vaniz W F. 2008. Rotenone: An essential but demonized tool for assessing marine fish diversity. *BioScience*, *58*(2), 165–170. https://doi.org/10.1641/B580211

Shelton A O, O'Donnell J L, Samhouri J F, Lowell N, Williams G D and Kelly R P. 2016. A framework for inferring biological communities from environmental DNA. *Ecological Applications*, *26*(6), 1645–1659. https://doi.org/10.1890/15-1733.1

Taylor M S and Hellberg M E. 2005. Marine radiations at small geographic scales: Speciation in Neotropical reef gobies (*Elacatinus*). *Evolution*, *59*(2), 374-385.

Thomsen P F and Willerslev E. 2015. Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*(December), 4–18. https://doi.org/10.1016/j.biocon.2014.11.019

Tornabene L, Valdez S, Erdmann M V and Pezold F L. 2016. Multi-locus sequence data reveal a new species of coral reef goby (Teleostei: Gobiidae: *Eviota*), and evidence of Pliocene vicariance across the Coral Triangle. *Journal of Fish Biology*, *88*(5), 18a11–1834. https://doi.org/10.1111/jfb.12947

Tornabene L, Ahmadia G N, Berumen M L, Smith D J, Jompa J and Pezold F. 2013. Evolution of microhabitat association and morphology in a diverse group of cryptobenthic coral reef fishes (Teleostei: Gobiidae: *Eviota*). *Molecular Phylogenetics and Evolution*, *66*(1), 391–400. https://doi.org/10.1016/j.ympev.2012.10.014

Tornabene L, Ahmadia G N and Williams J T. 2013. Four new species of dwarfgobies (Teleostei: Gobiidae: *Eviota*) from the Austral, Gambier, Marquesas and Society Archipelagos, French Polynesia. *Systematics and Biodiversity*, 11, 363-380.

https://doi.org/10.1080/14772000.2013.819822

Tornabene L, Greenfield D W and Erdmann M V. 2021. A review of the *Eviota zebrina* complex, with descriptions of four new species (Teleostei, Gobiidae). *ZooKeys*, *1057*, 149–184. https://doi.org/10.3897/zookeys.1057.66675

Tornabene L, Valdez S, Erdmann M and Pezold F. 2015. Support for a "Center of Origin" in the Coral Triangle: Cryptic diversity, recent speciation, and local endemism in a diverse lineage of reef fishes (Gobiidae: *Eviota*). *Molecular Phylogenetics and Evolution*, *82*(PA), 200–210. https://doi.org/10.1016/j.ympev.2014.09.012

Troyer E M, Coker D J and Berumen M L. 2018. Comparison of cryptobenthic reef fish communities among microhabitats in the Red Sea. *PeerJ*, *2018*(6). https://doi.org/10.7717/peerj.5014

Tsuji S, Ushio M, Sakurai S, Minamoto T and Yamanaka H. 2017. Water temperature-dependent degradation of environmental DNA and its relation to bacterial abundance. *PLoS ONE*, *12*(4), 1–13. https://doi.org/10.1371/journal.pone.0176608

Turner C R, Uy K L and Everhart R C. 2015. Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation*, *183*, 93–102. https://doi.org/10.1016/j.biocon.2014.11.017

Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen P F, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp G H, Geniez P, Pont D, Argillier C, Baudoin J M, … Dejean T. 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, *25*(4), 929–942. https://doi.org/10.1111/mec.13428

Victor B C. 2015. How many coral reef species are there? Cryptic diversity and the new molecular taxonomy. Ecology of Fishes on Coral Reefs (ed. C. Mora), pp. 76-87. Cambridge University Press, Cambridge.

Weigt L, A, Baldwin C, C, Driskell A, Smith D, G Ormos A and Reyier E A. 2012. Using SNA barcoding to assess Caribean reef Fish biodiversity: Expanding taxonomic and geographic coverage. *PLoS ONE* 7(7): e41059. doi:10.1371/journal.pone.0041059

Winterbottom R, Hanner R H, Burridge M and Zur M. 2014. A cornucopia of cryptic species - A DNA barcode analysis of the gobiid fish genus Trimma (Percomorpha, Gobiiformes). *ZooKeys*, *381*, 79–111. https://doi.org/10.3897/zookeys.381.6445

Winterbottom R, Hanner R H, Erdmann M V, Choffe K, Zur M and Steinke C. 2020. The cryptic cornucopia revisited – an extended analysis of the COI gene in the gobiid fish genus *Trimma* (Percomorpha: Gobiiformes). *Journal of the Ocean Science Foundation*, 36, 91–132. https://doi.org/10.5281/zenodo.4403739

Zhang S, Zhao J and Yao, M. (2020). A comprehensive and comparative evaluation of primers for metabarcoding eDNA from fish. *Methods in Ecology and Evolution*, *11*(12), 1609–1625. https://doi.org/10.1111/2041-210X.13485

CHAPTER 1 FIGURES

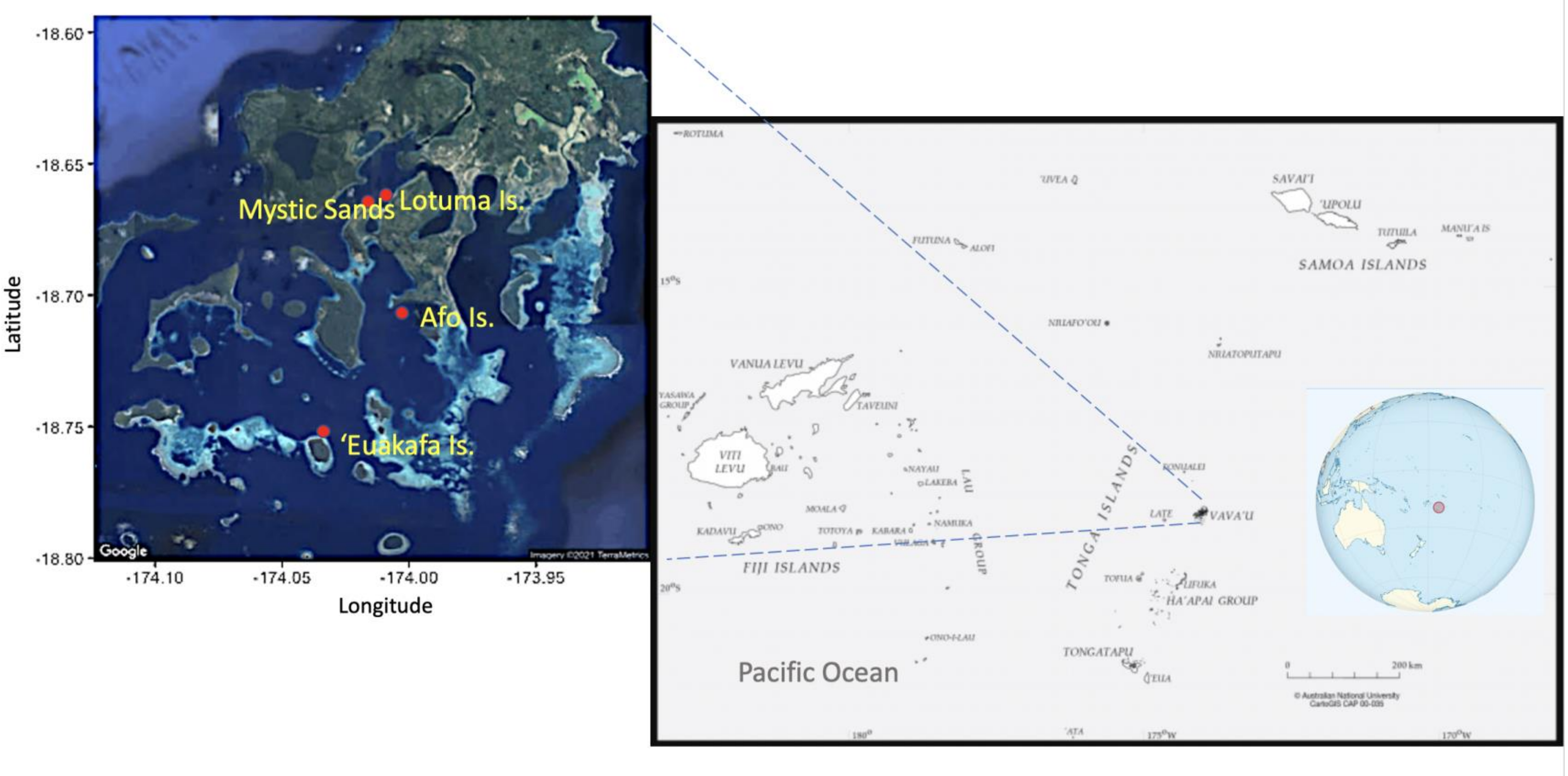


Fig. 1. Sampling sites in the Vava'u Island group (red dots in the inset map).

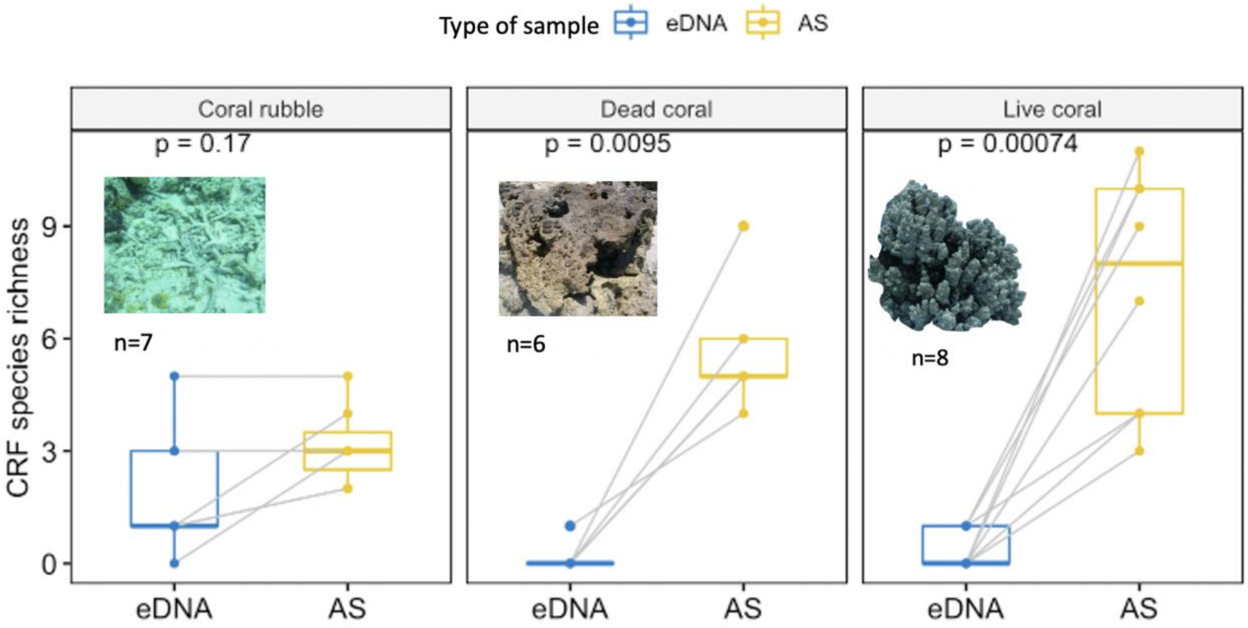


Figure 2. Cryptobenthic reef fishes (CRF) species richness of eDNA and anesthetic stations (AS) samples for the three microhabitats sampled. Grey lines show CRF species richness for the eDNA and corresponding AS samples.

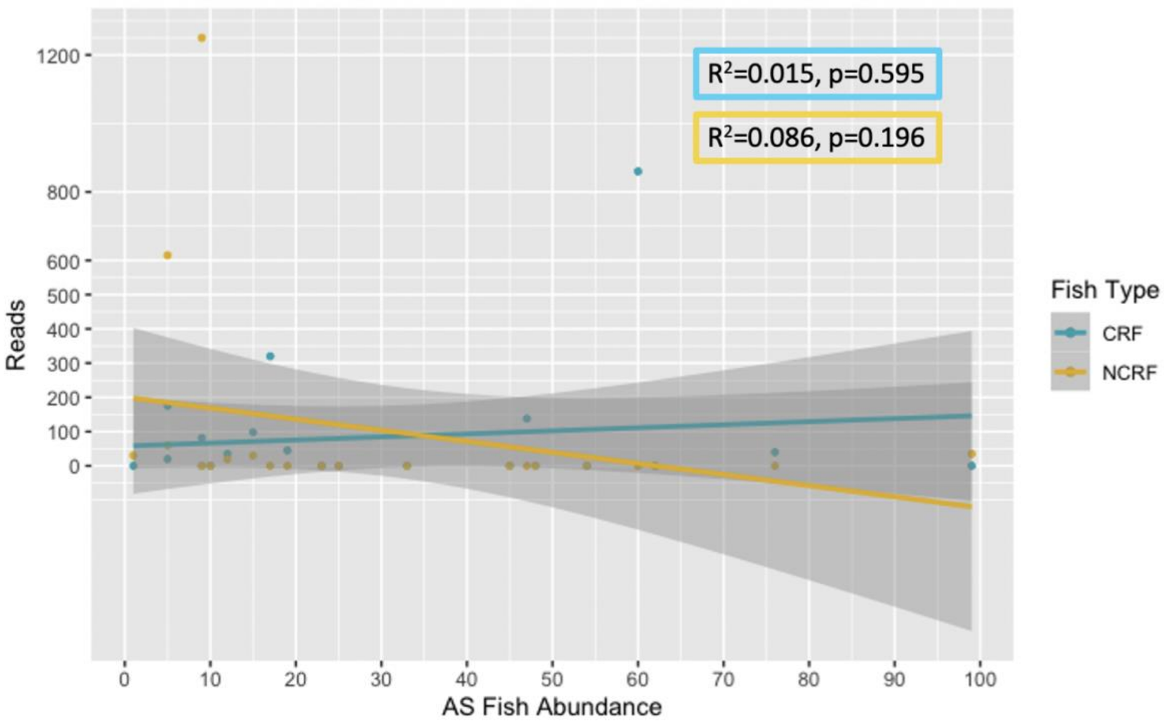
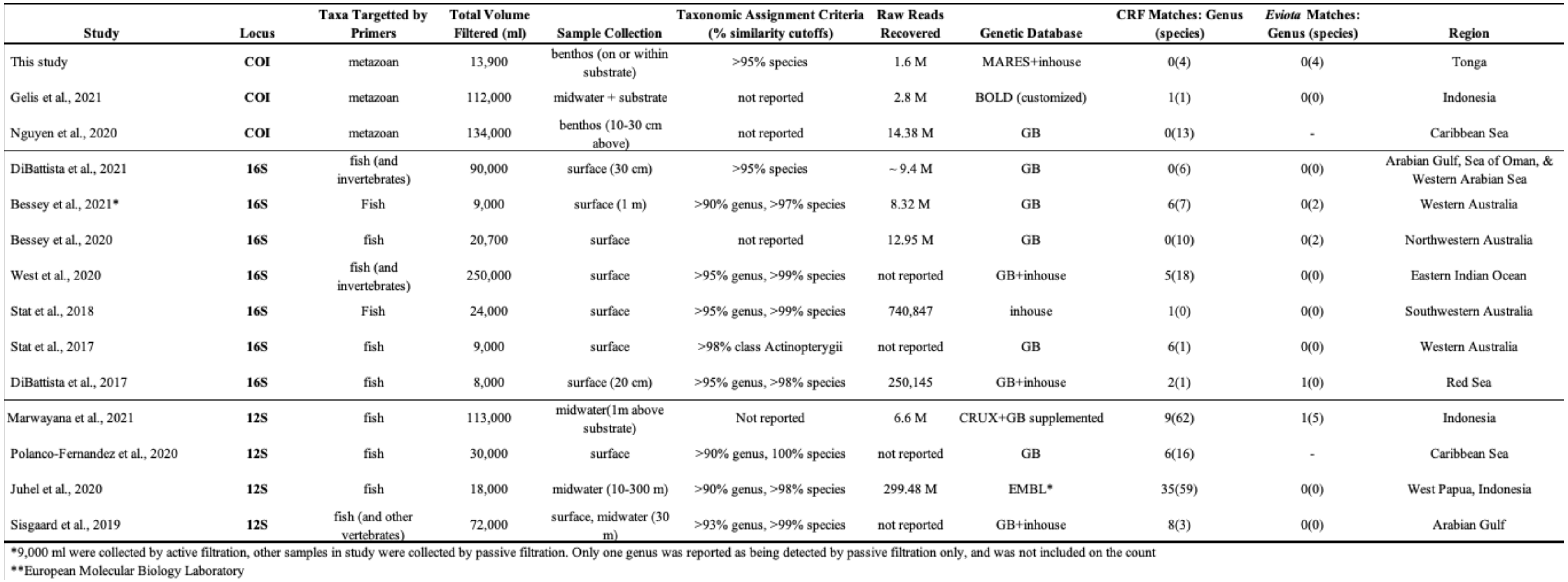


Figure 3. Read abundance for cryptobenthic reef fishes (CRF) and non-CRF in relation to the number of fishes collected from cospatial anesthetic stations (AS).

CHAPTER 1 TABLES

Table 1. Studies published since 2017 using environmental DNA samples (eDNA) to detect fishes in tropical environments. Studies conducted in the Caribbean region, (Polanco-Fernandez et al., 2020, and Nguyen et al., 2020) with no *Eviota* representatives are blank.



Study	Locus	Taxa Targetted by Primers	Total Volume Filtered (ml)	Sample Collection	Taxonomic Assignment Criteria (% similarity cutoffs)	Raw Reads Recovered	Genetic Database	CRF Matches: Genus (species)	*Eviota* Matches: Genus (species)	Region
This study	**COI**	metazoan	13,900	benthos (on or within substrate)	>95% species	1.6 M	MARES+inhouse	0(4)	0(4)	Tonga
Gelis et al., 2021	**COI**	metazoan	112,000	midwater + substrate	not reported	2.8 M	BOLD (customized)	1(1)	0(0)	Indonesia
Nguyen et al., 2020	**COI**	metazoan	134,000	benthos (10-30 cm above)	not reported	14.38 M	GB	0(13)	-	Caribbean Sea
DiBattista et al., 2021	**16S**	fish (and invertebrates)	90,000	surface (30 cm)	>95% species	~ 9.4 M	GB	0(6)	0(0)	Arabian Gulf, Sea of Oman, & Western Arabian Sea
Bessey et al., 2021*	**16S**	Fish	9,000	surface (1 m)	>90% genus, >97% species	8.32 M	GB	6(7)	0(2)	Western Australia
Bessey et al., 2020	**16S**	fish	20,700	surface	not reported	12.95 M	GB	0(10)	0(2)	Northwestern Australia
West et al., 2020	**16S**	fish (and invertebrates)	250,000	surface	>95% genus, >99% species	not reported	GB+inhouse	5(18)	0(0)	Eastern Indian Ocean
Stat et al., 2018	**16S**	Fish	24,000	surface	>95% genus, >99% species	740,847	inhouse	1(0)	0(0)	Southwestern Australia
Stat et al., 2017	**16S**	fish	9,000	surface	>98% class Actinopterygii	not reported	GB	6(1)	0(0)	Western Australia
DiBattista et al., 2017	**16S**	fish	8,000	surface (20 cm)	>95% genus, >98% species	250,145	GB+inhouse	2(1)	1(0)	Red Sea
Marwayana et al., 2021	**12S**	fish	113,000	midwater(1m above substrate)	Not reported	6.6 M	CRUX+GB supplemented	9(62)	1(5)	Indonesia
Polanco-Fernandez et al., 2020	**12S**	fish	30,000	surface	>90% genus, 100% species	not reported	GB	6(16)	-	Caribbean Sea
Juhel et al., 2020	**12S**	fish	18,000	midwater (10-300 m)	>90% genus, >98% species	299.48 M	EMBL*	35(59)	0(0)	West Papua, Indonesia
Sisgaard et al., 2019	**12S**	fish (and other vertebrates)	72,000	surface, midwater (30 m)	>93% genus, >99% species	not reported	GB+inhouse	8(3)	0(0)	Arabian Gulf

*9,000 ml were collected by active filtration, other samples in study were collected by passive filtration. Only one genus was reported as being detected by passive filtration only, and was not included on the count

**European Molecular Biology Laboratory

Table 2. Sampling effort and site coordinates for the eDNA and anesthetic stations. Last row for "Total" information includes in parenthesis the combined volume of water (ml) filtered for eDNA analysis from each microhabitat.

	Site coordinates		Cospatial eDNA and anesthetic stations			
Site	Latitude	Longitude	Live coral	Dead coral	Coral rubble	**Totals**
Lotuma Island	18°39'43.0"S	174°00'32.0"W	4	1	1	**6**
Mystic Sands	18°39'53.9"S	174°00'57.7"W	2	4	2	**8**
Afo Island	18°42'30.1"S	173°59'47.3"W	2	-	3	**5**
Euakafa Island	18°45'14.7"S	174°02'12.9"W	-	1	1	**2**
Total samples (ml of water)			**8(5,500)**	**6(2,800)**	**7(5,400)**	**21(13,900)**

Table 3. Summary of eDNA sequences recovered from entire study.

	Number of Reads	Percent (from initial raw reads)	Distinct ASVs	Unique taxa (fish species)
Total raw sequences	1,686,370	100	-	-
Marine eukaryotes*	313,184	18.6	1,211	347
Actinopterygii	3,851	0.2	39	22
Cryptobenthic Fishes	1,812	0.1	14	4

(*) Using MARES + Eviota database

Table 4. Fish species collected from cospatial anesthetic stations (AS) and eDNA samples. For the eDNA samples, only the species with >95% identity from the bioinformatic processing were included. For the AS all fish collected were included.

		eDNA samples where species were detected	AS where species were collected	Reads in eDNA samples	Fishes in AS	Available COI seqs	
CRFs - Family	**CRFs - Species**					**Genus**	**Species**
Apogonidae	*Cercamia eremia*	0	1	0	1	Y	Y
	Fowleria marmorata	0	1	0	1	Y	Y
Blenniidae	*Atrosalarias fuscus*	0	2	0	2	Y	Y
	Ecsenius bicolor	0	1	0	1	Y	Y
Gobiidae	*Amblygobius sphynx*	0	1	0	1	Y	Y
	Asterropteryx semipunctata	0	10	0	53	Y	Y
	Cabillus lacertops	0	1	0	1	Y	Y
	Cabillus cf. lacertops	0	2	0	2	Y	Y
	Eviota albolineata	0	1	0	1	Y	Y
	Eviota cf. punctulata	1	2	8	4	Y	Y
	Eviota cometa	0	2	0	5	Y	Y
	Eviota Fiji sp. 1	4	0	404	0	Y	Y
	Eviota herrei	0	2	0	4	Y	Y
	Eviota Papua New Guinea sp. 1	2	0	8	0	Y	Y
	Eviota prasites	0	4	0	41	Y	Y
	Eviota sebreei	0	4	0	6	Y	Y
	Eviota smargadus	0	1	0	1	Y	Y
	Eviota Tonga sp.1	8	11	1,392	291	Y	Y
	Eviota Tonga sp. 2	0	3	0	4	Y	Y
	Eviota Tonga sp. 3	0	1	0	1	-	-
	Eviota sp. (juvenile)	0	3	0	3	-	-
	Eviota sparsa	0	9	0	31	Y	Y
	Eviota tigrina	0	12	0	129	Y	Y
	Fusigobius iframaculatus	0	3	0	3	Y	Y
	Fusigobius cf. aureus	0	2	0	4	Y	N
	Gobiidae sp (juvenile)	0	10	0	33	-	-
	Grallenia sp.	0	1	0	1	N	N
	Koumansetta rainfordi	0	7	0	14	Y	Y
	Trimma benjamini	0	7	0	32	Y	Y
Oxudercidae	*Gnatholepis cauerensis*	0	1	0	1	Y	Y
Plesiopidae	*Plesiops Oxycephalus*	0	1	0	1	Y	N
Pseudochromidae	*Cypho purpurascens*	0	3	0	4	Y	Y
Syngnathidae	*Festucalex gibssi*	0	1	0	1	Y	N
	CRF Totals	17	13*	1,812	677		
Non-CRFs - Family	**Non-CRFs - Species**						
Acanthuridae	*Nasso brevirostris*	1	0	1,105	0	Y	Y
Chaetodontidae	*Heniochus chrysostomus*	1	0	595	0	Y	Y
Labridae	*Halichoeres trimaculatus*	1	0	35	0	Y	Y
	Iniistius melanopus	1	0	30	0	Y	Y
	Labridae sp	0	1	0	1	-	-
Pomacentridae	*Chromis viridis*	1	0	45	0	Y	Y
	Pomacentrus burroughi	0	3	0	3	Y	Y
	Pomacentrus callainus	1	0	9	0	Y	Y
	Pomacentrus lepidogenys	1	0	6	0	Y	Y
	Pomacentrus maafu	1	0	5	0	Y	Y
	Pomacentrus moluccensis	1	0	14	0	Y	Y
	Pomacentrus microspilus	1	0	4	0	Y	Y
	Pomacentrus nagasakiensis	1	0	16	0	Y	Y
	Pomacentrus sp. (juvenile)	0	3	0	7	-	-
Scombridae	*Rastrelliger kanagurta*	1	0	30	0	Y	Y
Serranidae	Serranidae sp. (juvenile)	0	1	0	1	-	-
Scorpaenidae	*Scorpaenodes hirsutus*	0	1	0	1	Y	Y
Synodontidae	*Saurida gracillis*	1	0	145	0	Y	Y
Unidentified	Unidentified	0	2	0	4	-	-
	Non-CRF Totals	13	11*	2,039	NA		

* Some of the counts may be overlaps from same sample.

CHAPTER 1 SUPPLEMENT

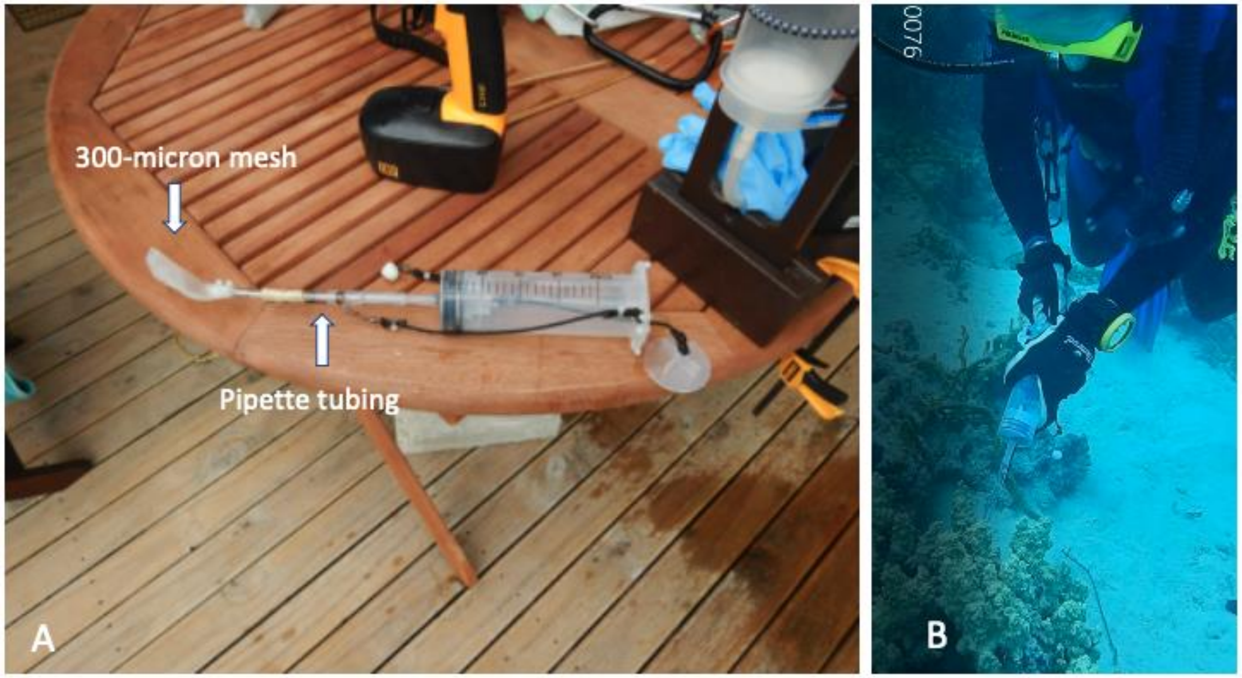


Figure S1. (A) Modified 200ml syringe used for *in situ* eDNA samples showing the pipette tubing extension. (B) sampling within the crevices using the modified syringe. (Photo credit Ray Buckley)

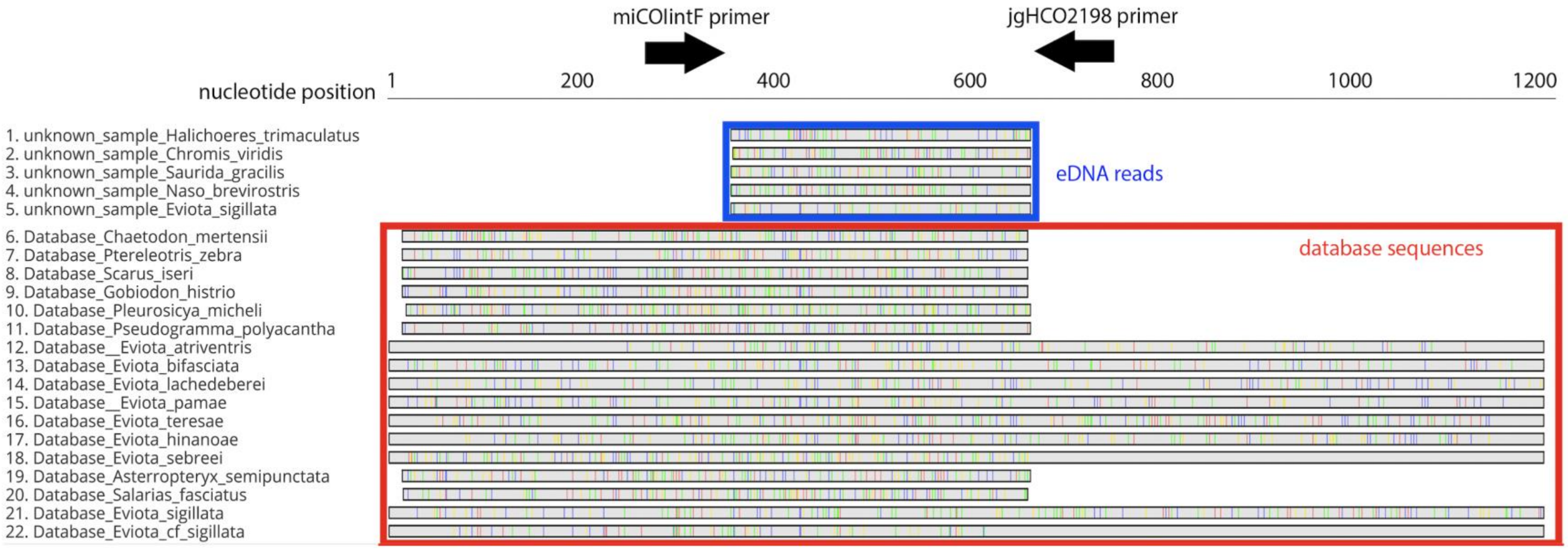


Figure S2. Subset of reads from eDNA samples (1-5) showing the alignment within an example set of typical COI sequences (6-22) in the used reference database.

Table S1. Fish families and species collected. Number of specimens in anesthetic stations (AS), and percent occurrence in the eDNA samples in the three microhabitats sampled.

		Live coral (8)		**Coral rubble (7)**		**Dead coral (6)**		**Total (21)**
CRFs - Family	**CRFs - Species**	**Fishes in AS**	**% Occurrence in eDNA samples**	**Fishes in AS**	**% Occurrence in eDNA samples**	**Fishes in AS**	**% Occurrence in eDNA samples**	**Fishes collected (AS)**
Apogonidae	*Cercamia aremia*					1		**1**
	Fowleria marmorata	1						**1**
Blenniidae	*Astrosalarias fucus*	2						**2**
	Ecsenius bicolor	1						**1**
Gobiidae	*Amblygobius sphynx*					1		**1**
	Asterropteryx semipunctata	5		44		4		**53**
	Cabillus lacertops					1		**1**
	Cabillus cf. lacertops			1		1		**2**
	Eviota albolineata					1		**1**
	Eviota cf. punctulata	1			8	3		**4**
	Eviota cometa	5						**5**
	Eviota Fiji sp. 1				31			
	Eviota herrei	2				2		**4**
	Eviota Papua New Guinea sp. 1				30			
	Eviota prasites	41						**41**
	Eviota sebreei	5				1		**6**
	Eviota smargadus					1		**1**
	Eviota Tonga sp.1	78	100	135	31	78	100	**291**
	Eviota Tonga sp. 2					4		**4**
	Eviota Tonga sp. 3	1						**1**
	Eviota sp. (juvenile)	2				1		**3**
	Eviota sparsa	25		3		3		**31**
	Eviota tigrina	29		11		89		**129**
	Fusigobius inframaculatus	2				1		**3**
	Fusigobius cf. aureus	4						**4**
	Gobiidae sp (juvenile)	14		7		12		**33**
	Grallenia sp.					1		**1**
	Koumasetta rainfordi	14						**14**
	Trimma benjamini	19				13		**32**
Oxudercidae	*Gnatholepis caurensis*					1		**1**
Plesiopidae	*Plesiops Oxycephalus*					1		**1**
Pseudochromidae	*Cypho purpurascens*	2				2		**4**
Syngnathidae	*Festocalex gibsii*			1				**1**
Total (CRFs in ASs and ocurrences in eDNA)		**253**	**3**	**202**	**13**	**222**	**1**	**677**
Non-CRFs - Family	**Non-CRFs - Species**							
Acanthuridae	*Nasso brevirostris*		50					
Chaetodontidae	*Heniochus chrysostomus*				16.6			
Labridae	*Halichoeres trimaculatus*						20	
	Iniistius melanopus						20	
	Labridae sp			1				**1**
Pomacentridae	*Chromis viridis*						20	
	Pomacentrus burroughi			1				**1**
	Pomacentrus callainus						20	
	Pomacentrus lepidogenys						20	
	Pomacentrus maafu				16.6			
	Pomacentrus moluccensis				16.6			
	Pomacentrus microspilus				16.6			
	Pomacentrus nagasakiensis				16.6			
	Pomacentrus sp. (juvenile)			1		1		**2**
Scombridae	*Rastrelliger kanagurta*				16.6			
Serranidae	Serranidae sp. (juvenile)	1						**1**
Scorpaenidae	*Scorpanodes hirsutus*	1						**1**
Synodontidae	*Saurida gracillis*		50					
Unidentified	Unidentified	2					5	**7**
Total (Non-CRFs in ASs and ocurrences in eDNA)		**4**	**2**	**3**	**6**	**1**	**5**	**13**

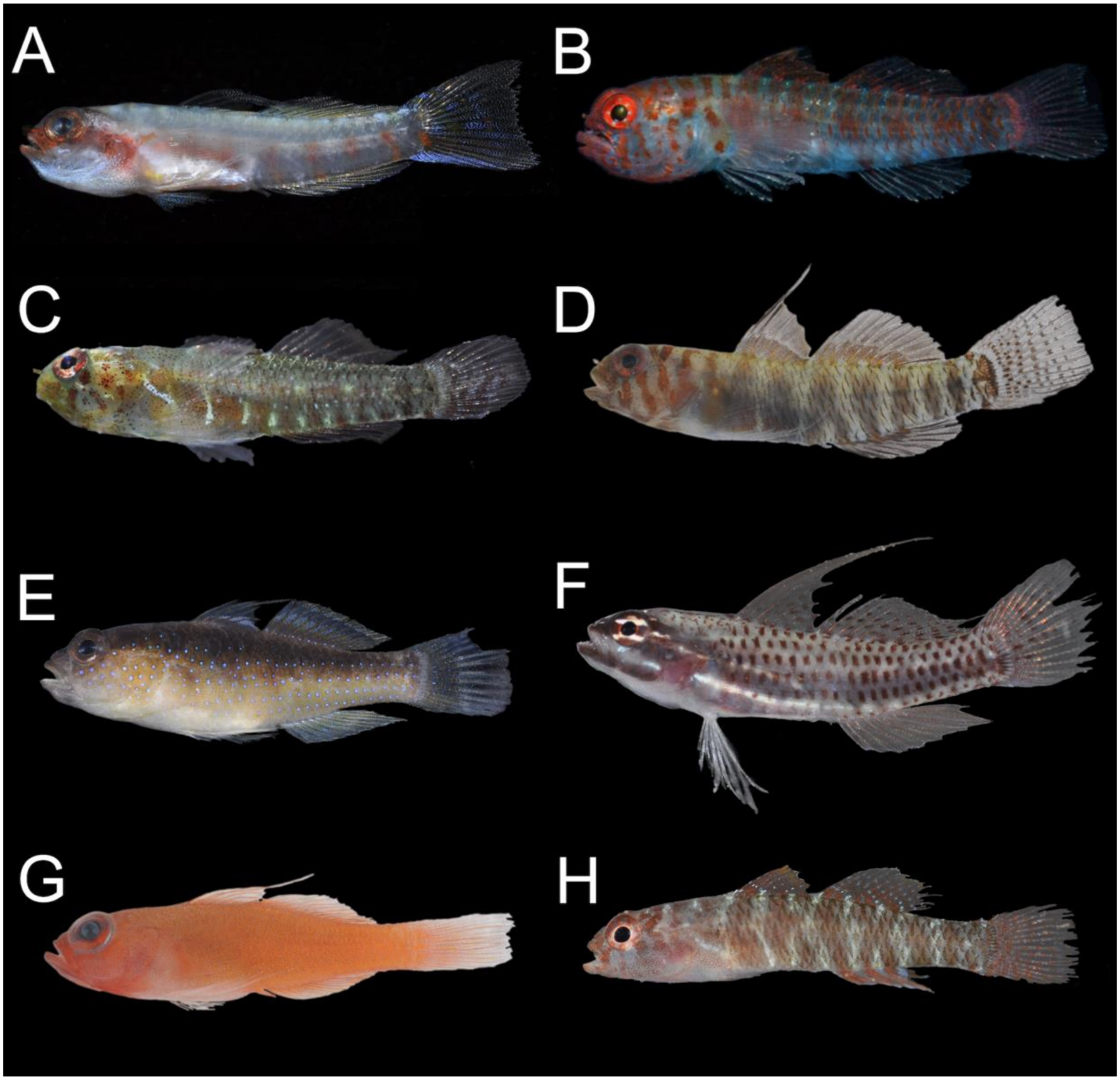
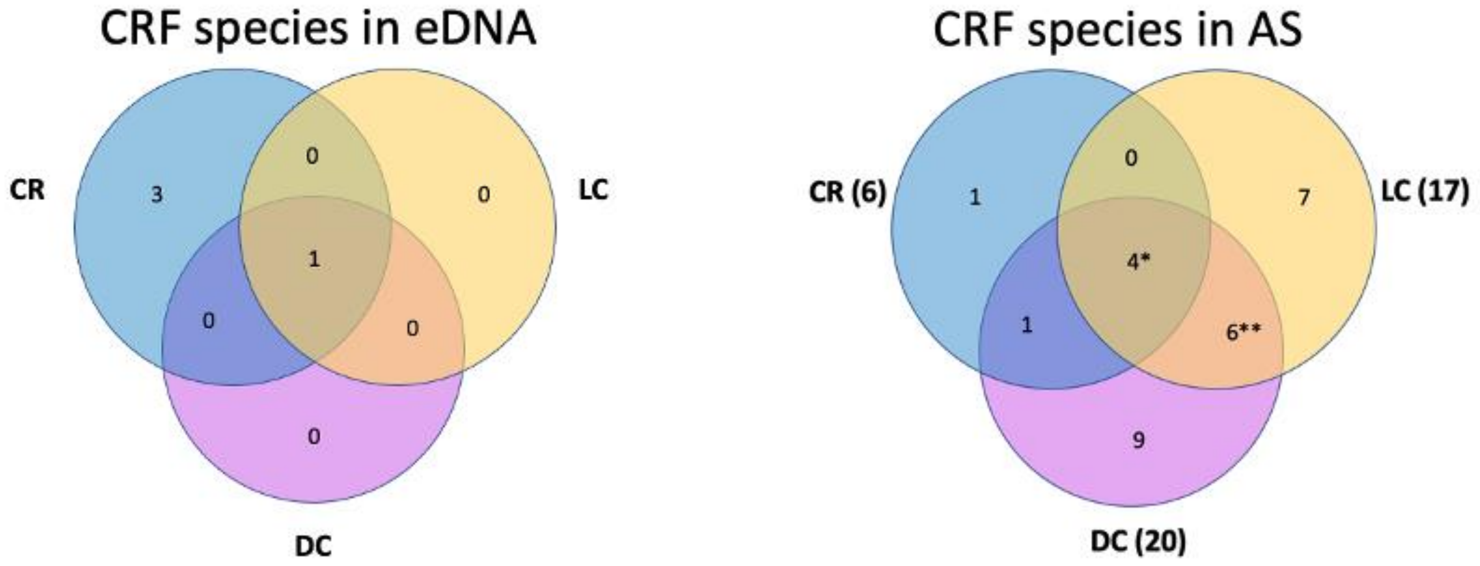


Figure S3. Representative CRF species detected in eDNA samples (A-C) and from anesthetic stations (D-H). (A) *Eviota* Tonga sp.1; (B) *Eviota* Fiji sp. 1; (C) *Eviota cf. punctulata*; (D) *Eviota tigrina*; (E) *Asterropteryx semipunctata*; (F) *Eviota prasites*; (G) *Trimma benjamini*; (H) *Eviota sparsa*.



CRF species in eDNA

CRF species in AS

*Several (33) unidentified Gobiidae juveniles were also collected from the three microhabitats

** Several (3) unidentified *Eviota* juveniles were also collected from LC(2) and DC(1) microhabitats

Figure S4. CRF species richness in eDNA samples and in Anesthetic stations (AS) showing for each microhabitat type, coral rubble (CR), live coral (LC), and dead coral (DC). For AS, the numbers in parentheses by microhabitat types represents the total number of species for that microhabitat.

Site description

We chose four sites (Figure 1) characterized by flat or low-slope profile bottoms, with a relatively homogenous mix of live coral bommies, dead coral rock, coral rubble, and sand substrates, at 3-7m depth. All sites are protected from high swell or high currents under normal weather conditions. Two of the sites, Lotuma Island (18°39'43.0"S 174°00'32.0"W) and Mystic Sands (18°39'53.9"S 174°00'57.7"W) are in close proximity to one another (0.8 km) and near populated areas, while Afo Island (18°42'30.1"S 173°59'47.3"W) and Euakafa Island (18°45'14.7"S 174°02'12.9"W) are separated from the other two sites by about 8 and 11km respectively, and about 6 km apart from each other. The latter two sites/islands are unpopulated, with Euakafa Island being the farthest away from the main populated areas. All sites except for Mystic Sands were located outside of the established Special Management Areas (SMA) to avoid interference with the Tongan Government efforts to restore and sustain village fishery resources (Smallhorn-West et al., 2020). For the Utungake SMA, of which Mystic Sands is part of, permission was granted from the village chief, and CRF captures were made in the SMA area outside of the delimited “No Take Zone”.

At each site we chose three different microhabitat types: live coral bommies (LC), dead coral rock (DC), and coral rubble (CR). Dead coral rock still had the structural complexity of live coral, as well as a biotic layer of mostly encrusting algae and sponges, whereas coral rubble habitat consisted of a dense matt of small fragments of broken coral skeleton on otherwise flat surface. Sand habitats were not included because previous exploratory surveys had shown the near absence of CRF in sand habitats when compared to the other microhabitat types. To standardize CRF collections from live coral heads, only structurally similar branching *Porites sp.* and *Montipora sp.* coral heads were selected. These two corals were present at all the sites as they are resilient corals that can grow fast and colonize areas of substrate that have experienced disturbances after natural or anthropogenic impacts (Barnhill & Bahr, 2019; Donner & Carilli, 2019). Constraints of time, boat availability, extent of the coral degradation, and distance of some of the sites from the base of operation, dictated the number of samples taken from each microhabitat type at each site.

Sample collections

eDNA samples were either collected at the substrate surface (when interstices were absent), or within the interstices of the coral rubble or the coral heads, using 200 ml syringes modified by having the tip extended with a 6 cm long, 1 cm diameter length of plastic tubing tipped with a small section of 300 µm mesh, to avoid sucking in debris and large organisms (Figure S1). This prevented obstructions that would interfere with collecting the water sample or foul the sample by overrepresentation in the sequenced data of larger organisms like micro-invertebrates. The total open surface of this mesh (combined surface of all square openings), was over four times larger than the mouth opening of the syringe opening, maximizing this way the intake of water without interfering with the eDNA signal in the water sampled. The tube and mesh were removed from each syringe after sample collection and the syringe was capped to isolate the water sample. The syringes were immediately placed in jars previously filled with fresh water to further prevent contamination from the surrounding sea water. The use of 200 ml syringes allowed us for the rapid collection of eDNA samples, thus reducing the potential for including water from the surrounding environment. All eDNA samples were placed in a cooler when brought to the surface and each sample was filtered as soon as possible (within 3 hours of collection) through a sterile 250 ml disposable filter funnel with a 0.45 Micron, 47 mm diameter nitrocellulose mixed ester (MCE) membrane filter (Sterlithech Corporation, Kent, WA) using a peristaltic pump powered by a hand drill. Due to time constraints and the haphazard distribution and availability of the three sampled microhabitats, we had different total volumes filtered for each station and microhabitat type, with 5,500 ml, 5,400 ml, and 2,800 ml filtered from live coral, coral rubble, and dead coral respectively. Each filter disc was preserved in a labeled 2 ml plastic micro-vial filled with 900 µl of Longmire's solution (Longmire et al., 1997). To fit the filter disc in the 2ml vials, the filters were folded in half three times with the use of two flat tip tweezers. All the equipment that was reused (syringes, filter funnels, and tweezers) was sterilized between sampling, filtering, and filter storage operations, by soaking it in baths of 10% bleach solution, and thorough rinsing in tap water boiled for ten minutes. Disposable nitrile gloves were used and replaced for every filtered sample and whenever handling the eDNA equipment.

When a haphazardly selected microhabitat type was located, a cylindrical mesh enclosure (open at the top and the bottom) with lead line at the bottom and float line on the top, was placed over the microhabitat to sequester as best as possible all CRF. The enclosure is designed to be able to adapt to the benthic contour of the sampled microhabitat and encompass an area of approximately 1 m^2. A canvas tarp with lead weights in each corner was then placed over the enclosure and a concentrated solution of quinaldine sulfate anesthetic powder diluted in sea water at 1 part powder to 2 parts water (Neiffer & Stamper, 2009), was delivered using squeezable plastic jars inside the enclosure under the tarp. The tarp was gently pushed down to mix the anesthetic and to evenly distribute it in the whole volume of the enclosure. For some of the live coral heads that were too large for the mesh enclosure, only the tarp was used. After about a minute, when many of the fishes were observed to be responding to the anesthetic, the tarp was lifted slowly and all the CRF were suctioned into a labeled plastic jar using an airlift device designed specifically for this study.

Specimens from each habitat station were humanely euthanized with MS222 and placed in seawater cooled with jars of ice to keep them from deteriorating and to preserve the specimen's live coloration. They were photographed in a glass or acrylic photo tank to aid in later detailed morphological analysis. Specimens were photographed with a Nikon D90 digital camera, Nikon speedlight SB 800 external strobes, and an Oshiro 60mm f/2.8 2:1 LD UNC Ultra-Macro lens. After the photographs were taken, each specimen was placed in individual 2 ml cryovials with unique label identifiers and filled with 95% ethanol for later genetic analysis.

Back at the University of Washington's Fish Collection lab, all the CRF specimens that could not be identified at the time of collection were further identified to the lowest taxonomic level possible by the main author and co-author Luke Tornabene, both with extensive expertise in CRF systematics (i.e., specimen identification using taxonomic keys, describing new CRF species and writing keys for them). A combination of the specimens' live coloration photographs taken when collected, examination of morphological features, as well as the use of the most updated available taxonomic keys for the genera *Eviota* (Greenfield, 2021; Greenfield &

Winterbottom, 2016; Carpenter & Niem, 2001), *Cabillus* (Kovacic & Bogorodsky, 2017), *Festucalex* (Syngnathidae, Carpenter & Niem, 2001), *Grallenia* (Allen & Hammer, 2018), *Fusigobius* (Randall, 2001), *Gnatholepis* (Larson & Buckle, 2012), and *Koumansetta* (Kovacic, et al. 2018), aided in the taxonomic identification. The main morphological features were examined using a Zeiss Discovery V20 SteREO microscope. These included, among others: dorsal and anal fin spines and ray counts, head sensory pore patterns, presence/absence of bifurcated pectoral fin rays, and pelvic fin features like number of segmentations and branching on the fourth pelvic fin ray.

Specimen permits and ethics statement

Permits where obtained in advance with the assistance of the collaborating NGO VEPA, from the Ministry of Fisheries of the Government of Tonga. Specimens are currently being archived at the University of Washington Fish Collection (UWFC) at the Burke Museum of Natural History and Culture. Procedures for collecting and euthanizing specimens, followed the protocols set forth by the University of Washington's Office of Animal Welfare and IACUC protocol 4468-01.

eDNA lab processing - Protocols of the two-step PCR; gel electrophoresis; PCR amplicon clean-up; barcoding PCR; gel electrophoresis; PCR normalization and pooling

A dedicated lab with sterilized work surfaces was used to perform DNA extractions from the eDNA samples. Each whole filter disc corresponding to individual sample replicates were extracted and kept as a separate sample for the subsequent amplification and sequencing processing. Prior to the extraction, a 'bead beating' processing (Djurhuus et al., 2017) with slight modifications was used to help separate DNA fragments from the filter. The glass beads (BioSpec Products, Bartlesville, Oklahoma, USA) were added to the vials that contained the filters in the Longmire's solution, and Qiagen ATL buffer was added to the vials after the bead beating process. A DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) was then used for the extraction following the manufacturer's protocols. All extracted samples, including negative controls, were quantitated using the Invitrogen Qubit 4 Fluorimeter (Invitrogen, Carlsbad, California, USA).

A 5' adaptor sequence was added to the reverse and forward primers to serve as indexing during Illumina sequencing. The Promega PCR Master Mix specifications (Promega catalog # M5133, Madison, WI) were followed to mix each 25µl PCR reaction, including: 12.5µl Master Mix, 0.5 µM of each primer, 1.0µl of gDNA, and 10.5µl DNase/RNase-free H_2O. For each PCR, controls were incorporated. The conditions for amplification were initial denaturation at 94 °C for 2 minutes, followed by 45 cycles of 15 seconds at 94 °C, 30 seconds at 50 °C, and 1 minute at 72 °C, and a final elongation at 72 °C for 10 minutes. For amplicon size determination, PCR efficiency, and to visualize each reaction, a 2% agarose gel was used, using 5µl per reaction. The amplicons were cleaned by incubation with an Exo1/SAP during 30 minutes at 37 °C, followed by inactivation at 95 °C during 5minutes, and then stored at -20 °C. To give each sample a unique 12-nucleotide index sequence, a second round of PCR was performed. For the indexing PCR, 2 µl template DNA from the cleaned amplicon resulting from the initial PCR, was mixed with Promega Master mix, and 0.5 µM of each primer. The conditions for amplification of this second PCR were: Initial denaturation of 95 °C for 3 minutes followed by 8 cycles of 95 °C for 30 sec, 55 °C for 30 seconds and 72 °C for 30 seconds. Indexed amplicons from each sample were cleaned and normalized in SequalPrep Normalization Plates (Life Technologies, Carlsbad, CA). Each PCR amplicon (25µl) was purified and normalize using the Life Technologies SequalPrep Normalization kit (cat#A10510-01) according to the manufacturer's protocol. Finally, the samples were pooled together by adding 5µl of each normalized sample to the pool. All the sample library pools were sent for sequencing to the CU Boulder BioFrontiers Sequencing Center (San Diego, CA), using an Illumina Mi Seq, with a v2 500-cycle kit (cat# MS-102-2003). Quality control measures prior to sequencing were conducted at the sequencing center.

Bioinformatic and taxonomic assignments

The pipeline used (https://github.com/ramongallego/Nextera_Dada2) consisted of a wrapper that used cutadapt (v2.1 - Martin, 2016) to remove PCR primer sequences and DADA2 (Callahan et al., 2016 -https://benjjneb.github.io/dada2/tutorial.html), a denoising algorithm that produces amplicon sequence variants (ASVs), avoiding clustering. These ASVs are equivalent to unique sequences and are preferable to the operational taxonomic units (OTUs), representing true biological units at a much higher level of resolution (Callahan et al., 2017). The parameters used for quality filtering were: (1) retain only the first 200 base pairs in each direction before joining until the sequences' quality lowers to 2; (2) eliminate all the sequences with unresolved bases; (3) discard sequences in which forward, and reverse do not overlap by at least 12 base pairs; (4) chimeric sequences were discarded.

The script used to produce a taxonomical ID relies on two thresholds of number of identical bases: the minimum threshold limits all credible matches, the maximum threshold, if cleared, keeps only those matches above it. For this study, the minimum threshold was 85%, and maximum threshold was 95%. For a given query sequence, if they were matches with a similarity above 95%, only those matches were kept. if there were none above that, but some between 95 and 85%, then those were kept. If all matches were below 85%, no matches were kept. Of all matches passing these criteria the LCA was calculated and give that annotation to the query sequence

A local BLAST was performed using bash scripts, to query the ASVs obtained from the quality filtering against a custom database.The parameters used for BLAST were, percent_identity = 85, word_size = 30, evalue = 1e-30, maximum_matches = 50, and culling = 5. The use of MARES made it possible to bring the taxonomical annotation (taxid) of the organism of origin of each reference sequence. Combined with the relationship amongst taxa, it was possible to recover the taxonomical annotation of each BLAST match and assign a taxonomical annotation to the query sequences.

References

Allen GR, Hammer MP. 2018. *Grallenia larsonae*, a new species of sandgoby (Pisces: Gobiidae) from northern Australia. *Journal of the Ocean Science Foundation,* 31, 87–96

Barnhill KA, Bahr KD. 2019. Coral resilience at Malauka’a fringing reef, Kāne’ohe Bay, O’ahu after 18 years. *Journal of Marine Science and Engineering*, *7*(9). https://doi.org/10.3390/jmse7090311

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Callahan BJ, McMurdie PJ, Holmes SP 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal*, *11*(12), 2639–2643. https://doi.org/10.1038/ismej.2017.119

Carpenter K, Niem V. 2001. *FAO species identification guide for fishery purposes. The living marine resources of the Western Central Pacific. Volumes 1-6. 2001. Carpenter, K.E. and V.H. Niem (eds). Rome, FAO*.

Donner SD, Carilli J. 2019. Resilience of Central Pacific reefs subject to frequent heat stress and human disturbance. *Scientific Reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-40150-3

Djurhuus A, Port J, Closek CJ, Yamahara KM, Romero-Maraccini O, Walz KR, Goldsmith DB, Michisaki R, Breitbart M, Boehm AB, Chavez FP. 2017. Evaluation of filtration and DNA extraction methods for environmental DNA biodiversity assessments across multiple trophic levels. *Frontiers in Marine Science*, *4*(OCT), 314. https://doi.org/10.3389/fmars.2017.00314

Greenfield D, Winterbottom R. 2016. A key to the dwarfgoby species (Teleostei: Gobiidae: *Eviota*) described between 1871 and 2016. *Journal of the Ocean Science Foundation*, *24*, 35–90. https://doi.org/10.5281/zenodo.219620

Greenfield D. 2021. Addendum to the 2016 key to the dwarfgobies (Teleostei: Gobiidae: *Eviota*). *Journal of the Ocean Science Foundation*, *38*, 1–12. https://doi.org/10.5281/zenodo.4458248

Kovacic M, Bogorodsky SV. 2017. Two species of *Cabillus* (Perciformes: Gobiidae) and the first record of *Cabillus mocrophthalamus* from the Western Indian Ocean. *Zootaxa*, *3717*(2): 179-194. http://dx.doi.org/10.11646/zootaxa.3717.2.4

Kovacic M, Bogorodsky SV, Mal AO, Alpermann TJ. 2018. Redescription of the genus *Koumansetta* (Teleostei: Gobiidae), with description of a new species from the Red Sea. *Zootaxa* 4459(3): 453-48. https://doi.org/10.11646/zootaxa.4459.3.3

Larson HK, Buckle DJ. 2012. A revision of the goby genus *Gnatholepis* Bleeker (Teleostei, Gobiidae, Gobionellinae), with description of a new species. *Zootaxa 3529*: 1-69. https://doi.org/10.11646/zootaxa.3529.1.1

Longmire J, Malbi M, Baker R. 1997. Use of “lysis buffer” in DNA isolation and its implications for museum collections. *Museum of Texas Tech University*, 1–4.

Neiffer DL, Stamper MA. 2009. Fish sedation, anesthesia, analgesia, and euthanasia: Considerations, methods, and types of drugs. *ILAR Journal*, *50*(4), 343–360. https://doi.org/10.1093/ilar.50.4.343

Randall JE. 2001. Five new Indo-Pacific Gobiid fishes of the genus *Coryphopterus*. *Zoological Studies*, *40*(3): 206-225.

Smallhorn-West PF, Sheehan J, Malimali S, Halafihi T, Bridge TCL, Pressey RL, Jones GP. 2020. Incentivizing co-management for impact: mechanisms driving the successful national expansion of Tonga's Special Management Area program. *Conservation Letters*, *13*(6), 1–10. https://doi.org/10.1111/conl.12742.

Chapter 2

Species delimitation in the *Eviota sigillata* complex, a widely distributed group of cryptobenthic coral reef fishes

In preparation for submission to: *Molecular Biology and Evolution*

Marta C. Gómez-Buckley[1], Carmen Pedraza Marrón[2], Mark V. Erdmann[3,4], and Luke M. Tornabene[1]

[1] School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA USA
[2] Department of Life and Environmental Sciences, University of California, Merced, CA USA
[3] Conservation International Aotearoa, University of Auckland, New Zealand
[4] California Academy of Sciences, San Francisco, CA USA

ABSTRACT

Cryptobenthic reef fishes (CRF) are the smallest vertebrates on coral reefs but represent about 40% of the fish species and about 50% of fish abundance in coral reef ecosystems. Their diversity can be explained by their extremely limited dispersal abilities and short generation times (promoting allopatric speciation) coupled with their ability to partition microhabitats at a very fine scale. Importantly, for some groups of CRF, their small size, cryptic nature, and conserved morphology have resulted in many undetected cryptic species, which may require a genome-wide species delimitation approach to discern how many species are present. One of the most species-rich groups of CRF, the genus *Eviota*, has 132 species described to date, is widely distributed from the Red Sea to Hawaii and French Polynesia, and is known to comprise numerous cryptic species. We focused on the *Eviota sigillata* complex which is represented by two nominal species described by morphological characters, yet preliminary genetic data suggest the presence of multiple cryptic lineages. Here we use molecular data from mitochondrial DNA and genome-wide SNP data generated via double digest restriction site associated sequencing (ddRADseq), in combination with morphological data to infer the number of species in the *E. sigillata* complex. Specifically, we constructed phylogenetic trees and conducted several types of single-locus and multilocus species delimitation analyses and compared these to groupings based on morphology, as well as their geographic distribution. Overall, we recovered evidence for the presence of 9-13 lineages within the *E. sigillata* species complex, with genetic lineages corresponding well with the biogeography of the group. We further confirmed that the original morphological diagnostic characters used for the separation of the two nominal species were not useful for distinguishing each of the nine clades in the complex but may be helpful in diagnosing groups of species. Overall, our study sheds light onto the patterns of speciation within CRF and provides a glimpse of the tremendous hidden diversity that still remains in coral reef fishes.

INTRODUCTION

Tropical coral reefs are home to an estimated one third of all marine fish species (~6000 species; Spalding and Grenfell 1997; Eschmeyer et al. 2010; Brandl et al. 2018), despite covering <1% of the ocean floor. Nevertheless, many coral reef biodiversity surveys tend to disregard a significant portion of the community, failing to account for small and inconspicuous species (Brandl et al.

2018; Brandl et al. 2019; Bessey et al. 2023). Cryptobenthic reef fishes (CRF) are the smallest vertebrates on coral reefs and belong to at least 17 different families of fishes (Brandl et al. 2018). CRF represent about 40% of the fish species present in coral reef ecosystems, and about 50% of fish abundance (Depczynski and Bellwood 2003; Brandl et al. 2018). They are typically less than 5 cm long, and the majority live very close to the coral substrate, often hiding within the crevices of live corals as well as coral rubble (Depczynski and Bellwood 2004; Brandl et al. 2018). Due to their high turnover rates, they transfer substantial energy as a constantly replenishing prey source for larger reef inhabitants (Brandl et al. 2019), making them an integral functional component of coral reef ecosystems. These rapid rates of replenishment, and their ability to partition microhabitats at a very fine scale (due to their small size and association with the benthos), may explain their high diversity and high speciation rates (Winterbottom et al. 2014; Greenfield 2017; Brandl et al. 2019) For example, the genus *Gobiodon* (Harrison 1989) demonstrates extreme niche partitioning, with many species having species-specific associations with host acroporid corals (Munday et al. 1997; Munday 2001). Similar fine-scale niche partitioning is also present in other gobiid genera such as *Bryaninops* and *Pleurosicya* (Depczynski and Bellwood 2004; Herler 2007). In addition to their fine-scale ecological partitioning, many species have short lifespans and high turnover rates (e.g., the goby *Eviota sigillata* has the shortest known vertebrate lifespan at only 59 days: Depczynski and Bellwood 2005). This could translate into high rates of molecular evolution which, when coupled with their limited dispersal abilities, could explain why CRF also show high levels of allopatric speciation and endemism (Tornabene et al. 2015; Dias et al. 2019).

CRF are difficult to collect and identify due to their small size and cryptic nature. They are a notoriously challenging group for finding diagnostic phenotypic characters for taxonomic classification, especially when color photographs are not available, which is the case for many CRF species. Because of this, CRF lag far behind larger reef fishes in rates of species discovery and description. Nevertheless, Brandl et al. (2018) estimated that around the year 2031, the projected number of CRF species described will equal or surpass the number of larger reef fish species. Within the family Gobiidae, the most species-rich family of CRF, the genera *Eviota* (Jenkins 1903) and *Trimma* (Jordan and Seale, 1906; Winterbottom et al. 2020) make up a considerable portion of CRF species in most microhabitat types in the tropical Indo-Pacific region. *Eviota* currently comprises 132 species (Greenfield et al. 2024) and is widely distributed from the Red Sea to Hawaii and French Polynesia (Greenfield 2017). The first species assigned to the genus was in 1871 (Klunzinger 1871), but few taxa were described for the next century. In the 1970s-80s, Lachner and Jewett (Jewett publishing earlier in her career as Karnella) continued specimen collections and established the main morphological features used to delimit species within the genus (Lachner and Karnella 1978; Lachner and Karnella 1980) and described several dozen species. More recently, Greenfield and Winterbottom (2016) published an exhaustive key of the 111 described species at the time of publication, with an addendum to this key raising the number to 123 species (Greenfield 2021). In the last 3 years, a further 9 species of *Eviota* have been described (Greenfield et al. 2024).

Molecular analysis started to be included as a tool for *Eviota* taxonomy in 2013, with the description of four new species from French Polynesia (Tornabene et al. 2013a), using partial mitochondrial cytochrome c oxidase subunit I (COI) sequences and a segment of the nuclear gene protease III (Ptr). Recent phylogenetic analysis of the genus *Eviota* based on COI and Ptr markers has uncovered a multitude of species complexes, including cases in which a single nominal species contains multiple genetically distinct groups with very subtle or non-existent differences in morphology or pigmentation (i.e., cryptic species). These complexes display numerous biogeographic patterns, ranging from sister species that are distributed on either side of well-known, separate biogeographical barriers (*E. sebreei* and *E. punyit*; Tornabene et al. 2016), to complexes that present many geographically restricted species occurring in allopatry, sympatry, or microallopatry (e.g., *E. nigriventris* and *E. bifasciata* complexes; Tornabene et al. 2015). For all these examples, there is usually congruence between morphological and molecular characters, as well as their geographic distributions, with little ambiguity regarding species boundaries. Nevertheless, for complexes with taxa at early stages of speciation where it may be difficult to distinguish between separate species and population structure, other approaches may be required.

To date, most molecular systematic studies on CRF, and all such studies on *Eviota*, have relied on very few numbers of loci (typically one or two genes). Specifically, COI continues to be a preferred marker for this group because it usually presents little within-species variation and high between-species variation (‘the barcode gap’), a desired quality when studying species with a rapid evolutionary clock (Tornabene et al. 2015). Species delimitation analyses that rely on a single (or a few) markers are either tree-based approaches like the Generalized Mixed Yule-Coalescence (GMYC, Pons et al. 2006; Monaghan et al. 2009; Fujisawa and Barraclough 2013), or distance-based species delimitation methods that rely on identifying the ‘barcode gap’ like the Automatic Barcode Gap Discovery (ABGD, Puillandre et al. 2012), and the more recent

Assemble Species by Automatic Partitioning (ASAP, Puillandre et al. 2021). Nevertheless, in some cases, the ‘barcode gap’ metric is not clear-cut. Factors like incomplete lineage sorting, hybridization, or introgression may lead to overlap in genetic distances between individuals of the same species and those of different species, making it hard to distinguish between intraspecific and interspecific variation and thus delimit species. This ‘blurry boundary’ has been referred to as the ‘anomaly zone’ (Degnan and Rosenberg 2006). For some recently diverged or closely related groups evolving in an ‘anomaly zone’ it may be necessary to incorporate methods that rely on genome-wide datasets to resolve species delimitation questions, as well as using more sophisticated models of species delimitation, such as the Multispecies Coalescent model (MSC, Rannala et al. 2020).

The development of techniques like Restriction-site Associated DNA sequencing (RADseq) has made it feasible to recover thousands of single nucleotide polymorphism (SNPs) from loci throughout the entire genome at a relatively low cost (Miller et al. 2007; Hohenlohe et al. 2010; Andrews et al. 2016; Meek and Larson 2019). Despite the argument by some (Lowry et al. 2017) that RADseq subsampling of the whole genome may miss informative regions for species delimitation purposes, others argue that this large random sample of the genome should include coding, non-coding, and regulatory regions that would inform of important evolutionary changes for the studied taxa (Catchen et al. 2017). Other important advantages of using RADseq are that it can be used when a reference genome is not available, which is the case for most CRF, and that it is statistically robust for identifying fine genetic structure (Piñeros et al. 2022). Many methods sampling multiple loci for delimiting species utilize the MSC model, which is also widely used in phylogenomic studies and for estimating population structure (Liu et al. 2019).

The MSC model originated from Kingman’s (1982) coalescent theory and is used to build phylogenetic hypotheses by applying a likelihood function equation to determine the probability of a coalescent history within a lineage. The MSC accounts for incomplete lineage sorting (individual gene lineages may not have had enough time to fully “coalesce”) by recognizing that due to the stochasticity of molecular coalescent events, gene trees may not match species trees (Degnan and Rosenberg 2006). A computational efficient approach is to jointly estimate the species tree from an entire multi-locus dataset without inferring gene trees, using methods like SNAPP (Bryant et al. 2012). Following species-tree inference based on genome-wide SNP data, species delimitation can be carried out using Bayes Factor Delimitation (BFD*, Leaché et al. 2014; Leaché and Ogilvie 2016; Leaché and Bouckaert 2018). BFD* utilizes SNAPP to compute marginal likelihoods for various hierarchical configurations of individuals grouped into predetermined populations (i.e., species assignments into different numbers of taxa). We then can compare the marginal likelihoods of the different species assignments and the number of taxa in each model through the assessment of Bayes Factors (Kass and Raftery, 1995). The BDF* method is widely used and has proven effective in delimitations cases for recently diverged groups. For the silverside species complex in Mexico, Piñeros et al. (2022) ascertained the presence of three lineages that had previously been over-split using morphological and mtDNA data.

One group of CRF that is ideal for applying MSC species-delimitation approaches using genome-wide data is the *Eviota sigillata* complex. It includes two nominal species, *E. sigillata* (Jewett and Lachner 1983), described from the Seychelles, and *E. shimadai* (Greenfield and Randall 2010), described from the Ryukyu Islands of Japan. *E. sigillata* is associated with a variety of microhabitats from sand to coral rubble to branching corals (Tornabene et al. 2013b) and has been found throughout the entire Indo-Pacific, ranging from the Seychelles (type locality) in the west, north to the Ogasawara Islands (Japan), east to Micronesia and Melanesia to Wallis and Futuna, and south to Western Australia and the Great Barrier Reef (Froese and Pauly 2019). Nevertheless, preliminary COI data show high levels of genetic variation within the species complex, with a mixture of patterns that could indicate multiple distinct and undescribed species (highly likely, given the wide distribution of the complex), or that may simply indicate population structure within the two nominal species, or both. The two nominal species are most notably distinguished from each other by the different cephalic sensory-canal pores distribution and numbers, and their patterning of red banding across the body. Nevertheless, these morphological characters are frequently not obvious, are often variable within a single geographic location, and are nearly impossible to detect in juvenile specimens <7 mm long, making morphology-based diagnoses extremely challenging. Clarifying the taxonomy of this complex thus requires an integrative approach that incorporates genome-wide data and MSC analyses as additional line of evidence.

The aim of this study was twofold: (1) assess the number of species in the *Eviota sigillata* complex, a recently diverged group that spans the entire Indo-west Pacific.
For this objective, we ascribed to the genealogical species concept or reciprocal monophyly (Baum and Shaw 1997; De Queiroz 2007); and (2) assess the utility of commonly used, putatively diagnostic phenotypic characters to identify the different species delimited in Objective 1. In addition, we also discuss the biogeographic and speciation patterns of *E. sigillata* complex in the context of other known *Eviota* complexes.

METHODS

Sample Collection

A total of 123 specimens in the *E. sigillata* complex were collected from locations across the Indo-Pacific region (Figure 1; see Table S1 for specific location of 102 samples passing the bioinformatic filtering). Specimens were determined to belong to the *E. sigillata* complex if they keyed out to *E. sigillata*, *E. shimadai,* or an ambiguous combination of the two using the key from Greenfield and Winterbottom (2016) or very closely resembled live photographs of these two species from the taxonomic literature. All specimens were captured after being anesthetized using 5% quinaldine sulphate dissolved in seawater or a solution of 5% clove oil dissolved in ethanol dispersed from a squirt bottle (see Tornabene et al., 2013). Specimens were photographed in the field and then preserved in 95% ethanol. We also used two tissue samples of *E. sigillata* from two specimens from the South African Institute for Aquatic Biodiversity (SAIAB) that were available for genomic analysis and that were collected from the type locality of the described *E. sigillata* in the Seychelles (SAIAB #77969-406 and #77969-407). We also extracted DNA from a tissue sample of an *E. shimadai* specimen collected from the type locality and donated by the National Museum of Nature and Science in Tokyo.

Molecular Protocols

DNA was extracted from pectoral fin clips using Qiagen DNeasy Blood and tissue kits (Qiagen, Inc.) following the manufacturer’s protocols. A 53-sample subset from the 123 total samples was selected for initial sequencing for a segment of the mitochondrial gene cytochrome-c oxidase subunit I (COI), using the primers GobyL6468 and GobyH7696 (Thacker 2003) and FishF1 and

FishR1for a smaller subset of the samples (Ward et al. 2005). PCR protocols were the same as those used by Tornabene et al. (2016). Sequences were aligned in Geneious v.6.0.6 (Biomatters; www.geneious.com). The final alignment consisted of 1,142 bp of COI, with some samples having a shorter ~650 bp fragment (Table 1).

To capture SNP data, we used ddRADseq. Including two enzymes to ‘double digest’ the DNA provides a more precise selection of fragments throughout the genome that are close to the target size (Peterson et al. 2012). It also allows for reproducibility of results if additional ddRADseq data from the studied group is added later to the existing dataset. DNA extractions were sent to the NGS labs at the University of Wisconsin for sequencing and library preparation. Prior to sequencing, an optimization process conducted at the same lab facilities was used to select the restriction enzymes that would work best for our set of samples. For the enzyme optimization, we selected eight representative samples from across the genus *Eviota*. We sent the DNA extractions from these specimens to create ddRADseq libraries, employing various combinations of enzymes including ApeKI, PstI/MspI, PstI/BfaI, NsiI/MspI, and NsiI/BfaI. Subsequently, the libraries were analyzed using an Agilent Tapestation 4200 to assess their fragment sizes and profiles. The combination of the NsiI and MspI enzymes was ultimately chosen due to its favorable characteristics, including a smooth profile and a fragment size >300 bp. Libraries were sequenced on Illumina NovaSeq6000 using paired-end 150 bp reads

ddRADseq libraries were prepared following the methodology outlined in Elshire et al. (2011), with only minor adjustments. To summarize, 100 ng of DNA underwent digestion using NsiI and MspI enzymes from New England Biolabs in Ipswich, MA. Following digestion, barcoded adapters suitable for Illumina sequencing were incorporated through ligation using T4 ligase from New England Biolabs. Subsequently, 123 samples with adapter-ligated DNA were combined and amplified to generate libraries suitable for sequencing, with the removal of adapter dimers achieved through SPRI bead purification. The final libraries' quality and quantity were evaluated using the Agilent Tapestation from Agilent Technologies in Santa Clara, CA, and the Qubit® dsDNA HS Assay Kit from Life Technologies in Grand Island, NY, respectively.

To assemble the ddRADseq raw data, we used the Stacks 2.4.1 pipeline (Catchen et al. 2013) on the University of Washington's High Performance Computer Cluster. First, the samples were demultiplexed using the *process_radtags* program for the paired-end reads, including the -c (*clean data, remove any read with an uncalled base*), -q (*discard reads with low quality scores*), and -r (*rescue barcodes and Rad-Tag cut sites*) flags. We used the R packages RADstackshelpR 0.1.0 (https://github.com/DevonDeRaad/RADstackshelpR) to calculate summary statistics from Stacks output vcf files, and to determine the optimal parameters for assembling putative RAD loci de novo (Mastretta et al. 2015, Paris et al. 2017). These assembly parameters selected were '-m=4' (minimum number of reads required to form a stack), '-M=2' (number of mismatches allowed between stacks within individuals), and '-n=3' (number of mismatches allowed between loc catalog building). To reduce computational burden and as instructed in DeRaad's RADstackshelpR, we used a subsample of representative individuals from the 10 samples to run this parameter optimization pipeline. With the resulting optimal parameters, we ran the *denovo_map.pl* program for the 123 samples. Each of the analysis methods required specific levels of filtering and data curation (Table 1). For instance, the --write-single-snp flag was run for all the species delimitation and the population structure analysis in order to select only the first SNP at each locus, whereas we used all variant sites for phylogenetic analysis of ddRAD data. The R package SNPfiltR 1.0.1 (DeRaad 2021) aided the visualization of the proportion of missing data per sample, and in setting an acceptable completeness cutoff per sample that would retain enough SNPs to still allow for robust inferences from each of the performed analyses.

Phylogenetic analyses

From the COI sequence data, we generated a maximum likelihood phylogenetic tree using the GTR+G RAxML-NG v1.1 0 software (Kozlov et al. 2019) on CIPRES Science Gateway online high-performance computer resource (Miller et al. 2010). Support for nodes in the ML tree was assessed with 100 bootstrap replicates. The tree was visualized using FigTree v1.4.4 (Rambaut 2012). Downstream COI species-delimitation analyses require a time-calibrated phylogeny. We generated an ultrametric COI tree with the program BEAST 2 (Bouckaert et al. 2019). We used the 'GTR+G' site model and the 'Optimized Relaxed Clock' model to estimate the mean clock rate during the analysis. The calibration prior on the age of the root of the tree was set to 1.5 million years using a normal distribution with a mean of 1.5 and standard deviation of 0.5, based on the stem age of the *E. sigillata* complex estimated by Tornabene et al. (2016).

For the ddRADseq data, we selected a filtered dataset containing 104 individuals (101 sigillata complex and three non-sigillata outgroup samples) with a combined 4,591 loci (RAD sequence) and 660,013 bp (Table 1). This dataset resulted after setting an initial filter specifying a minimum of 20% of individuals required to process a locus. The program SNPfiltR would be run for the remaining data until only 40% of missing data by samples and SNPs was accomplished for the remaining 104 taxa and 4,591 loci. With this dataset, we generated a maximum likelihood tree (from here on referred to as ddRADseq tree), using RAxML-NG v1.1 (Kozlov 2018) under the GTR+G model. We used the CIPRES (Miller et al. 2010) supercomputer portal to run the analysis. Support for nodes was assessed with 100 bootstrap replicates. The trees were visualized using FigTree v1.4.4 (Rambaut 2012).

Single-locus species delimitation analyses

For the COI data, we used the tree-based Generalized Mixed Yule Coalescent (GMYC, Pons 2006) and the distance-based Assemble Species by Automatic Partitioning (ASAP, Puillandre et al. 2020) methods to explore if species delimitation results from the single locus dataset would agree with the results from the species delimitation analyses with the genome-wide dataset. The GMYC is a likelihood method that delimits species by fitting within and between-species branching models to reconstruct gene trees. Using a time-calibrated single-locus tree, GMYC fits a model containing a single species and compares this to the multiple species models using a likelihood ratio (LR) test. For each model, GMYC reports the estimated number of clusters (i.e. species containing more than one terminal taxon in the dataset) and number of entities (i.e. the estimated total number of species including those made up of a single taxon), along with their respective confidence intervals. To perform the GMYC, we used the ultrametric COI tree with the function 'gmyc' in the R package 'splits' (Ezard et al. 2009; Fujisawa and Barraclough 2013). The ASAP method (Pulliandre et al. 2021) is based on pairwise genetic distances and uses an ascending hierarchical clustering algorithm that merges sequences into '*groups*'. After each merging step, the assignment of all sequences into '*groups*' is named a '*partition*'. Through a recursive splitting process an asap-score (combination of probability and barcode-gap width metrics) is generated and used to rank the partitions. ASAP does not use a phylogenetic tree and requires a FASTA file of all sample sequences as input into their graphical web interface (https://bioinfo.mnhn.fr/abi/public/asap/#). We selected the '*simple distance*' (p-distance) as the method to calculate pairwise distance. The lowest asap-generated score is the one that determines the best species model (Puillandre et al. 2021).

Coalescent species delimitation

For our genome wide SNP data, we employed Bayes Factor Delimitation with BFD* program (Leaché et al. 2014) to assess and rank various species delimitation models within the context of a multispecies coalescent framework. In essence, BFD* entails conducting SNAPP analyses with different species counts and individual species assignments, estimating the marginal likelihood for each model, and then comparing Bayes factors based on runs that rank each model tested. This approach provides a quantitative way to delimit species across multiple unlinked loci and is in alignment with the genealogical species concept (Baum and Shaw 1995) which assumes no gene flow as criteria for defining species.

The BFD* analysis was conducted in SNAPP v. 1.3.0 (part of BEAST2, Bouckaert et al. 2019) on a subset of individuals from Tonga, Fiji, and Samoa. A more stringent filtering was necessary for this dataset, since the analysis was focused on a smaller set of data that clustered together in a closely related group. An initial filter when running *populations* was set at -R 0.5, meaning that a minimum of 50% of individuals across populations was required to process a locus. Further filtering was accomplished by using VCFtools (Danecek et al. 2011) with a minor allele frequency of – maf 0.05, meaning that the minor allele is present in at least 5% of the individuals. The program SNPfiltR would be run for the remaining data, until there were only

18% of missing data by sample and by SNP, which left 29 individuals (two of these were from the nearest outgroup to the Tonga, Fiji, and Samoa clade) and 2,476 loci (Table 1). We further filtered the dataset using the program TASSEL (Bradbury et al. 2007) to remove taxa and loci, such that only loci present in at least one individual per putative species in our most species-rich model (Tonga, Fiji, Samoa, and outgroup) were included (see details in supplementary section). The final filtered dataset consisted of 29 individuals and 1,640 loci. We compared Bayes Factors for models with two, three, and four (three + outgroup) species. We followed the guidelines specified in Leaché et al. (2014) for parameter setting. For ***Model Parameters,*** we used *Mutation Rates* for U (Instantaneous rate of mutation from the 0 allele to the 1 allele) and V (Instantaneous rate of mutation from the 1 allele to the 0 allele) = 1.0 (as recommended in Leaché and Bouckaert 2018), with a *Coalescence Rate* = 0.01 (checking the *Sample* and *Use Log Likelihood Correction* boxes). For ***Priors***, we used the *Gamma* model, with *Alpha* and *Beta* values = 2.0 and 100.0 respectively. For the ***MCMC***, we used a Chain Length = 10,000,000. To determine the marginal likelihood values for each species model, we ran multiple SNAPP analyses with the software BEAST2 (Bouckaert et al. 2019) in the CIPRES supercomputing portal (Miller et al. 2010). SNAPP requires users to specify *a priori* assignments of taxa as proxy for potential species, thus we used geography to divide our taxa. The three models we ran in BEAST2 were as follows: Run A = 4 species (Samoa, Fiji, Tonga, and outgroup), Run B = 3 species (Fiji+Tonga, Samoa, and outgroup), and Run C = 2 species (Samoa+ Fiji+Tonga, and outgroup). Visualization of results and confirmation of mixing and convergence was done through the program Tracer 1.7.1 (Rambaut et al. 2018). The final step in this process was to compare the marginal likelihood estimates for each species model using the Bayes Factors (BF, Kass and Raftery 1995), calculated with the equation: $BF = 2 \times (MLE_1 - MLE_0)$, where MLE_0 is the marginal likelihood of model 0 and MLE_1 is the marginal likelihood of model 1. The scale for the BF values identifies what models should be consider as follows: $0 < BF < 2$ not supported, $2 < BF < 6$ positive evidence, $6 < BF < 10$ strong support, and $BF > 10$ decisive.

Population structure analysis

To investigate the fine-scale genetic structure among a closely-related complex of lineages that could potentially represent new species, or perhaps early-diverging lineages in the 'anomaly zone' (i.e. clades of samples from Tonga, Fiji, and Samoa – see above), we took a population genetic approach. For this analysis, we used the same taxa as for the SNAPP/BFD* analysis. The only difference is that the outgroup taxa were removed to avoid introducing confounding genetic diversity not representative of the studied group. The final filtered subset consisted of 27 individuals from the three above-mentioned populations genotyped at 4,235 loci (Table 1). For this subset, we used the program Structure 2.3.4 (Pritchard et al. 2000). This program uses a Bayesian clustering algorithm to assign individuals to genetic clusters based on their genotype data. This is done by examining the rate of change in likelihood values when different numbers of clusters (k) are present without prior knowledge of cluster membership of individual samples. The goal is to identify the k value that maximizes the rate of change of the likelihood value (Δk, Evanno et al. 2005). The peak of the Δk on the resulting graph (of k values vs. Δk) indicates this most likely number of genetic clusters or populations in the dataset. An advantage of using Structure for our data is that SNPs have lower variability than multiple allele loci genomic data (i.e., microsatellite data), thus requiring much smaller sample sizes per putative population (Porras-Hurtado et al. 2013). For the number of genetic clusters (k), we tested a range of values from 1-6 to identify the optimal number of clusters that explained the genetic variation of our dataset. This range was selected based on the phylogenetic analyses showing three distinct clades each from Tonga, Fiji, and Samoa. For the burn-in and MCMC replicates, we conducted 10 runs for each k value with a burn-in period of 12,500 iterations followed by 50,000 MCMC replicates to ensure convergence and consistency of results. We assumed the “admixture” model, which allows individuals to have mixed ancestry from multiple genetic clusters. We employed the “independent allele frequencies” model, which assumes that the allele frequencies within clusters are uncorrelated. After running Structure, the program’s output files were processed using Structure Harvester web server (Earl and vonHoldt 2012).

Morphology

The morphological dataset included 59 specimens with a range in standard length (SL) of 9.17-17.53 mm (average = 12.29 mm). To evaluate patterns of phenotypic variation, we compiled 10 linear morphometric and 6 meristic traits (Figure 5). These linear measurements and counts were chosen because they were used in taxonomic studies for species descriptions (Lachner and Karnella 1978, 1980) and for identification keys (Greenfield and Winterbottom 2016) for the genus *Eviota*. We used a Zeiss Discovery V20 SteREO microscope, controlled with the ZEN 2.3 Pro software to have a photographic record of morphological features and to aid in measurements and meristic counts. The complete morphology dataset can be found in the Supplementary Data 1.

We used Principal Component Analysis to examine morphological variation among our specimens using R v4.0.3 and packages FactoMineR (Lê et al. 2008), factoextra (Kassambara and Mund 2017), dplyr (Yarberry 2021), tidyr (Wickham and Wickham 2017), ggplot2

(Wickham 2016), and ggrepel (Slowikowski 2024). We selected groups based on the genome-wide data (ddRAD) which provided finer genetic structure. The variables selected for morphometric PCA plots were head length, distance from mouth to first dorsal fin spine, distance from mouth to second dorsal fin spine, distance from mouth to anal fin spine, body depth, caudal peduncle length, caudal peduncle depth, eye diameter, snout length, and upper jaw length (Figure 5). Standard length (SL) was recorded but not included in the PCA because it is known to be a confounding variable when analyzing fish morphometric relationships. We used raw measurements and log-transformed the data before doing the PCA, as recommended by Orr and Hawkins (2007).

For comparisons amongst groups from the ddRADseq tree, we focused on sensory pore patterns, and two meristic variables, second dorsal and anal fin ray counts. These are all diagnostic characters that separate the two nominal species in the complex, *E. sigillata* and *E. shimadai*, and are routinely used for diagnosing other species of *Eviota* as well.

RESULTS

Phylogenetic analysis

The ultrametric COI tree (Figure 2B) recovered two main clades. One clade several strongly supported genetic lineages with individuals from Tonga, Maldives, and the Coral Triangle. The other strongly supported clade, included three subclades, two from the Coral Triangle, and one with samples Tonga and Fiji, and Samoa.

The ddRADseq tree generated based on 104 taxa, 4,591 loci, and 85,608 variant sites resolved at least nine distinct clades with 100% bootstrap support values (Figure 3). All nine clades are deeply divergent from one another, with little variation within clades, and they each belong to a relatively restricted geographic region (Figure 1). Clade I contains specimens exclusively from Tonga. Clades II and III comprise specimens from the Maldives and the Seychelles, and specimens from Indonesia, respectively. Clade IV contains individuals from Indonesia, Timor-Leste, and Brunei. Clade V contains individuals from Indonesia, Papua New Guinea, and the Solomon Islands. Clade VI contains specimens from Indonesia and Solomon Islands, Clade VII from the Samoan Archipelago (Western Samoa and American Samoa), and Clade VIII from Fiji. Clade IX contains specimens exclusively from Tonga and includes numerous nested clusters that are not well supported. The internal branch lengths between Clades VII, VIII, and IX are considerably shallower than those between other clades in the tree, and thus were subjected to downstream species delimitation analyses (see below).

When comparing the ddRADseq and the ultrametric COI trees, many of the same groups are recovered in both with strong support (Figure S1). The main differences are that several of the between-clade relationships are poorly supported in the ultrametric COI tree and differed from that of the ddRADseq tree (Figure S1). Also, the ultrametric COI tree has a mix of different branch lengths, making it difficult to clearly visualize the number of distinct clades (e.g. some lineages appear to be in an ‘anomaly zone’ between species and populations), whereas the ddRADseq tree shows as many as 9 deeply divided groups, except for Clades VII, VIII and IX.

Single-locus species delimitation

The GMYC analysis of the COI tree identified a total of 13 'entities' representing 13 putative species (Figure 2B), with a 95% confidence interval spanning 13 to 18 species. The LR test yielded a statistically significant p-value of 0.004089, indicating strong support for a model with >1 species. Some of the species were represented by single specimens, whereas others were composed of clades with multiple individuals that had varying levels of support from the species delimitation analysis. The best estimate for a temporal cutoff where the branching pattern transitions from a Yule or speciation process to that of intraspecific variation was at ~180k years before present (shown by the red line in Figure 2A), but the confidence interval for that cutoff spanned ~100-180k years. This indicates a moderate 'anomaly zone', or area of uncertainty regarding the intra- vs interspecific branching patterns. ASAP ranked a 13 species model the best supported, as determined by the lowest ASAP score value of 2.00 (Table 2, Figure 2C). Similarly to the GYMC results, many 'species' in the ASAP analysis were represented by single specimens. In general, most species recognized in each analysis were geographically restricted to single islands or island chains. Several of the putative species in the ASAP analysis were phylogenetically nested within other putative species (Figure 2).

Structure analysis and multispecies coalescent species delimitation of genomic data

We performed a Structure analysis of the SNP data in the monophyletic lineage containing specimens from Tonga, Fiji, and Samoa, as the branching patterns among the different subclades of this group were shallower than those found between the other major clades on the ddRADseq tree but might still represent recently diverged species (Table S2). Fst values for lineage comparisons ranged from 0.056-0.255 (Table S2). Among the structure plots for the k=1-6 clusters, Structure Harvester selected a k=2 as the best-fitting model (Figure 4). This model assigned individuals from the Samoan archipelago as a separate cluster, with the Fiji and Tonga individuals being assigned the same cluster (Figure 4). On the other hand, results from the SNAPP+BFD* analysis ranked first (BF support value of 4,371.79) the species model with Samoa, Tonga, and Fiji as separate groups (Table 3).

Morphology

We examined six discrete morphological characters that are known to be diagnostic for species of *Eviota* (Figure 5). After examining the variation of these characters within and among the clades in the ddRADseq tree, the only characters that showed promise as diagnostic characters for this complex (e.g. variable between clades, but mostly conserved within clades) were pore patterns and counts of soft rays in the second dorsal and anal fins. Additional meristic counts that we initially considered informative included pectoral fin ray, caudal fin ray, and branched caudal fin ray counts, as well as other discrete characters such as pectoral ray bifurcation and pelvic fin segmentations were also recorded. However, after examining the variation of these characters among the ddRADseq tree groups, we found no consistency within clades or differences between clades, so these traits were not considered in further analyses.

The only two head pore patterns observed here corresponded to the two patterns present in the two described species (Figure 5). The type series of *E. sigillata* (Jewett & Lachner 1983) has two distinguishable interorbital pores, while the other pattern seen in *E. shimadai* (Greenfield and Randall 2010) has a single enlarged interorbital pore. Of all the 79 specimens examined for pore pattern, 50% had pattern A, 30% had pattern B, and 20% were either damaged or were too small for the canals and pores to be developed. The other two diagnostic meristic characters were the second dorsal fin and anal fin ray counts. Fin ray counts ranged from 8-9 in the second dorsal fin and 7-8 in the anal fin, which agrees with ranges reported in the taxonomic literature (Jewett and Lachner 1983; Greenfield and Randall 2010; Greenfield and Winterbottom 2016). Except for individuals from Clades I and II, where both pore patterns were present within those lineages, all other lineages consistently presented either one or two interorbital pores. Within Clade II, the specimens from the Seychelles (type locality for *E. sigillata*) had two pores (SAIB406 and SAIB407; Table 4) as the described *E. sigillata*, while specimens from the Maldives presented one pore (ESH19 and ESH17, with ESH20 presenting non developed pores) as the described *E. shimadai* from the Ryukyu Islands in Japan. Like pore patterns, counts of anal fin and dorsal fin rays were generally consistent within lineages but varied across lineages (Table 4, Figure 3B). Morphometric data among the species belonging to the nine lineages from the ddRADseq tree overlapped heavily, as seen in the PCA plot for the principal components 1 and 2 (Figure 6) and were of little use in differentiating putative species.

DISCUSSION

Cryptobenthic reef fishes possess many ecological traits that make them an ideal group for studying fine-scale patterns of speciation in the sea, including small home ranges, reduced dispersal capabilities, and extremely short generation times (Depczynski and Bellwood 2005; Tornabene et al. 2013b, 2015, 2016; Brandl et al. 2018). The aim of this study was to examine the species boundaries of the *Eviota sigillata* complex, a widespread group of recently divergent cryptobenthic reef fishes (Greenfield 2017), using molecular and morphological data, and implementing species-delimitation approaches. Specifically, we wanted to estimate the number of species in the *Eviota sigillata* complex, and evaluate the utility of commonly used, putatively diagnostic phenotypic characters to identify the different species. Our analyses revealed that like many other groups of *Eviota* and other cryptobenthic fishes, species richness in the *E. sigillata* complex was previously underestimated.

How many species are in the Eviota sigillata complex?

Prior species delimitation studies for complexes within the *Eviota* genus were based on just morphology or a combination of morphology and a few genetic markers (Greenfield et al. 2014, 2017, 2018, 2019; Tornabene et al. 2015, 2016, 2021 Erdmann et al. 2023; Hanahara 2023). With the addition of genome-wide data, our results show the presence of at least nine species within the *E. sigillata* complex (Figure 3). Support for the number of species found with the six different tools employed here (GMYC, ASAP, molecular phylogenies from COI and ddRADseq, Structure, BFD, and morphology) ranged from five, based on groups with discrete combinations of diagnostic morphological features, to as many as 13 species based on the two single-loci mtDNA methods, GMYC and ASAP (Figure 3). Discrepancies in inferred species boundaries are to be expected in species delimitation research (Carsten et al. 2013; Luo et al. 2018). Even though our approach also inferred a different number of species for each method used (Figure 8), perhaps the most unambiguous pattern found here was that of the 9 strongly supported clades delimited from the ddRAD-generated tree (Figure 3, and S1). Additionally, each of the nine clades in the ddRADseq tree corresponded to specific geographic regions. The ddRAD RAxML tree thus provided a foundation for us to evaluate the other species delimitation approaches.

We used the GYMC method (Pons et al. 2006) to analyze our mitochondrial data, which has been widely implemented and tested since its development (Monaghan et al. 2009; Fujiwasa and

Barraclought 2013). Here the GMYC method yielded 13 “entities” or species. Nevertheless, some ambiguities were observed on the COI tree used for the GMYC analysis, which raise uncertainty about the results of the analysis and may suggest the presence of over-splitting. One of them was related to patterns of branch lengths, which ultimately determine species delimitation in the GMYC analysis. Only one clade, Clade I (Tonga), includes a distinct long branch at the crown of the group with all members of the clade showing little-to-no genetic variation between them; this pattern shows a distinct separation between a speciation or yule process (the long branch), and population level variation (short branches) is typically interpreted as being the evidence for a lineage being ‘good’ species. Yet for most of the other clades, this pattern was absent, and it was unclear what genetic variation could be attributed to inter- vs intraspecific variation. Another source of ambiguity was related to the presence of lineages that were identified as being distinct species despite only being represented by one or two individuals (Figure 2). For instance, sample ES3 from Indonesia was split into a separate species from its sister clade (Figure 2), all of which were sampled from other Indonesia and Brunei locations. The same could be said for sample FB7 from the Solomons, which GMYC considered a separate species from the sister species also from the Coral Triangle. Similarly, two samples from Tonga, VMP087 and VMP111, are clustered together as a single species by GMYC and appear as sister clade to the remainder of Tonga and Fiji samples. In all these instances, the single or pairs of individuals identified as species in the GYMC analysis were deeply nested in other clades in the ddRADseq phylogeny, casting doubt on their validity. Indeed, empirical as well as simulation studies suggest that the GMYC method may overestimate species diversity (Carsten et al. 2013; Miralles and Vences 2013; Dellicour and Flot 2015), and 13 species may be indeed over-splitting species in the *E. sigillata* complex.

The ASAP method (Puillandre et al. 2020) produced similar results to the ones from GMYC in terms of the number of delimited species. While it did not identify the ES3 Indonesian sample as a separate species, it instead identified a different singleton, ESH9 from Indonesia, as a separate species. Once again, this sample was deeply nested in a larger clade of Coral Triangle samples in the ddRADseq phylogeny, casting doubt on the ASAP results. Both the GMYC and the ASAP methods are recognized as tools that can be used as the first line of evidence for delimiting species but should be corroborated/supplemented with other lines of evidence like morphology, ecology, or biogeography (Fujisawa and Barraclough 2013; Blair and Bryson 2017). If taxonomic decisions solely rely on these methods, the approach should be conservative when deciding on the number of species (Carsten et al. 2013).

In contrast to the COI tree, the ddRADseq tree showed very clear clades with long branch lengths separating them into well-resolved lineages. All the singletons or pairs of taxa identified as distinct species in the COI analyses were deeply embedded within clades in the ddRADseq tree. The Samoa, Fiji, and Tonga subgroups were the only clades with shorter branch lengths in the ddRADseq tree, hinting at recent divergence that might represent population structure or the early stages of speciation. Thus, we investigated these subgroups using population genomic and multispecies coalescent species delimitation analyses (BFD). The population genomics analyses we performed for these subgroups with Structure separated them into a Samoa population and a Fiji-Tonga population. The F_{st} values among the three populations showed the lowest level of genetic differentiation for Fiji-Tonga (0.0559), when compared to higher values between Samoa-Fiji (0.255) and Samoa-Tonga (0.111). Nevertheless, it has been documented that Structure tends to over select k=2 values (Janes et al. 2017), and it is possible that hierarchical relationships may exist among these three groups which may underestimate their number (Kalinowski 2011), giving us more confidence in the results from the BFD method, which splits the subgroups into three different populations based on the three locations. Our confidence is based not only on the bias of Structure to select two populations but it is also because it is common practice to use clustering approaches like Structure as a guidance tool for exploratory purposes that will be followed by methods that will complement results and that are specifically designed for species delimitation (Leachè and Fujita 2010; Carsten et al. 2013), like the SNAPP/BDF* method. Nevertheless, it is reasonable to think that our three groups may be incipient species in their early stages of speciation. On the other hand, the disagreement between the two genomic analyses in the number of groups present may indicate that we may be observing population-level variation, species-level structure, or a combination of both due to the speciation process being not an instantaneous but an extended process (Sukumaran and Knowles 2017).

The morphological analysis identified five morphotypes based on combinations of head pore pattern, dorsal fin rays, and anal fin rays. The ranges of fin-ray counts we observed across the entire complex were originally described as intraspecific phenotypic plasticity in the original descriptions of *E. sigillata* and *E. shimadai*. However, our results indicate that rather than being completely plastic, they show some phylogenetic signal and correspond to distinct clades, or groups of clades, ultimately suggesting that they may indeed be taxonomically informative. While not every strongly supported clade in our ddRADseq tree had its own unique combination of morphological traits, it is also possible that phenotypic traits other than the ones used in our study could separate the more closely related groups. Subtle features like the presence of tricuspid teeth, variations in eye coloration of live specimens, or subtle differences in live/fresh body coloration could eventually prove to be informative, as they have in other dwarfgobies (Greenfield et al. 2018; Tornabene et al. 2021; Erdmann et al. 2023). Underwater photos are unfortunately not available for all the lineages here, and preliminary examination of coloration has been inconclusive. Nevertheless, the agreement between the topology of the nine clades shown in the ddRadseq tree and the distribution of some phenotypic characters suggests that the genetic differences here may indeed reflect species boundaries.

Taxonomic status of the E. sigillata complex

While we currently conservatively recognize nine species in the *Eviota sigillata* complex (Figure 3), discerning which of them, if any, correspond to *E. sigillata* and *E. shimadai* is challenging. The two nominal species in the complex were described using only morphological traits. All the holotypes and paratypes were preserved in formalin, so no genetic information is available. The main differences described for the two species are the pore pattern and dorsal/anal ray counts. The only five specimens that have subsequently been collected from near the *E. sigillata* type locality in the Indian Ocean, and that we successfully ddRAD sequenced, clustered in the well-supported Clade II. Specimens from this clade have fin ray counts (9/8) that aligned with the types of *E. sigillata* but had a mixture of pore patterns. Thus, if this clade is indeed *E. sigillata*, then there is even more confusion regarding the characters that differentiate it from *E. shimadai*. The DNA extracted from tissue samples of two *E. shimadai* specimens from the type locality in Japan, and from a tissue sample that the Tokyo Museum provided from one of the *E. shimadai* paratypes, were too degraded and unfortunately could not be successfully sequenced. This, coupled with the variation in dorsal and anal ray counts in the type series of *E. shimadai*, prevented us from determining which lineage in our study corresponds to that species. Thus, conservatively we have at least seven distinct new species if the two nominal species are included in our study, or nine new species if they are not present in our dataset. Further sampling in the type localities is necessary to help clarify the identification of *E. shimadai* and *E. sigillata*.

Biogeographic patterns and speciation

Among the several *Eviota* complexes that have been studied to date, there are examples of complexes with relatively restricted biogeographic ranges (e.g., complexes entirely within the Coral Triangle) and others that are broadly distributed (e.g., complexes that span nearly the entire Indo-Pacific Region), with each of these complexes displaying a mix of allopatric and sympatric species (Tornabene et al. 2013a; Tornabene et al. 2015; Erdmann et al. 2023). Our results demonstrate that, like other complexes of dwarfgobies, the *Eviota sigillata* group shows a mosaic of geographic patterns including both allopatry and sympatry. In the present study, we found two distinct genetic lineages in Tonga (Clades I and IX). These are in fact beyond sympatric; at our sampling site in Vava'u, Tonga, they are syntopic, occurring in the same coral rubble habitats and collected in the same 1 m^2 sampling event. Other examples of dwarfgobies from the same species complex sometimes occurring in sympatry include *E. punyit* and *E. sebreei*, which are sister species that segregate into separate coral colonies in their area of geographic overlap (Tornabene et al. 2016). It is expected that recently diverged sister species would experience ecological segregation in areas of geographic overlap, maintaining prezygotic reproductive isolation, but in the case here, the two syntopic Tonga clades are not sister species; Clade IX is more closely related to specimens more than 10,000 kilometers away from the Seychelles and Maldives (Clade II) than to their Tongan cohabitants (clade I). This is similar to the *E. zebrina* complex, where *E. gunawanae* and *E. longirostris* occur in sympatry in Fakfak, Indonesia, but the two are not sister species (Tornabene et al. 2021). In the example here, Clade I specimens in Tonga are less abundant than Clade IX specimens (main author's personal observation). Furthermore, these two clades have different distinct coloration patterns (Figure S2), which may perpetuate reproductive isolation via mate selection. These changes in coloration patterns and behaviors may trigger species-specific mating rituals, which have been observed in other *Eviota* (Sunobe 1998; Sunobe and Nakazono 1999) that would further help establish reproductive isolation and reinforce species boundaries. Another possible explanation is that genetic variants that evolved in isolation formed a founding population, until a second wave of individuals from an original population arrived. An example of this type of "independent founder events" has been hypothesized for *Stegastes faciolatus* (Pomacentridae) colonizing the Eastern Islands and Hawaii. *Eviota* sister species like *E. punyit* and *E. sebreei* are considered an example of allopatric speciation followed by a range expansion into sympatry (Tornabene et al. 2016). On the contrary, sympatric sister species *E. dorsopurpurea* and *E. brahmi* (*E. nigriventris* complex), both occur off Papua New Guinea, may have undergone sympatric, or microallopatric speciation (Tornabene et al. 2015).

Our study recovered a monophyletic group containing Clades VII-IX, with specimens from the Samoan Archipelago, Fiji, and Tonga respectively. These three clades are also morphologically unique (morphotype V) and geographically distinct from the rest of the *E. sigillata* complex. They are the easternmost members of the complex, and their sister Clade VI consists of specimens from the Coral Triangle. These three clades are the most recently diverged species in the *sigillata* complex and are likely endemic to their respective archipelagos. This pattern resembles that of the *Eviota guttata* complex, where specimens from Fiji (*E. teresae*) were distinct from those from Samoa (*E. taeiae*) (Erdmann et al. 2023). Fiji, Tonga and Samoa possess high levels of endemism in damselfishes (Pomacentridae, Drew and Amatangelo 2017), which share some life history characteristics with gobies and other cryptobenthic fishes such as limited dispersal abilities (small bodies, small home ranges, benthic eggs, 24-26 day larval duration; Brandl et al. 2018), with the example of anemonefishes (a clade of damselfishes), that have been found to settle close to or in their natal anemones (Jones et al. 2005).

The Coral Triangle (CT) region, known for having the greatest levels of diversity for the entire Indo-Pacific, is represented in the ddRADseq by 27 specimens belonging to Clades III-VI. Several other species complexes within *Eviota* also show high species richness in the CT. For example, in the *E. bifasciata* and *E. nigriventris* complexes, there are five and six distinct lineages, respectively, with all lineages almost exclusively occurring within the CT, suggesting speciation at fine geographic scales (Tornabene et al. 2015). Conversely, there are complexes like the *E. zebrina* complex, with eight species distributed from the Red Sea to the Tongan Archipelago, with four species occurring entirely within the CT (Tornabene et al. 2021). In the *E. sebreei* complex (2 species), *E. sebreei* and *E. punyit* both have broad geographic distributions spanning much of the western Pacific Ocean and Indian Ocean/Red Sea, respectively, but both species overlap in the CT (Tornabene et al. 2016). Patterns from our current study and these past studies suggest that the high richness of *Eviota* in CT, and for reef fishes in general, is a result of many biological (life history, behavioral, habitat) and physical factors (oceanography, geology) (Tornabene et al. 2015, 2016; Greenfield 2017), promoting a combination of multiple speciation modes (Bellwood and Meyer 2009; Bowen et al. 2013).

Clade II in our *sigillata* complex had specimens that were exclusively from the Indian Ocean, from near the *E. sigillata* type locality in the Seychelles, and from the Maldives (Figure 1). Similarly, *E. zebrina* (part of the *E. zebrina* complex) was restricted to the Seychelles and the Maldives, with other species in the complex occurring from Pacific localities or restricted to the Red Sea (Tornabene et al. 2021). In the *E. guttata* complex, *E. guttata* is restricted to the Red Sea, *E. teresae*, *E. taeiae* and *E. albolineata* occur in the Pacific, and another species (yet to be confirmed genetically) may be exclusive to the Indian Ocean (Erdman et al. 2023). There are many fish species that are endemic to the Indian Ocean (Borsa et al. 2016; Fricke et al. 2018), with the high level of endemism present in the Red Sea serving as an "evolutionary incubator" to the Indian Ocean (DiBattista et al. 2013). Biogeographic barriers like the Mid-Indian Ocean and the Indo-Pacific Barrier (present during the Pleistocene), may have been conducive to the present segregation of some marine fish species (Gaither et al. 2015; Hodge and Bellwood 2016), including the Indian Ocean clade here. The Indian Ocean is one of the regions that remains under-sampled for the *Eviota* genus. The *Eviota* specimens collected before the era of genomics relied only on morphological traits for species descriptions (Jewett and Lechner 1983). It is necessary to increase efforts to sample for *Eviota* specimens and for cryptobenthic fishes in general, to better understand patterns of diversity in the region.

CONCLUSION

Our integrative approach to delineating the number of species in the *Eviota sigillata* complex provides great insights on how the different methods can guide us in ascertaining the most likely number of lineages for this group. We discovered at least seven hidden species within the *Eviota* *sigillata* complex. Despite some disagreement amongst the methods used, the phylogenetic tree generated with the genome-wide ddRADseq dataset, in combination with the BDF* method, gave robust evidence for the 9 different species in this clade. This study confirms our suspicion that species richness in dwarfgobies is extremely underestimated. To elucidate all this vast hidden diversity will require concerted efforts that will result in robust museum collections where voucher specimens and metadata are archived for easy information retrieval. Voucher specimens must be collected in under-sampled or never sampled before locations, to understand the complete distributions of these species. A fundamental part of these collections is quality live/fresh coloration pictures to aid in proper taxonomic identifications or variations within species. Fresh specimens must be preserved in 90%ETOH or frozen for genetic analysis, avoiding degradation of the genetic material as much as possible. The wide distribution of dwarfgobies throughout the Indo-Pacific and their rapid rates of population turnover, make this group great models for the study of speciation at multiple geographic scales. They provide great insights on how sympatric and allopatric speciation are shaping the evolution and the distribution of this enigmatic group of coral reef fishes. This study established the foundation for future endeavors to unlock the true scale of species diversity within the entire genus *Eviota* by incorporating genome-wide data, one complex at a time.

ACKNOWLEDGEMENTS

We thank David Greenfield for helpful discussions regarding *Eviota* systematics, and Diego Vaz and Chris Goatley for their help with morphological analyses. We thank Valentine and Motu Vaeoso (Department of Natural Resources, and Coral Reef Advisory Group respectively) in American Samoa for assistance in the field. We thank the people of Vava'u, Tonga and their government agencies, and Karen Stone for facilitating fieldwork. Ray Buckley and Richard Ball provided technical assistance in the field. Emily McFarland, Sam Ghods, Katlyn Fuentes, Sydney Schumaker, and Junda (David) Zhao helped with DNA extractions. Many thanks to Katherine Maslenikov for assistance with cataloging specimens, as well as Katherine Rogers and

Isaiah Rothschild. We acknowledge the Tokyo Museum of Natural History, the California Academy of Sciences, the South African Institute of Aquatic Biodiversity, the National Museum of Natural History: Smithsonian Institution, and the Bishop Museum for their help with specimen loans. We are grateful to Adam Leaché, Steven Roberts, Sam White, Bryan Briones, and Hayden Davis for their assistance with bioinformatics and genetic analyses. This project was in part funded by the Hall Conservation Genetics Research Fund (University of Washington), 2018 Lerner-Gray Memorial Fund of the American Museum of Natural History, the American Academy of Underwater Sciences Foundation, the Seattle Rho Chapter of Washington Women Educators, the William and Dorothy T. Gilbert Ichthyology Research Fund, and NSF Award #1827013.

REFERENCES

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics. 17(2):81–92. doi:10.1038/nrg.2015.28.

Baum DA, Shaw KL. 1995. Genealogical perspectives on the species problem. Experimental and molecular approaches to plant biosystematics, 53(289-303), 123-124.

Bellwood DR, Meyer CP. 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. https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2699.2008.02029.x.

Bessey C, Depczynski M, Goetze JS, Moore G, Fulton CJ, Snell M, Parsons SK, Berry O, Wilson S. 2023. Cryptic biodiversity: A portfolio-approach to coral reef fish surveys. Limnology & Ocean Methods. 21(10):594–605. doi:10.1002/lom3.10567. https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lom3.10567.

Blair C, Bryson RW. 2017. Cryptic diversity and discordance in single-locus species delimitation methods within horned lizards (Phrynosomatidae: *Phrynosoma*). Molecular Ecology Resources. 17, 1168–1182 doi: 10.1111/1755-0998.12658

Borsa P, Durand J-D, Chen W-J, Hubert N, Muths D, Mou-Tham G, Kulbicki M. 2016. Comparative phylogeography of the western Indian Ocean reef fauna. Acta Oecologica. 72:72–86. doi:10.1016/j.actao.2015.10.009. https://linkinghub.elsevier.com/retrieve/pii/S1146609X15300333.

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 15(4):e1006650. doi:10.1371/journal.pcbi.1006650. https://dx.plos.org/10.1371/journal.pcbi.1006650.

Bowen BW, Rocha LA, Toonen RJ, Karl SA. 2013. The origins of tropical marine biodiversity. doi:https://doi.org/10.1016/j.tree.2013.01.018.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 23(19):2633–2635. doi:10.1093/bioinformatics/btm308. https://academic.oup.com/bioinformatics/article/23/19/2633/185151.

Brandl SJ, Goatley CHR, Bellwood DR, Tornabene L. 2018. The hidden half: ecology and evolution of cryptobenthic fishes on coral reefs. Biological Reviews. 93(4):1846–1873. doi:10.1111/brv.12423.

Brandl SJ, Tornabene L, Goatley CHR, Casey JM, Morais RA, Côté IM, Baldwin CC, Parravicini V, Schiettekatte NMD, Bellwood DR. 2019. Demographic dynamics of the smallest marine vertebrates fuel coral reef ecosystem functioning. Science. 364(6446):1189–1192. doi:10.1126/science.aav3384.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. 2012. Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. Molecular Biology and Evolution. 29(8):1917–1932. doi:10.1093/molbev/mss086. https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mss086.

Carstens BC, Pelletier TA, Reid NM, Salter JD. 2013. How to fail at species delimitation. Molecular Ecology 22, 4369–4383 doi: 10.1111/mec.12413

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology. 22(11):3124–3140. doi:10.1111/mec.12354. https://onlinelibrary.wiley.com/doi/10.1111/mec.12354.

Catchen JM, Hohenlohe PA, Bernatchez L, Funk WC, Andrews KR, Allendorf FW. 2017. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. Molecular Ecology Resources. 17(3):362–365. doi:10.1111/1755-0998.12669.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics. 27(15):2156–2158. doi:10.1093/bioinformatics/btr330. https://academic.oup.com/bioinformatics/article/27/15/2156/402296.

Degnan JH, Rosenberg NA. 2006. Discordance of Species Trees with Their Most Likely Gene Trees. PLoS Genet. 2(5):e68. doi:10.1371/journal.pgen.0020068. https://dx.plos.org/10.1371/journal.pgen.0020068.

Dellicour S, Flot JF. 2015. Delimiting species-poor data sets using single moleculecular markers: A study of barcode gaps, haplowebs and GMYC. Systematic Biology. 64(6):900-908. https://doi.org/10.1093/sysbio/syu130

Depczynski M, Bellwood DR. 2003. The role of cryptobenthic reef fishes in coral reef trophodynamics. Marine Ecology Progress Series. 256:183–191. doi:10.3354/meps256183.

Depczynski M, Bellwood DR. 2004. Microhabitat utilization patterns in cryptobenthic coral reef fish communities. Marine Biology. 145(3):455–463. doi:10.1007/s00227-004-1342-6.

Depczynski M, Bellwood David R. 2005. Shortest recorded vertebrate lifespan found in a coral reef fish. Current Biology. 15(8):288–289.

De Queiroz K. 2007. Species Concepts and Species Delimitation, Systematic Biology, 56(6): 879 – 886. DOI: 10.1080/10635150701701083

DeRaad D. 2021. SNPfiltR: an R package for interactive and reproducible SNP filtering. doi:10.22541/au.163976415.53888836/v1.

Dias RM, Lima SMQ, Mendes LF, Almeida DF, Paiva PC, Britto MR. 2019. Different speciation processes in a cryptobenthic reef fish from the Western Tropical Atlantic. Hydrobiologia. 837(1):133–147. doi:10.1007/s10750-019-3966-z.

DiBattista JD, Berumen ML, Gaither MR, Rocha LA, Eble JA, Choat JH, Craig MT, Skillings DJ, Bowen BW. 2013. After continents divide: comparative phylogeography of reef fishes from the Red Sea and Indian Ocean. Journal of Biogeography. 40(6):1170–1181. doi:10.1111/jbi.12068. https://onlinelibrary.wiley.com/doi/10.1111/jbi.12068.

Drew JA, Amatangelo KL. 2017. Community assembly of coral reef fishes along the Melanesian biodiversity gradient. PLoS ONE. 12(10):e0186123. doi:10.1371/journal.pone.0186123. https://dx.plos.org/10.1371/journal.pone.0186123.

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genet Resour. 4(2):359–361. doi:10.1007/s12686-011-9548-7. http://link.springer.com/10.1007/s12686-011-9548-7.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE. 6(5). doi:10.1371/journal.pone.0019379.

Erdmann MV, Greenfield DW, Tornabene L. 2023 Jun 6. *Eviota taeiae*, a new dwarfgoby (Teleostei: Gobiidae) of the Eviota guttata complex from Samoa. doi:10.5281/ZENODO.8011284. https://zenodo.org/record/8011284.

Eschmeyer WN, Fricke R, Fong JD, Polack DA. 2010. Marine fish diversity: history of knowledge and discovery (Pisces). Zootaxa. 2525(1). doi:10.11646/zootaxa.2525.1.2. https://mapress.com/zt/article/view/zootaxa.2525.1.2.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*, 14(8):2611-2620.
Ezard T, Fujisawa T, Barraclough TG. 2009. splits: SPecies' LImits by Threshold Statistics. R package version 1.0-14/r31.

Fricke R, Mahafina J, Behivoke F, Jaonalison H, Léopold M, Ponton D. 2018. Annotated checklist of the fishes of Madagascar, southwestern Indian Ocean, with 158 new records.

Froese R, Pauly D. 2019. Fishbase. World Wide Web electronic publication. FishBase.:www.fishbase.org.

Fujisawa T, Barraclough TG. 2013. Delimiting Species Using Single-Locus Data and the Generalized Mixed Yule Coalescent Approach: A Revised Method and Evaluation on Simulated Data Sets. Systematic Biology. 62(5):707–724. doi:10.1093/sysbio/syt033. https://academic.oup.com/sysbio/article/62/5/707/1684890.

Gaither MR, Bernal MA, Coleman RR, Bowen BW, Jones SA, Simison WB, Rocha LA. 2015. Genomic signatures of geographic isolation and natural selection in coral reef fishes. Molecular Ecology. 24(7):1543–1557. doi:10.1111/mec.13129.

Greenfield DW, Erdmann MV, Teitelbaum A. 2024. *Eviota bacata*, a new dwarfgoby (Teleostei: Gobiidae) from New Caledonia. Journal of the Ocean Science Foundation, 41:14–21. doi: https://doi.org/10.5281/zenodo.11114207.

Greenfield D. 2017. An overview of the dwarfgobies, the second most speciose coral-reef fish genus (Teleostei: Gobiidae: Eviota). Journal of the Ocean Science Foundation. 29:32–54. doi:10.5281/zenodo.1115683.

Greenfield D. 2021. Addendum to the 2016 key to the dwarfgobies (Teleostei: Gobiidae: *Eviota*). Journal of the Ocean Science Foundation. 38:1–12. doi:10.5281/zenodo.4458248.

Greenfield DW, Tornabene L. 2014. *Eviota brahmi* n. sp. from Papua New Guinea, with a redescription of *Eviota nigriventris* (Teleostei: Gobiidae). Zootaxa 3793(1):133-146. doi.org/10.11646/zootaxa.3793.1.6

Greenfield DW, Tornabene L, Erdmann MV, Pada DN. 2019. *Eviota gunawanae*, a new microendemic dwarfgoby from the Fakfak Peninsula, West Papua, Indonesia (Teleostei: Gobiidae). Journal of the Ocean Science Foundation, 32, 57–67. doi: https://doi.org/10.5281/zenodo.2616753 urn:lsid:zoobank.org:pub:FBBF2441-6587-4C73-9F04-CB76E47D90A1

Greenfield DW, Tornabene L, Gómez-Buckley M, Erdmann MV. 2018. *Eviota maculosa*, a new dwarfgoby from the Western Pacific Ocean (Teleostei: Gobiidae). Journal of the Ocean Science Foundation, 31, 18–31. doi: http://dx.doi.org/10.5281/zenodo.1445810 urn:lsid:zoobank.org:pub:5ABD20DD-B260-4696-8903-D32A67C7BEB1

Greenfield D, Randall JE. 2010. Four new fishes of the genus *Eviota* from the Western Pacific, with clarification of *Eviota guttata* and *Eviota albolineata* (Teleostei: Gobiidae). https://www.researchgate.net/publication/267332956.

Greenfield D, Winterbottom R. 2016. A key to the dwarfgoby species (Teleostei: Gobiidae: Eviota) described between 1871 and 2016. Journal of the Ocean Science Foundation. 24(April):35–90. doi:10.5281/zenodo.219620.

Hanahara N. 2023. Cryptic diversity of *Eviota* (Teleostei: Gobiidae) and their habitat use in the shallow waters of Okinawa Island. Marine Biodiversity 53:61 https://doi.org/10.1007/s12526-023-01369-6

Harrison IJ. 1989. Specialization of the gobioid *palatopterygoquadrate* complex and its relevance to gobioid systematics. Journal of Natural History. 23(2):325–353. doi:10.1080/00222938900770211. http://www.tandfonline.com/doi/abs/10.1080/00222938900770211.

Herler J. 2007. Microhabitats and ecomorphology of coral- and coral rock-associated gobiid fish (Teleostei: Gobiidae) in the northern Red Sea. Marine Ecology. 28(SUPPL. 1):82–94. doi:10.1111/j.1439-0485.2007.00165.x.

Hodge JR, Bellwood DR. 2016. The geography of speciation in coral reef fishes: the relative importance of biogeographical barriers in separating sister-species. Journal of Biogeography. 43(7):1324–1335. doi:10.1111/jbi.12729. https://onlinelibrary.wiley.com/doi/10.1111/jbi.12729.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. Begun DJ, editor. PLoS Genet. 6(2):e1000862. doi:10.1371/journal.pgen.1000862. https://dx.plos.org/10.1371/journal.pgen.1000862.

Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrel JC, Cullingham, Andrew RL. 2017. The K=2 conundrum. Molecular Ecology. 26:3594–3602. DOI: 10.1111/mec.14187

Jenkins OP. 1903. Report on collections of fishes made in the Hawaiian Islands, with descriptions of new species. Bulletin of the U.S. Fish Commission. 22:415–511.

Jewett SL, Lachner EA. 1983. Seven new species of the Indo-Pacific genus *Eviota* (Pisces: Gobiidae). Proceedings of the Biological Society of Washington. 96:780–806. https://www.biodiversitylibrary.org/part/45377.

Jones GP, Planes S, Thorrold SR. 2005. Coral reef fish larvae settle close to home. Current Biology. 15(14):1314–1318. doi:10.1016/j.cub.2005.06.061. https://linkinghub.elsevier.com/retrieve/pii/S0960982205007128.

Jordan DS, Seale A. 1906. The fishes of Samoa. Description of the species found in the archipelago, with a provisional check-list of the fishes of Oceania. Bulletin of the Beureau of Fisheries. Vol. 25. US Government Printing Office.

Kalinowski S.T. 2011. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity, 106(4): 625-632.

Kass RE, Raftery AE. 1995. Bayes factors. Journal of the American statistical association, 90(430), pp.773-795.

Kassambara A, Mundt F. 2017. Package ‘factoextra’. Extract and visualize the results of multivariate data analyses, 76(2).

Kingman JFK. 1982. The coalescent. Stochastic Processes and their Applications 13 (1982) 235-248

Klunzinger CB. 1871. Synopsis der Fische des Rothen Meeres. II. Theil. Verhandlungen der K.-K. zoologischbotanischen Gesellschaft in Wien. 21: 441-688.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 35(21):4453–4455. doi:10.1093/bioinformatics/btz305. [accessed 2024 Apr 15]. https://academic.oup.com/bioinformatics/article/35/21/4453/5487384.

Lachner EA, Karnella SJ. 1978. Fishes of the genus *Eviota* of the Red Sea with descriptions of three new species (Teleostei, Gobiidae). Smithsonian Contributions to Zoology. doi:10.5479/si.00810282.286.

Lachner EA, Karnella SJ. 1980. Fishes of the Indo-Pacific Genus *Eviota* with Descriptions of Eight New Species (Teleostei: Gobiidae) Smithsonian Contributions to Zoology – Number 315.

Lê S, Josse J, Husson F. 2008. FactoMineR : An *R* Package for Multivariate Analysis. J Stat Soft. 25(1). doi:10.18637/jss.v025.i01. http://www.jstatsoft.org/v25/i01/.

Leaché A, Ogilvie HA. 2016. Bayes Factor Delimitation of Species (*with genomic data; BFD*): A Tutorial and Worked Example.

Leaché AD, Bouckaert RR. 2018. Species Trees and Species Delimitation with SNAPP: A Tutorial and Worked Example.

Leaché, AD, Fujita MK. 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). Proceedings of the Royal Society B 277:3071-3077.

Leaché AD, Fujita MK, Minin VN, Bouckaert RR. 2014. Species delimitation using genome-wide SNP Data. Systematic Biology. 63(4):534–542. doi:10.1093/sysbio/syu018.

Liu L, Anderson C, Pearl D, Edwards SV. 2019. Modern phylogenomics: Building phylogenetic trees using the multispecies coalescent model. In: Methods in Molecular Biology.

Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A. 2017. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Molecular ecology resources. 17(2):142–152. doi:10.1111/1755-0998.12635.

Luo D, Wan X, Liu J, Tong T. 2018. Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Stat Methods Med Res. 27(6):1785-1805. doi: 10.1177/0962280216669183. Epub 2016 Sep 27. PMID: 27683581.

Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñero D, Emerson BC. 2015. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Molecular ecology resources, 15(1):28-41.

Meek MH, Larson WA. 2019. The future is now: Amplicon sequencing and sequence capture usher in the conservation genomics era. Molecular Ecology Resources. doi:10.1111/1755-0998.12998.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE). New Orleans, LA, USA: IEEE. p. 1–8. http://ieeexplore.ieee.org/document/5676129/.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17(2):240–248. doi:10.1101/gr.5681207. http://genome.cshlp.org/lookup/doi/10.1101/gr.5681207.

Miralles A, Vences M. 2013. New Metrics for Comparison of Taxonomies Reveal Striking Discrepancies among Species Delimitation Methods in *Madascincus* Lizards. PLoS ONE 8(7): e68242. doi:10.1371/journal.pone.0068242

Monaghan MT, Wild R, Elliot M, Fujisawa T, Balke M, Inward DJG, Lees DC, Ranaivosolo R, Eggleton P, Barraclough TG, et al. 2009. Accelerated Species Inventory on Madagascar Using Coalescent-Based Models of Species Delineation. Systematic Biology. 58(3):298–311. doi:10.1093/sysbio/syp027. https://academic.oup.com/sysbio/article/58/3/298/1611030.

Munday PL. 2001. Fitness consequences of habitat use and competition among coral-dwelling fishes. Oecologia. doi:10.1007/s004420100690.

Munday PL, Jones GP, Caley MJ. 1997. Habitat specialisation and the distribution and abundance of coral-dwelling gobies. Marine Ecology Progress Series. 152(1–3):227–239. doi:10.3354/meps152227.

Orr JW and Hawkins S. 2008. Species of the rougheye rockfish complex: resurrection of *Sebastes melanostictus* (Matsubara, 1934) and a redescription of *Sebastes aleutianus* (Jordan and Evermann, 1898)(Teleostei: Scorpaeniformes). Fishery Bulletin 106(2):111-134.

Paris JR, Stevens JR, Catchen JM. 2017. Lost in parameter space: a road map for STACKS. Johnston S, editor. Methods Ecol Evol. 8(10):1360–1373. doi:10.1111/2041-210X.12775. [accessed 2024 Apr 15]. https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.12775.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE. 7(5). doi:10.1371/journal.pone.0037135.

Piñeros VJ, Del R. Pedraza-Marrón C, Betancourt-Resendes I, Calderón-Cortés N, Betancur-R R, Domínguez-Domínguez O. 2022. Genome-wide species delimitation analyses of a silverside fish species complex in central Mexico indicate taxonomic over-splitting. BMC Ecol Evo. 22(1):108. doi:10.1186/s12862-022-02063-0. https://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-022-02063-0.

Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP. 2006. Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. Hedin M, editor. Systematic Biology. 55(4):595–609. doi:10.1080/10635150600852011. https://academic.oup.com/sysbio/article/55/4/595/1677524.

Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu MV. 2013. An overview of STRUCTURE: Applications, parameter settings, and supporting software. Frontiers in Genetics. 4(MAY):1–13. doi:10.3389/fgene.2013.00098.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics. 155(2):945–959. doi:10.1093/genetics/155.2.945. https://academic.oup.com/genetics/article/155/2/945/6048111.

Puillandre N, Brouillet S, Achaz G. 2021. ASAP: assemble species by automatic partitioning. Molecular Ecology Resources. 21(2):609–620. doi:10.1111/1755-0998.13281. https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13281.

Puillandre N, Modica MV, Zhang Y, Sirovich L, Boisselier M -C., Cruaud C, Holford M, Samadi S. 2012. Large-scale species delimitation method for hyperdiverse groups. Molecular Ecology. 21(11):2671–2691. doi:10.1111/j.1365-294X.2012.05559.x. https://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2012.05559.x.

Rambaut A. 2012. FigTree v1.4.4. Available from: http://tree.bio.ed.ac.uk/software/figtree/

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Susko E, editor. Systematic Biology. 67(5):901–904. doi:10.1093/sysbio/syy032. https://academic.oup.com/sysbio/article/67/5/901/4989127.

Rannala B, Edwards SV, Leaché A, Yang Z. 2020. The Multispecies Coalescent Model and Species Tree Inference. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, Phylogenetics in the Genomic Era, chapter No. 3.3, pp. 3.3:1–3.3:21.

Shi J, Shimizu H, Zou C. 2010. Differentiation and population subdivision in *Picea mongolica* based on microsatellite analyses. Biochemical Systematics and Ecology. 1;38(6):1122-8.

Slowikowski K. 2024. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. https://ggrepel.slowkow.com/, https://github.com/slowkow/ggrepel.

Spalding MD, Grenfell AM. 1997. New estimates of global and regional coral reef areas. Coral Reefs. 16(4):225–230. doi:10.1007/s003380050078. http://link.springer.com/10.1007/s003380050078.

Sukumaran J, Knowles LL. 2017. Multispecies coalescent delimits structure, not species. Proc Natl Acad Sci USA. 114(7):1607–1612. doi:10.1073/pnas.1607921114. https://pnas.org/doi/full/10.1073/pnas.1607921114.

Sunobe T. 1998. Reproductive behavior in six species of *Eviota* (Gobiidae) in aquaria. Ichthyol Res. 45(4):409–412. doi:10.1007/BF02725195. http://link.springer.com/10.1007/BF02725195.

Sunobe T, Nakazono A. 1999. Alternative mating tactics in the gobiid fish, *Eviota prasina*. Ichthyological Research. 46(2):212–215. doi:10.1007/BF02675442. https://link.springer.com/10.1007/BF02675442.

Thacker CE. 2003. Molecular phylogeny of the gobioid fishes (Teleostei: Perciformes: Gobioidei). Molecular Phylogenetics and Evolution. 26(3):354–368. doi:10.1016/S1055-7903(02)00361-5. https://linkinghub.elsevier.com/retrieve/pii/S1055790302003615.

Tornabene L, Ahmadia GN, Williams JT. 2013a. Four new species of dwarfgobies (Teleostei: Gobiidae: *Eviota*) from the Austral, Gambier, Marquesas and Society Archipelagos, French Polynesia. Systematics and Biodiversity. doi:10.1080/14772000.2013.819822.

Tornabene L, Ahmadia GN, Berumen ML, Smith DJ, Jompa J, Pezold F, 2013b. Evolution of microhabitat association and morphology in a diverse group of cryptobenthic coral reef fishes (Teleostei: Gobiidae: *Eviota*). Mol. Phylogenet. Evol. 66:391–400.

Tornabene L, Greenfield DW, Erdmann MV. 2021. A review of the *Eviota zebrina* complex, with descriptions of four new species (Teleostei, Gobiidae). ZooKeys 1057:149–184. https://doi.org/10. 3897/zookeys.1057.66675.

Tornabene L, Valdez S, Erdmann M, Pezold F. 2015. Support for a “Center of Origin” in the Coral Triangle: Cryptic diversity, recent speciation, and local endemism in a diverse lineage of reef fishes (Gobiidae: Eviota). Molecular Phylogenetics and Evolution. 82(PA):200–210. doi:10.1016/j.ympev.2014.09.012. http://dx.doi.org/10.1016/j.ympev.2014.09.012.

Tornabene L, Valdez S, Erdmann MV, Pezold FL. 2016. Multi-locus sequence data reveal a new species of coral reef goby (Teleostei: Gobiidae: *Eviota*), and evidence of Pliocene vicariance across the Coral Triangle. Journal of Fish Biology. 88(5):1811–1834. doi:10.1111/jfb.12947.

Veron JE, Devantier LM, Turak E, Green AL, Kininmonth S, Stafford-Smith M, Peterson N. 2009. Delineating the coral triangle. Galaxea, Journal of Coral Reef Studies. 11(2):91-100.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN. 2005. DNA barcoding Australia’s fish species. Phil Trans R Soc B. 360(1462):1847–1857. doi:10.1098/rstb.2005.1716. https://royalsocietypublishing.org/doi/10.1098/rstb.2005.1716.

Winterbottom R, Hanner RH, Burridge M, Zur M. 2014. A cornucopia of cryptic species - A DNA barcode analysis of the gobiid fish genus *Trimma* (Percomorpha, Gobiiformes). ZooKeys. 381:79–111. doi:10.3897/zookeys.381.6445.

Winterbottom R, Hanner RH, Erdmann MV, Choffe K, Zur M, Steinke C. 2020. The cryptic cornucopia revisited – an extended analysis of the COI gene in the gobiid fish genus *Trimma* (Percomorpha : Gobiiformes). 132(2020):91–132.

Wickham H 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wickham H, Wickham MH. 2017. Package ‘tidyr’. Easily Tidy Data with’spread’and’gather ()’Functions.

Yarberry W. 2021. Dplyr. CRAN recipes: DPLYR, stringr, lubridate, and regex in R, pp.1-58.

CHAPTER 2 FIGURES

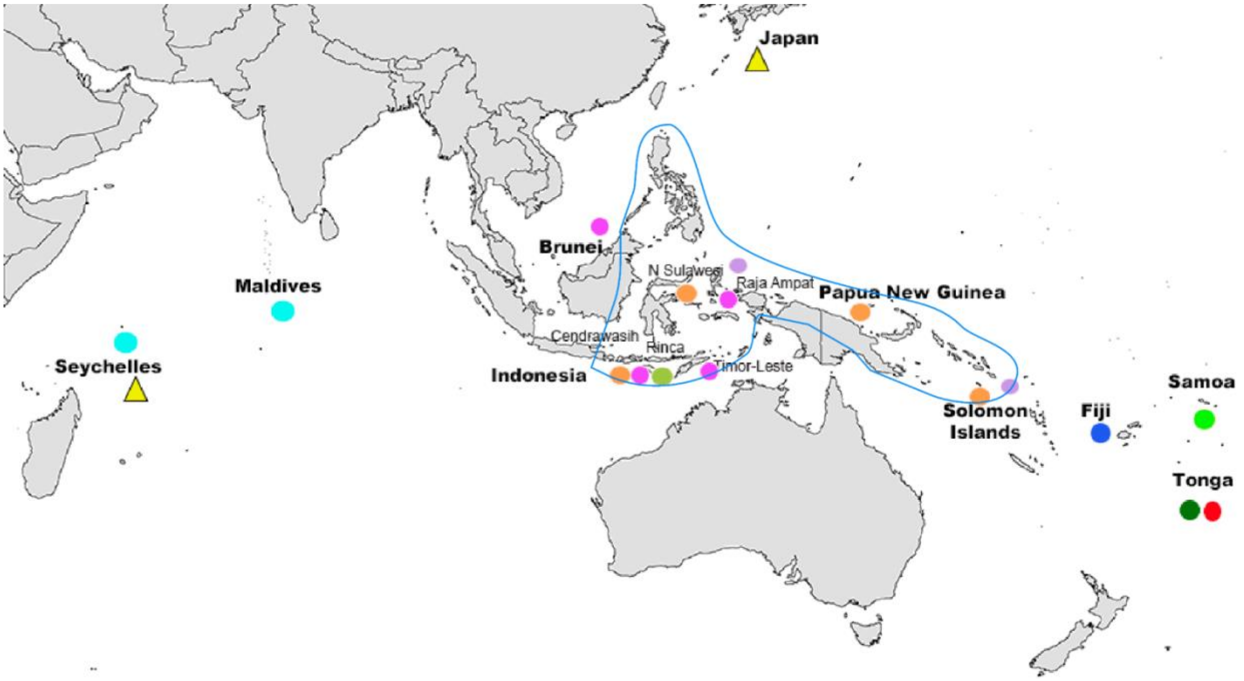


Figure 1. Sampling locations for specimens of the *Eviota sigillata* complex used in this study. Colors correspond to the clades on the genome-wide data RAxML tree (Figure 3). Also included (yellow triangles), are the type localities for the two nominal species in the complex, *E. sigillata* from the Seychelles, and *E. shimadai* from Japan. Area within the blue line demarcates the Coral Triangle (according to Veron et al. 2019).

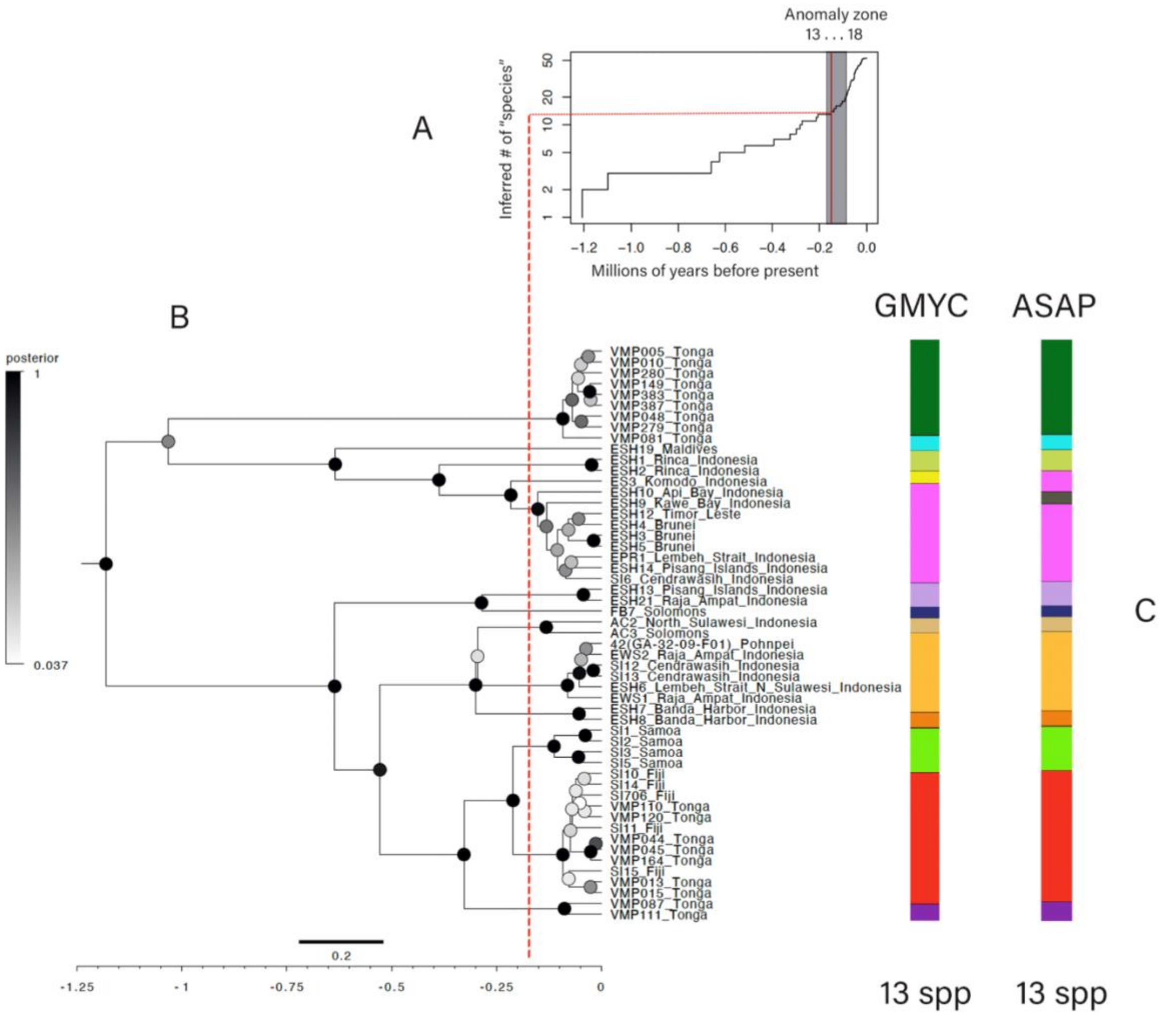


Figure 2. Generalized Mixed Yule-Coalescence (GMYC) plot (A) showing the best supported 13 species model (red discontinued line) and the “anomaly zone” in the grey rectangle. Ultrametric calibrated BEAST tree for the COI data (B). Color bars (C) represent the results of both GMYC and Assemble Species by Automatic Partitioning (ASAP) analysis, showing 13 species model represented by each of the colored boxes (colors are coded to match the ones used for the nine clades in the ddRADseq tree – Figure 3).

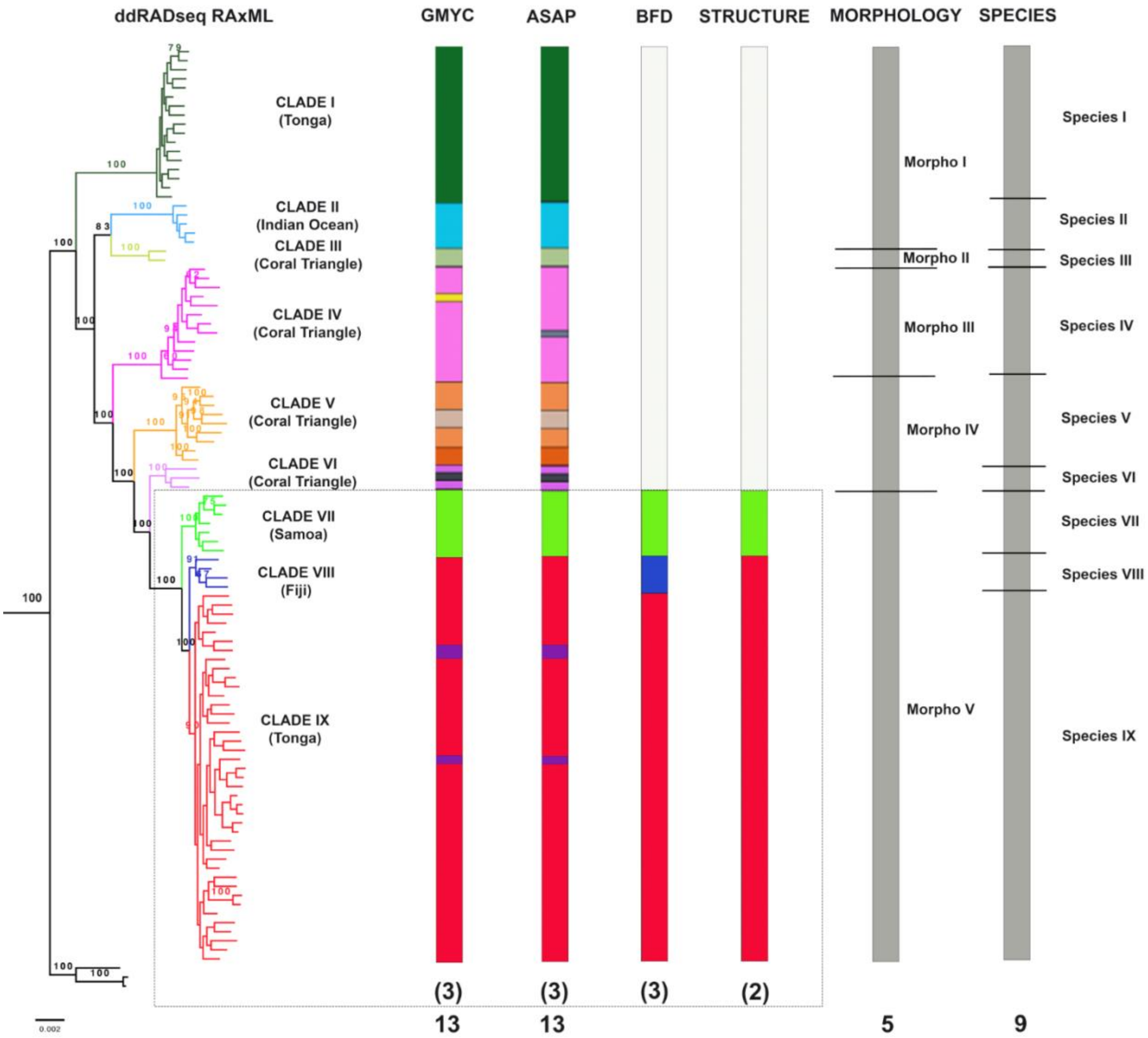


Figure 3. Phylogeny inferred from genome-wide SNP data, with the nine strongly supported, color coded clades (left). Node numbers represent bootstrap support values (only >59 values are included). The results from the species delimitation methods are represented by the bars with the number of ‘species’ for each method noted at the bottom of each column. The number of species from Tonga, Fiji and Samoa are included in parentheses. ASAP and GMYC results from the COI dataset are included here for comparison. The BFD and Structure methods were only used to analyze the Tonga, Fiji, and Samoa samples. The Morphology column depicts the morphotypes based on pore patterns and second dorsal and anal fin-ray counts.

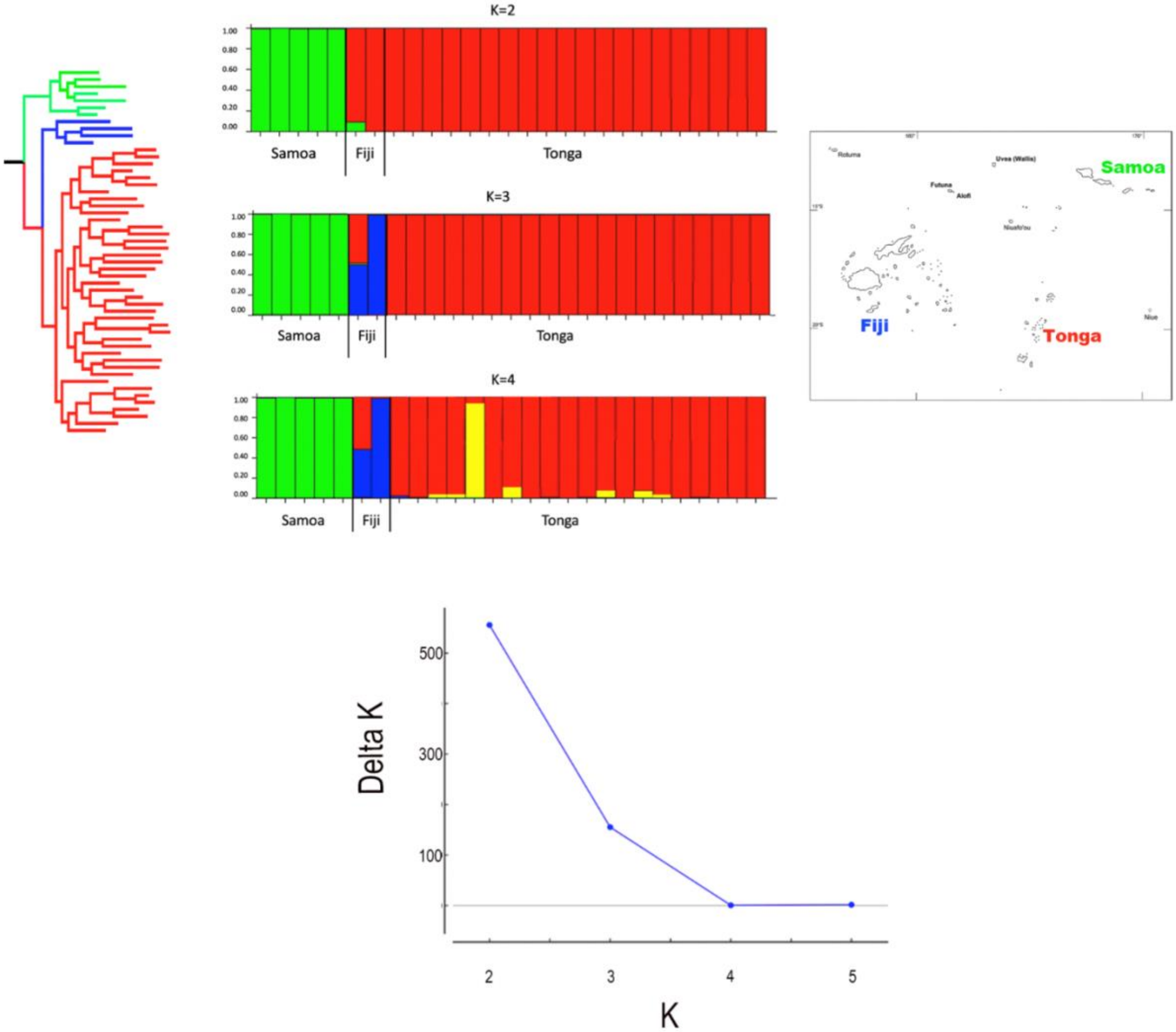
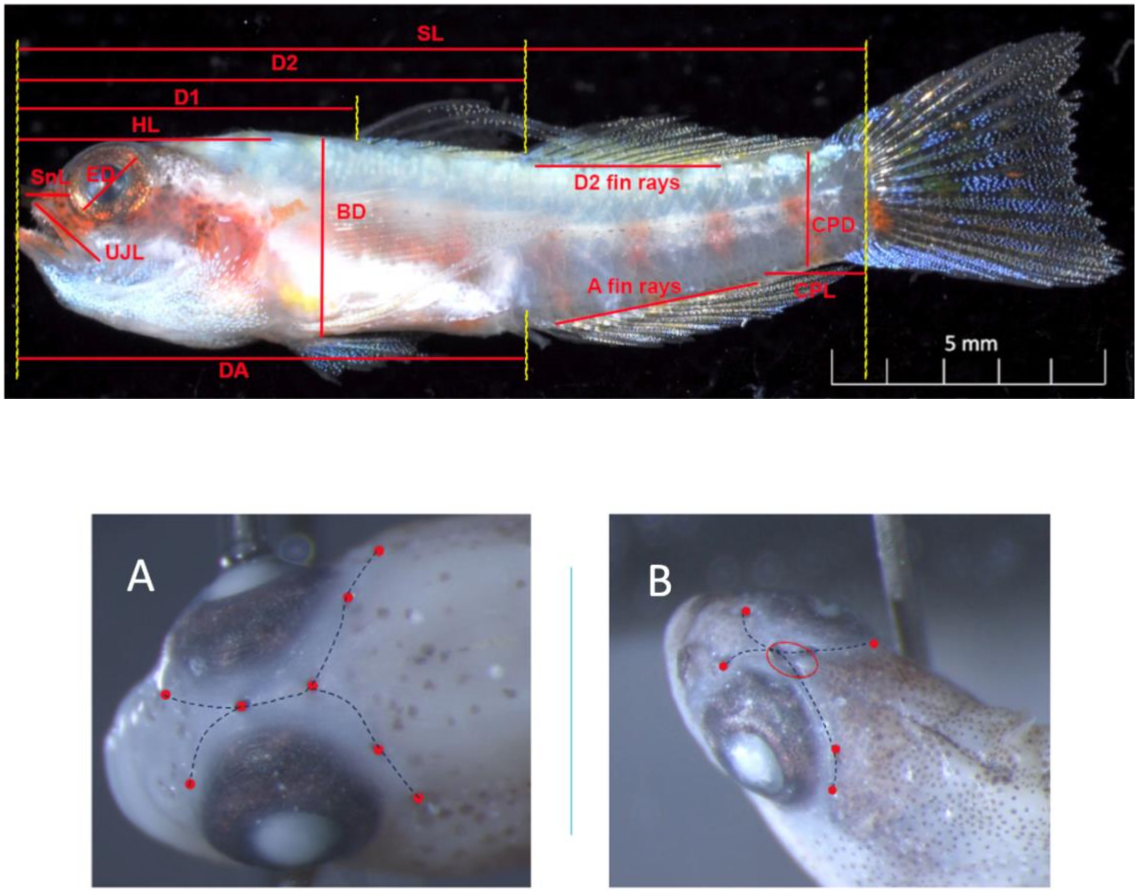


Figure 4. Structure plots for K values 2-4, for the genome wide SNP subset of samples from the Tonga, Fiji, and Samoa regions, alongside the ddRADseq RAxML tree for these samples. The bar plots represent the cluster membership coefficients for the 27 individuals genotyped at 1,491 loci. Each individual is represented by a vertical line partitioned into segments of different colors. The length of each segment indicates the posterior probability of membership for each group. Below, structure Harvester plot showing K=2 as the optimal value for this data subset.



E. sigillata (Seychelles)

E. shimadai (Japan)

Figure 5. Morphometric variables included in our analyses, and cephalic sensory-pore patterns present among examined specimens. Picture A shows the diagnostic character for the type series of *E. sigillata*, with two interorbital pores, while picture B, shows the diagnostic character for the type series of *E. shimadai*, with one interorbital central open depression.

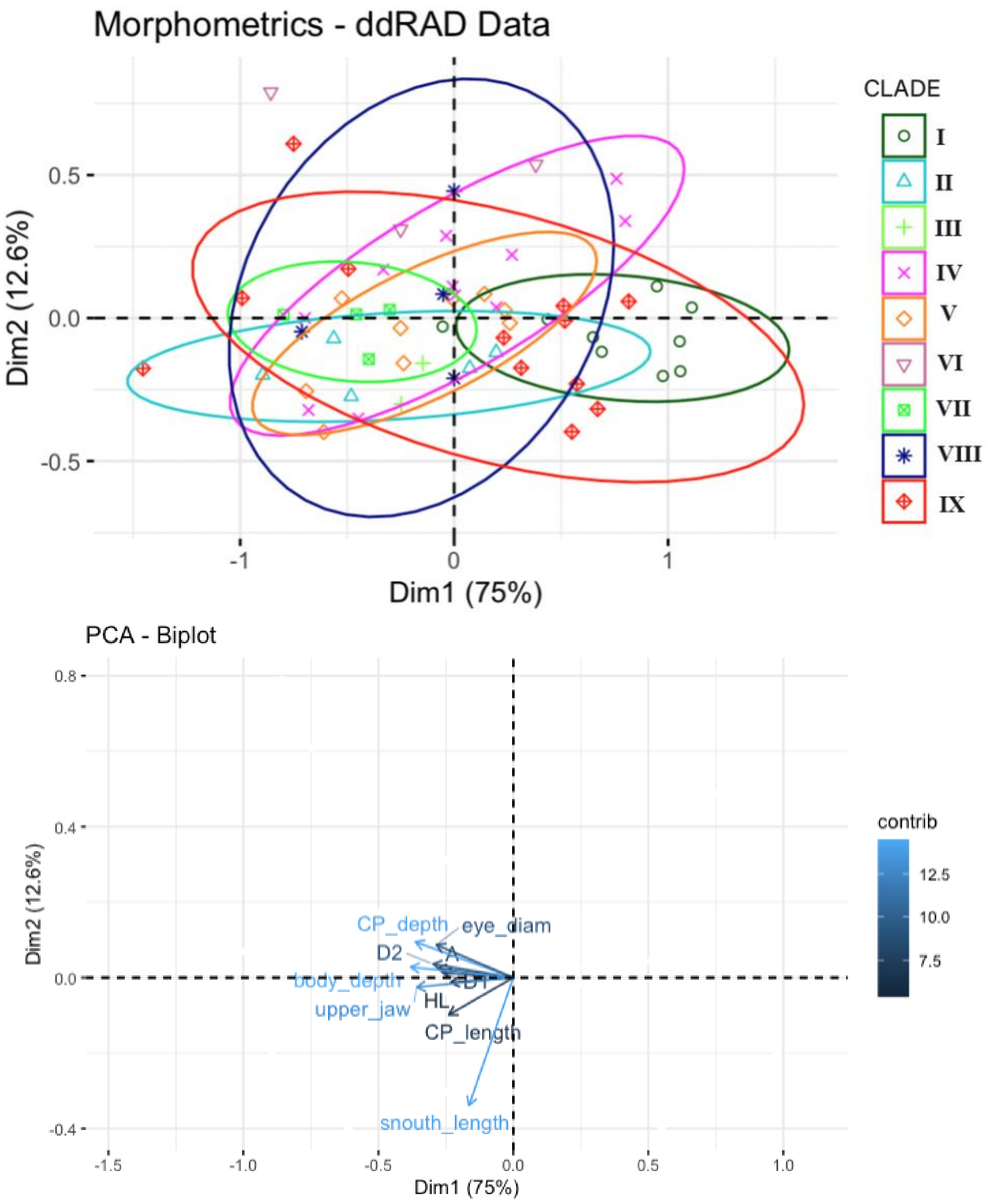


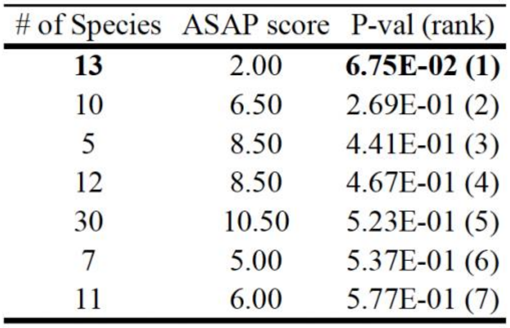
Figure 6. Principal component analysis and loadings for morphometrics of the nine clades from the genome-wide SNP RAxML tree. Abbreviations: A= distance from snout to base of annal fin spine; D1 & D2, distance from snout to base of first and second dorsal spin rays respectively; CP=Caudal peduncle; HL= head length.

CHAPTER 2 TABLES

Table 1. Molecular values of datasets for each analysis. The last three columns are the criteria used to filter datasets. Note that for the RAxML tree, all SNPs (vs. only the first SNP per locus) were included. For the SNAPP-BFD* dataset, extra filtering was performed to remove loci that had missing data for an entire lineage (putative species).

Analysis	Taxa	Loci	Variant sites analyzed (SNPs)	Total bp in phylogenetic analysis (incl. invariant)	Min taxa per locus	Max allele freq.	Max missing data per taxa
COI tree, ASAP, GMYC	53	1	328	1142	-	-	-
RADseq tree	104	4591	85608	660013	0.5	-	-
SNAPP-BFD*	29	1640	1640	-	0.5	0.5	0.8
Structure	27	4235	4235	-	0.5	0.2	0.75

Table 2. Results from the Assemble Species for Automatic Partitioning (ASAP) for the COI dataset, highest scoring model in bold.



# of Species	ASAP score	P-val (rank)
13	2.00	**6.75E-02 (1)**
10	6.50	2.69E-01 (2)
5	8.50	4.41E-01 (3)
12	8.50	4.67E-01 (4)
30	10.50	5.23E-01 (5)
7	5.00	5.37E-01 (6)
11	6.00	5.77E-01 (7)

Table 3. Bayes Factor Delimitation (BDF) results for *E. sigillata*, for the subgroup of Samoa, Fiji, and Tonga samples (n=27, 1,640 loci). MLE= Marginal likelihood estimate, BF= Bayes factor. The BFD best supports the model with three species.

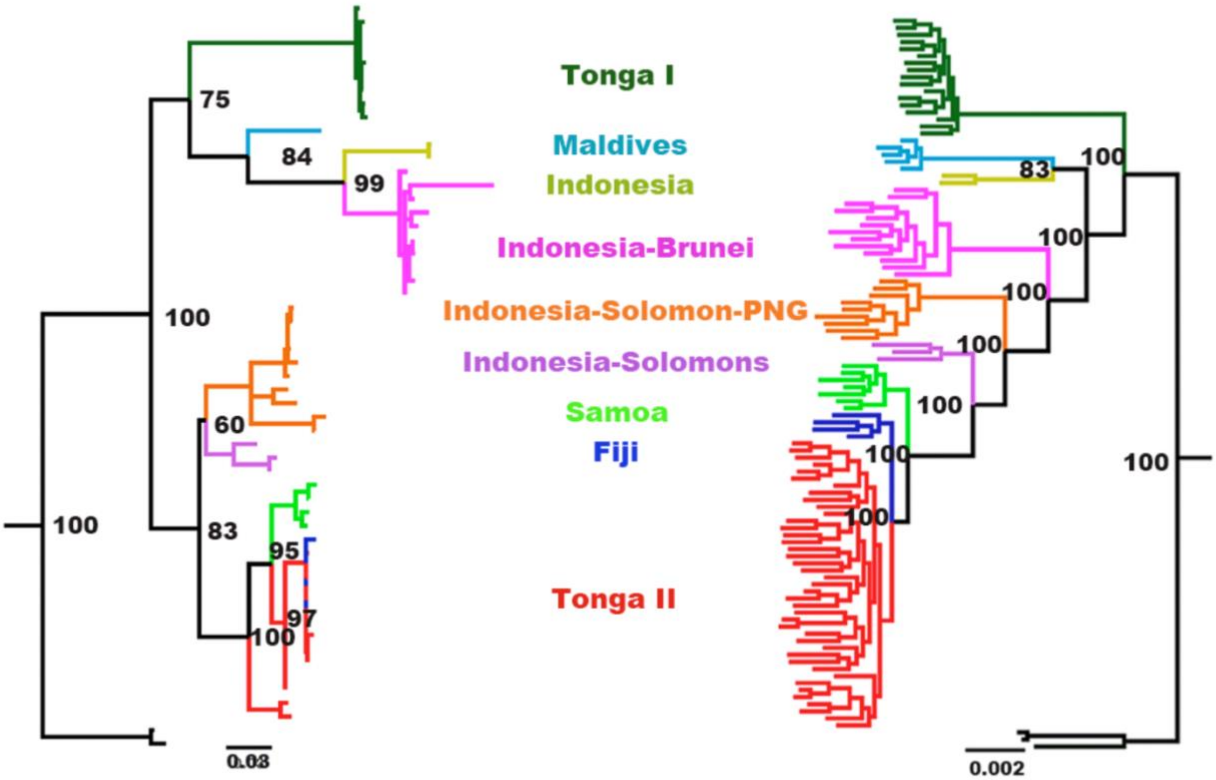
Model	Species	MLE	Rank	BF
Lump Samoa, Fiji, and Tonga	1	-30998.244	3	--
Lump Fiji and Tonga + Samoa	2	-28957.835	2	4080.818
Split Samoa, Fiji, and Tonga	**3**	**-28812.347**	1	**4371.794**

BF = 2 x (MLE1 - MLE0)

Table 4. Main meristic characters observed in specimens from Clades I-IX in the ddRADseq tree, with number of samples for each clade in parentheses. The two nominal species of the *Eviota sigillata* complex are also included as reference. Morphotypes in last column correspond to the ones presented in Figure 3. * 9/8 usual dorsal/anal fin-ray formula, but 8/7 observed with geographic variation (Jewett and Lachner 1983). ** 8/8 usual dorsal/anal fin-ray formula, but 8/7 or 9/8 also observed (Greenfield and Randall 2010). *** Papua New Guinea.

	Region	Interorbital pores	D2 fin rays	Anal fin rays	Morphotypes
*E. sigillata**	Seychelles (Type locality)	2	9	8	Holotype
*E. shimadai***	Ryukyu Islands, Japan (Type locality)	1	8	8	Holotype
Clade I (17)	Tonga I	2, 1	9	8	Morpho I
Clade II (5)	Maldives & Seychelles	2, 1	9	8	
Clade III (2)	Indonesia	2	8	7	Morpho II
Clade IV (13)	Indonesia - Brunei	1	8	8	Morpho III
Clade V (9)	Indonesia - Solomons - PNG***	1	9	8	Morpho IV
Clade VI (3)	Indonesia - Solomons	1	9	8	
Clade VII (7)	Samoa	2	9	8	Morpho V
Clade VIII (4)	Fiji	2	9	8	
Clade IX (41)	Tonga II	2	9	8	

CHAPTER 2 SUPPLEMENT



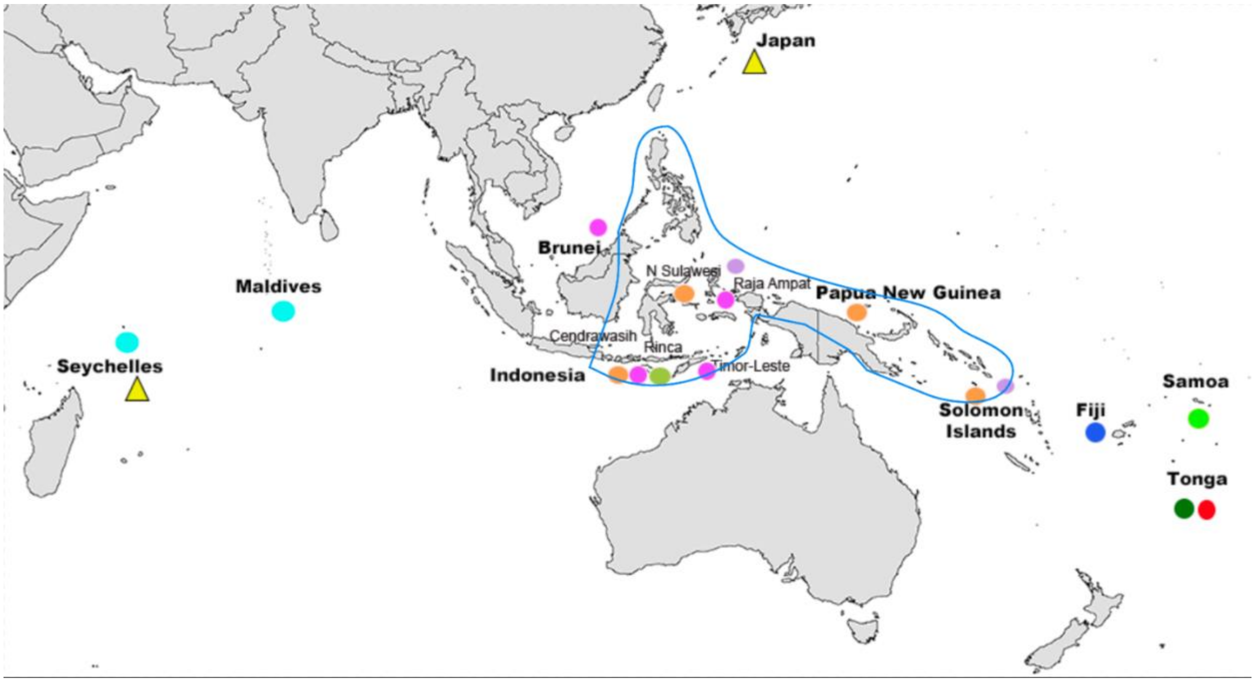


Figure S1. Maximum likelihood trees for both the mitochondrial and ddRADseq trees and location map for samples with corresponding colors for each clade. Yellow triangles mark the type locality for the two nominal species, *Eviota sigillata* (Seychelles) and *Eviota shimadai* (Japan).

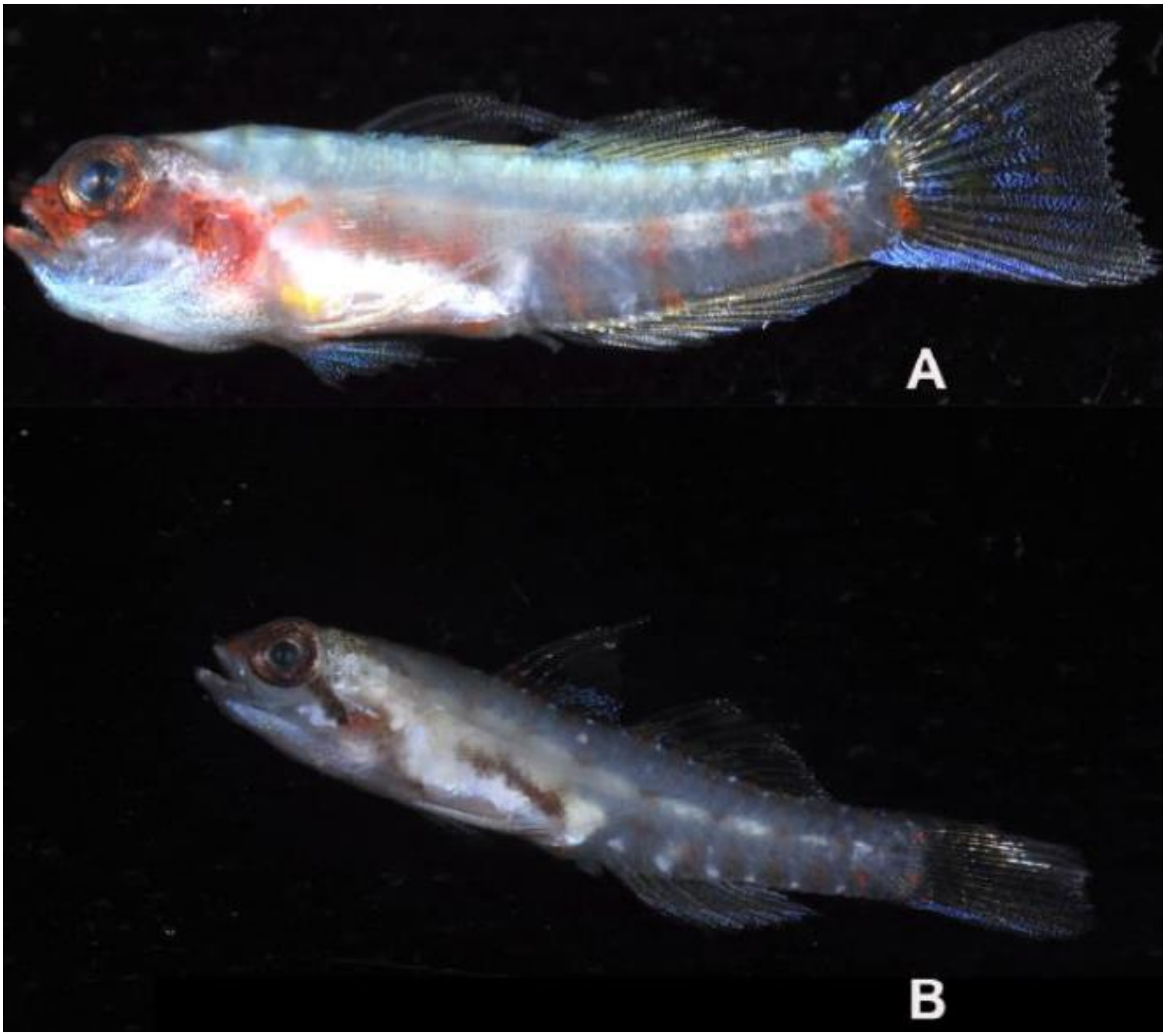


Figure S2. *Eviota cf. sigillata* representatives from Clade I (B) and Clade IX (A). Picture credit, Marta Gómez-Buckley.

Table S1. Geographic distribution of samples collected for this study, indicating the subsets used for the COI tree, ddRADseq tree, morphology, SNAP+BFD*, and Structure analyses. Three non-*sigillata* samples (two *E. rubriceps* and one *E. storthynx*) used as outgroup for the ddRAD tree were not included in this table.

CAT. NUM	individual_id n=102	COI_tree n=53	ddRAD_tree n=101	Morphology n=59	SNAP+BFD* n=29	Structure n=27	Country	Locality	Latitude	Longitude
UW 204159	VMP005	Y	Y	Y	N	N	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 204203	VMP050	N	Y	N	N	N	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 204201	VMP048	Y	Y	Y	N	N	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 204416	VMP279	Y	Y	Y	N	N	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 204300	VMP149	Y	Y	Y	N	N	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 204519	VMP383	Y	Y	Y	N	N	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 204523	VMP387	Y	Y	Y	N	N	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 204417	VMP280	Y	Y	Y	N	N	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 204164	VMP010	Y	Y	Y	N	N	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 204233	VMP081	Y	Y	Y	N	N	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 205286	CRF789	N	Y	N	N	N	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 205225	CRF728	N	Y	N	N	N	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 205198	CRF701	N	Y	N	N	N	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 205285	CRF788	N	Y	N	N	N	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 205284	CRF787	N	Y	N	N	N	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 205049	CRF050	N	Y	N	N	N	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 204989	CRF166	N	Y	N	N	N	Tonga	Vava'u_Utungake	-18.665179	-174.016049
forthcoming	ESH19	Y	Y	Y	N	N	Maldives	Laamu_Atoll	1.808733	73.351583
forthcoming	ESH17	N	Y	Y	N	N	Maldives	Kaafa-haa,_Laamu_Atoll	1.808733	73.351583
forthcoming	ESH20	N	Y	Y	N	N	Maldives	Haleema_Haa_Laamu_Atoll	1.833517	73.378233
SAIAB_77969_6	SAIAB_77969_6	N	Y	Y	N	N	Seychelles	Ile_Conception	13.503	-4.6641
SAIAB_77969_7	SAIAB_77969_7	N	Y	Y	N	N	Seychelles	Ile_Conception	14.006	-4.6641
forthcoming	ESH1	Y	Y	Y	N	N	Indonesia	Rinca	-8.628217	119.708317
forthcoming	ESH2	Y	Y	Y	N	N	Indonesia	Rinca	-8.628217	119.708317
forthcoming	SI5	Y	Y	Y	N	N	Indonesia	Cendrawasih	4.856633	114.271617
forthcoming	SI6	Y	Y	N	N	N	Indonesia	Cendrawasih	4.856633	114.271617
forthcoming	EPR1	Y	Y	N	N	N	Indonesia	Lenbeh Strait	1.503611	125.3925
forthcoming	ESH4	Y	Y	Y	N	N	Brunei	Chearnley	-4.856633	114.271617
forthcoming	ESH14	Y	Y	Y	N	N	Indonesia	Pisang_Islands	-2.677833	131.610167
forthcoming	ESH12	Y	Y	Y	N	N	Timor Leste	Beoli_Atauro_Island	-8.242883	125.611517
forthcoming	ESH9	Y	Y	Y	N	N	Indonesia	Kawe_Bay	-0.092067	130.125667
forthcoming	ESH3	Y	Y	Y	N	N	Brunei	Chearnley	-4.856633	114.271617
forthcoming	ESH5	Y	Y	Y	N	N	Brunei	Chearnley	-4.856633	114.271617
forthcoming	ESH10	Y	Y	Y	N	N	Indonesia	Api_Bay	-8.53255	119.650917
forthcoming	ES3	Y	Y	Y	N	N	Indonesia	Komodo	-8.53255	119.650917
forthcoming	ESH21(19)	N	Y	Y	Y	N	Indonesia	Pisang Islands_ Fakfak	-2.4382	131.622783
forthcoming	ESH25	N	Y	Y	N	N	Indonesia	North_Misool	-1.6226	129.925
forthcoming	AC2	Y	Y	Y	N	N	Papua New Guinea	North_Sulawesi	-10.59611	150.766944
forthcoming	AC3	Y	Y	Y	N	N	Solomons	Yadina	-9.083889	159.205833
forthcoming	ESH6	Y	Y	Y	N	N	Indonesia	Lembeh_Strait_N_Sulawesi	1.425833	125.185556
forthcoming	EWS1	Y	Y	Y	N	N	Indonesia	Raja_Ampat	0.501667	131.131389
forthcoming	EWS2	Y	Y	Y	N	N	Indonesia	Raja_Ampat	0.51	131.1255
forthcoming	42(GA-32-09-F01)	Y	N	N	N	N	Pohnpei	-	6.991758	158.241767
forthcoming	SI12	Y	Y	Y	N	N	Indonesia	Cendrawasih	-3.208483	135.083683
forthcoming	SI13	Y	Y	Y	N	N	Indonesia	Cendrawasih	-3.208483	135.083683
forthcoming	ESH7	Y	Y	Y	N	N	Indonesia	Banda_Harbor	-4.513233	129.89315
forthcoming	ESH8	Y	Y	Y	N	N	Indonesia	Banda_Harbor	-4.513233	129.89315
forthcoming	ESH13	Y	Y	Y	N	N	Indonesia	Pisang_Islands	-2.677833	131.610167
forthcoming	ESH21(18)	Y	Y	Y	N	N	Indonesia	Raja_Ampat	-0.795067	130.49805
forthcoming	FB7	Y	Y	Y	Y	N	Solomons	Lingatu	-9.1216	159.1562
forthcoming	SI1	Y	Y	Y	N	N	Western Samoa	Manono	-13.258611	-172.112222
forthcoming	SI2	Y	Y	Y	Y	Y	Western Samoa	Malitanua_wreck	NA	NA
forthcoming	SI3	Y	Y	Y	Y	Y	Western Samoa	Malitanua_wreck	NA	NA
forthcoming	SI4	Y	Y	Y	N	N	Western Samoa	Faleolo	-13.824444	-172.018333
UW 202654	SAM19120	N	Y	N	Y	Y	American Samoa	Utulei	-14.280656	-170.681709
UW 202654	SAM19128	N	Y	N	Y	Y	American Samoa	Utulei	-14.280656	-170.681709
UW 202654	SAM19129	N	Y	N	Y	Y	American Samoa	Utulei	-14.280656	-170.681709
forthcoming	SI10	Y	Y	Y	N	N	Fiji	Moala_Lau_Group	-18.537567	179.928683
forthcoming	SI14	Y	Y	Y	N	N	Fiji	Yagasa_Lagoon_Lau_Group	-19.19633	178.47425
forthcoming	SI7	Y	Y	Y	Y	Y	Fiji	Moala_Lau_Group	-18.537567	179.928683
forthcoming	SI11	Y	Y	Y	Y	Y	Fiji	Matuku_Lagoon_Lau_Group	-19.151917	179.745533
UW 204271	VMP120	Y	Y	Y	Y	Y	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 204197	VMP044	Y	Y	Y	N	N	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 204315	VMP164	Y	Y	Y	N	N	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 204198	VMP045	Y	Y	Y	Y	Y	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 204168	VMP014	N	Y	N	Y	Y	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 204169	VMP015	Y	Y	Y	Y	Y	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 204261	VMP110	Y	Y	Y	Y	Y	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 204167	VMP013	Y	Y	Y	Y	Y	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 204171	VMP017	Y	Y	Y	Y	Y	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 204224	VMP072	N	Y	Y	Y	Y	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 204239	VMP087	Y	Y	Y	N	N	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 204262	VMP111	Y	Y	Y	Y	Y	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 204467	VMP329	N	Y	Y	Y	Y	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 204468	VMP330	N	Y	N	N	N	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 204941	CRF111	N	Y	N	Y	Y	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 204262	VMP111B	N	Y	N	N	N	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 205200	CRF703	N	Y	N	Y	Y	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 205227	CRF730	N	Y	N	N	N	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 204934	CRF131	N	Y	N	Y	Y	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 205222	CRF725	N	Y	N	N	N	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 205236	CRF739	N	Y	N	Y	Y	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 204892	CRF200	N	Y	N	Y	Y	Tonga	Vava'u_Euakafa_Island	-18.753733	-174.035278
UW 204798	CRF500	N	Y	N	Y	Y	Tonga	Vava'u_Euakafa_Island	-18.753733	-174.035278
UW 205309	CRF812	N	Y	N	N	N	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 204917	CRF584	N	Y	N	N	N	Tonga	Vava'u_Euakafa_Island	-18.753733	-174.035278
UW 204805	CRF428	N	Y	N	N	N	Tonga	Vava'u_Euakafa_Island	-18.753733	-174.035278
UW 204806	CRF447	N	Y	N	N	N	Tonga	Vava'u_Euakafa_Island	-18.753733	-174.035278
UW 205302	CRF805	N	Y	N	N	N	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 205123	CRF019	N	Y	N	N	N	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 204898	CRF453	N	Y	N	N	N	Tonga	Vava'u_Euakafa_Island	-18.753733	-174.035278
UW 204801	CRF441	N	Y	N	N	N	Tonga	Vava'u_Euakafa_Island	-18.753733	-174.035278
UW 205306	CRF809	N	Y	N	N	N	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 204893	CRF068	N	Y	N	N	N	Tonga	Vava'u_Euakafa_Island	-18.753733	-174.035278
UW 204707	CRF219	N	Y	N	Y	Y	Tonga	Vava'u_Lotuma	-18.661926	-174.008925
UW 205292	CRF795	N	Y	N	Y	Y	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 205221	CRF724	N	Y	N	N	N	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 204939	CRF150	N	Y	N	Y	Y	Tonga	Vava'u_Utungake	-18.665179	-174.016049
UW 205291	CRF794	N	Y	N	N	N	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 205308	CRF811	N	Y	N	N	N	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 205266	CRF769	N	Y	N	N	N	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588
UW 205224	CRF727	N	Y	N	Y	Y	Tonga	Vava'u_Afo_Island	-18.708747	-173.996588

Table S2. Population genetics metrics for the subgroup of samples from Tonga, Fiji, and Samoa analyzed with the program Structure 2.3.4. Calculations were done with the program module Populations in the Stacks 2.4.1. Program.

Lineage	N	Heterozygosity (observed)	Heterozygosity (expected)	π	F_{is}	F_{st} (vs. Samoa)	F_{st} (vs. Fiji)	F_{st} (vs. Tonga)
Samoa	5	0.111	0.148	0.167	0.119	-	0.255	0.111
Fiji	2	0.115	0.097	0.145	0.045	0.255	-	0.056
Tonga	20	0.114	0.21	0.216	0.369	0.111	0.056	-

Chapter 3

Phylogenomic analysis of the dwarfgobies (Teleostei: Gobiidae: *Eviota* and *Sueviota*)

Marta C. Gómez-Buckley[1], Carmen Pedraza Marrón[2], Diego Vaz[3], Christopher Goatley[4], Mark V. Erdmann[5,6], David Greenfield[7], and Luke M. Tornabene[1]

[1] School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA USA

[2] Department of Life and Environmental Sciences, University of California, Merced, CA USA

[3]Natural History Museum, London, UK

[4]University of Southampton, UK

[5] Conservation International Aotearoa, University of Auckland, New Zealand

[6] California Academy of Sciences, San Francisco, CA USA

ABSTRACT

The goby genera *Eviota* and *Sueviota* (family Gobiidae) are commonly known as dwarfgobies, and collectively the two genera are among the most abundant and diverse groups of fishes on coral reefs. Despite the diversity, abundance, and ecological importance of this group, and the large and growing number of species described to date (132 *Eviota*, 9 *Sueviota)*, there is a lack of understanding of the phylogenetic relationships among the major clades of *Eviota,* poor knowledge of the relationships between dwarfgobies and other Gobiidae species, and no information on the placement of *Sueviota.* In addition, as is the case with most small reef fishes, a clear understanding of taxonomically informative phenotypic characters is also lacking. To resolve the evolutionary history of dwarfgobies, we inferred a time-calibrated phylogeny of the group using genome-wide data from 440 ddRADseq loci captured across 98 *Eviota* and *Sueviota* specimens, plus 66 specimens of other related gobies. We also assessed the distribution of 14 external and osteological morphological characters across the tree to assess which may be useful for diagnosing clades. Our results robustly established the non-monophyly of the genus *Eviota,* which was resolved into two separate clades, both of which were resolved within a lineage of other coral associated genera (*Gobiodon*, *Paragobiodon*, *Pleurosicya*, *Minysicya* and *Bryaninops*). One of the two clades is herein elevated to its own genus, *Eviotops*, a name which was previously considered synonymous with *Eviota*. Additionally, we established that the genus *Sueviota* is deeply nested within one of the *Eviota* clades and is herein synonymized with *Eviota.* We also found strong phylogenetic signal for 12 out of the 14 phenotypic traits examined, providing strong complementary support for the two recovered clades and establishing the validity of phenotypic traits that strongly correspond with genetic groupings that should aid in future taxonomic studies for this group.

INTRODUCTION

The goby genera *Eviota* and *Sueviota* (Jenkins 1903; Winterbottom and Hoese 1988) are commonly known as dwarfgobies (Greenfield 2017). Dwarfgobies (family Gobiidae) are extremely small fishes (< 2mm) that live association with a variety of habitats on coral reefs including live corals, dead coral, gravel, and sand (Tornabene et al. 2013a; Ahmadia et al. 2018). They have very short life cycles (<100 days, ~28 as pelagic larvae) which translates into very rapid turnover rates (Depczynski and Bellwood 2006). *Eviota* and *Sueviota* are superficially very similar in morphology, coloration, and ecology and are often difficult to differentiate *in situ*. *Eviota* (Jenkins 1903) is one of the most specious marine genera in tropical coral reef ecosystems

(Greenfield 2017). To date there are 132 described species of *Eviota* (Greenfield et al. 2024) and nine species of *Sueviota* (Peinemann et al. in review), with many additional new species of both genera being described each year. Along with the genus *Trimma* (Winterbottom et al. 2020), species of dwarfgobies are among the most abundant and diverse cryptobenthic fishes on reefs (Ahmadia et al. 2012; Atta et al. 2019; Gómez-Buckley et al. 2023), and together with other cryptobenthic reef fishes, they contribute substantially to the overall trophodynamics of coral reefs ecosystems throughout the tropical Indo-Pacific region (Brandl et al. 2018; Brandl et al. 2019).

Owing to their high species richness, abundance, and ecological importance, there has been a growing number of papers examining the systematics and evolution of dwarfgobies over the last two decades. Recent taxonomic descriptions of new species for the genus *Eviota* have included molecular analyses to infer phylogenetic relationships and delimit species boundaries (e.g. Greenfield et al. 2018; Tornabene et al. 2021). These studies typically focus their phylogenetic analyses on a single species complex and rely on a single mitochondrial marker (COI) or a combination of COI and the nuclear gene Ptr to examine taxonomic boundaries and geographic speciation patterns. A few studies have included broader taxonomic sampling, including Hanahara (2023), who studied the phylogenetic relationships of 11 *Eviota* species complexes around Okinawa Island (Japan) using COI, and Tornabene et al. (2013) and Ahmadia et al. (2018), who looked at habitat association across the genus using COI and Ptr. However, despite the ecological importance of this group as substantial food source for larger predators in coral reefs (Depczynski and Bellwood 2003; Brandl, et al. 2019), and the large and growing number of species described to date, there is a lack of understanding of the phylogenetic relationships among the major clades of *Eviota*, and no studies to date have incorporated information on the placement of *Sueviota*. Moreover, no studies on *Eviota* have included more than two loci. Thus, the evolutionary history of one of the most diverse, rapidly speciating, and ecologically important groups of reef fishes in the world remains poorly known (Greenfield 2017; Brandl et al. 2018).

The phylogenetic relationships within dwarfgobies have been more intensely studied than their position within the Gobiidae. Ahmadia et al. (2018) performed a phylogenetic analysis of 43 species of *Eviota* using COI and Ptr and confirmed two monophyletic groups within the genus; one contained species with branched pectoral fin rays, while the other group had unbranched pectoral fin rays. The monophyly of these two clades was also found in prior and subsequent studies, which each focused on specific species complexes and were solely based on two loci (Tornabene et al. 2013a, 2015, 2016, 2021; Tornabene and Greenfield 2016; Greenfield and Tornabene 2014; Greenfield et al. 2018; Erdmann et al. 2023). All these studies featured extremely limited sampling outside of *Eviota* (but see Harefa and Chen 2024 with just three *Eviota* species in their tree), and thus the monophyly of the genus was not rigorously tested.

The first study that incorporated some dwarfgobies in their analysis of the family Gobiidae (Thacker and Roje 2011) used just one species (*E. afelei*). This study showed *E. afelei* being sister to the genera *Gobiodon*, *Paragobiodon*, and *Bryaninops* – the latter genera, unlike dwarfgobies, all have highly specialized morphologies that are adapted for living in close symbiosis with different types of corals and sponges. As such, Thacker and Roje (2011) referred to this collective group as the Coral Gobies. Tornabene et al. (2013a) generated a 75 species phylogeny that included *E. albolineata* and other two Coral Goby genera (*Gobiodon* and *Pleurosicya*) that again were sister to *Eviota*. In another goby phylogeny (Agorretta et al. 2013) the one *Eviota* representative they used (*E. abax*) was sister species with *Bryaninops,* and the two were sister to *Gobiodon*, with the collective clade being referred to as "the *Gobiodon*-lineage". Later Thacker (2015) corroborated again the same placement of *Eviota* and relationship to other Coral Gobies in their time calibrated tree. The most recent Gobiidae phylogeny was generated by McCraney et al. (2020). They used public sequences of 827 taxa (65 of them in the "*Gobiodon*-lineage" and from that, 45 were *Eviota),* but those *Eviota* were represented in the matrix by only one gene (Ptr). The resolution within the Coral Gobies in this study was very low, and support for a monophyletic *Eviota* was equivocal. Specifically, there was little support for the interrelationships of *Gobiodon, Pleurosicya, Bryaninops*, and *Eviota*. Like previous studies, no *Sueviota* species were included in McCraney et al. (2020). Thus, much like the relationships within dwarfgobies, the relationships between dwarfgobies, other Coral Gobies, and the rest of >1500 species of Gobiidae are also not well understood.

The genus *Sueviota* (Winterbottom and Hoese 1988) is morphologically very similar to *Eviota*. One of the main differences is the presence of a branched fifth pelvic-fin ray (but see below), which for *Eviota* tends to be absent, or if present, it is smaller than the fourth ray and unbranched. A second distinguishing character is that for most species of *Sueviota*, the fifth pelvic rays of each fin are joined by a well-developed membrane forming a complete pelvic disk, while species of *Eviota* typically lack this membrane. Five species of *Sueviota* have been described since the recognition of the genus, bringing the total number of *Sueviota* species to nine (Pienemann et al. in review). However, some of the later described species were missing some of the above-mentioned diagnostic characters (Peinemann et al. in review). At the time of the genus' description (which included four *Sueviota* species), Winterbottom and Hoese (1988) were uncertain about the monophyly of *Sueviota*, or whether it was separate from or nested within *Eviota*. The status of this genus continues to be tenuous (Allen et al. 2016; Allen and Erdmann 2017; Greenfield and Randall 2017; Greenfield et al. 2019). To our knowledge, the *Sueviota* has not been included in molecular phylogenies, adding to the uncertainty regarding the monophyly of the genus, their relationship to other dwafgobies, and other Gobiidae more broadly.

Recently, genome-wide data from restriction site-associated DNA sequencing (RADseq, Andrews et al. 2016) has been successfully used for clarifying phylogenetic relationships among closely related species of dwarfgobies (Gómez-Buckley et al. in review; Chapter 2 of this dissertation). Specifically, 4,591 loci from RADseq data were able to fully resolve relationships within the *Eviota sigillata* species complex and identify the presence of at least 7 undescribed species. Other studies have successfully used RADseq data to resolve phylogenetic relationships of fishes at deeper levels, including phylogenetic analyses of snailfishes (Orr et al. 2019) and surfperches (Longo et al. 2015). This approach could therefore prove useful for resolving the phylogenetic relationship between *Eviota* and *Sueviota* and resolve if the genus *Eviota* is monophyletic or if a taxonomic revision is in order.

Genome-wide data may prove useful for resolving the evolutionary history of dwarfgobies. However, the vast majority of specimens of *Eviota* and *Sueviota* in collections lack genetic data, and the use of external phenotypic characters to diagnose genera, groups within genera, and species continue to be critical in fish systematics (Fujiwara et al. 2020; Greenfield and Erdman 2021; Greenfield et al. 2024). Furthermore, many of these phenotypic characters have not explicitly been tested for true phylogenetic signal, and thus may not be diagnosing actual monophyletic groups. Dwarfgoby morphology is relatively conserved, and there is a need to find better diagnostic characters that can help us identify the more than 132 species in the genus. The previous morphology-based studies have used a limited suite of available characters to organize them into groups. The genus-wide taxonomic keys available (Greenfield and Winterbottom 2016; Greenfield 2021), rely on the morphological character states described by the earlier taxonomic work for the genus by (Lachner and Karnella 1978, 1980; Jewett and Lachner 1983), including fin ray counts of the second dorsal and anal fins, the presence or absence of branched pectoral fin rays, the length of the pelvic-fin ray relative to the fourth, and the morphology of the male urogenital papilla. The arrangement pattern of the sensory pores on the head is also used in the keys and is a helpful diagnostic character at the species level (Greenfield 2017). Internal osteological characters, such as vertebral counts and the presence of tricuspid teeth in some species (Greenfield et al. 2018), can be informative when discerning different species or groupings. However, these characters have not been extensively used, mostly due to the difficulty in collecting this information. A detailed examination of dwarfgoby osteology by Vaz et al. (in review) revealed characters with potential for differentiating various groups of dwarfgobies and distinguishing them from closely related genera of Coral Gobies. Specifically, they found that the ligament connecting the maxilla to the preopercle was absent in all the *Sueviota* and *Eviota* specimens and present in all the other non-*Eviota*/*Sueviota* Coral Gobies examined. They also found that the posterior margin of the interopercle bone has a deep notch or wrench-like shape for the examined *Sueviota* and branched *Eviota* specimens. In contrast, the unbranched *Eviota* specimens exhibited a shallow notch, along with *Pleurosicya, Bryaninops,* and *Paragobiodon*. *Gobiodon* was the only genus with a straight-edge interopercle. The only study that was able to compare an extremely limited set of morphological characters to genetic groupings for a small number of *Eviota* taxa (20 spp.) found that branched versus non- branched pectoral fin rays perfectly aligned with the two clades generated from their molecular phylogeny, and that sensory pore patterns were homoplasious (Tornabene et al. 2013a). A more extensive examination of the relationship between dwarfgoby morphology and phylogeny may help reveal which phenotypic characters are informative in diagnosing monophyletic groups and informing classification.

Overall, since the description of the first *Eviota* species in 1871 to date, we still do not have a clear understanding of the phylogenetic relationships among the different groups within dwarfgobies, nor do we know whether the genera are monophyletic. This study is the first to use genome-wide data to clarify the phylogenetic relationships among members of both genera of dwarfgobies. Specifically, we use double digest restriction site associated (ddRADseq) data to examine the relationship between 170 samples from 64 *Eviota* species, 3 species of *Sueviota* and several genera of Coral Gobies and other gobiid genera. We specifically aim to answer (1): is the genus *Eviota* monophyletic with respect to other gobiid genera, specifically those of the Coral Gobies or the "*Gobiodon*-lineage" *sensu* Agorretta et al (2013); (2) what is the placement of the genus *Sueviota* within gobiid phylogeny; and (3) is there concordance between clades on the RAD tree and the distributions of potentially informative important morphological characters to assess the extent of phylogenetic signal or homoplasy in dwarfgoby morphology.

METHODS

Taxon sampling and DNA extraction

All the samples used in this study are part of a growing collection of *Eviota* voucher specimens housed at the California Academy of Sciences (CAS) and the University of Washington Burke Museum of Natural History and Culture UW Fish Collection (Table S1). The *Eviota* and *Sueviota* samples were collected using either clove oil or quinate solution at depths ranging from ~3-30 meters throughout the Indo-Pacific region. Many specimens were photographed in situ prior to collection, and all others had pictures of dead specimens taken immediately after capture and prior to preservation in 90-95% ethanol. All *Eviota and Sueviota* specimens were identified using Greenfield and Winterbottom (2016) and Peinemann et al. (in review) keys. The other Coral Gobies and the rest of the outgroup specimens were identified using the primary literature that presented those species descriptions

We selected taxa by prioritizing: (i) representatives of all species complexes across *Eviota* and *Sueviota* that have been identified by past research (e.g. the *E. zebrina, E. punctulata,* and *E. sigillata* complexes, etc. Table 1); (ii) species with unambiguous identifications based on published descriptions (Table 1); (iii) samples collected near the type locality of that species whenever possible. A total of 34 taxa were included in the analyses from other Coral Goby genera (*Gobiodon, Paragobiodon*, *Pleurosicya*, *Minysicya*, and *Bryaninops*) and from other gobioid genera (Table S1).

Molecular Methods

We extracted DNA from pectoral fin clips or muscle samples of 182 individuals representing 95 species (Table 2) using QIAGEN DNEasy Blood and Tissue Kits (Qiagen, Inc.). We used ddRADseq (Peterson et al. 2012) for capturing SNP data from the samples used in this study. DNA extractions were sent to the NGS labs at the University of Wisconsin for sequencing and library preparation, and we follow the same protocols as in Gómez-Buckley et al. (in review; Chapter 2 in this dissertation) for these two steps. We used the Fastqc software to check for sequence data quality and because adapter content was high, we use Trimmomatic v0.39 (Bolger et al. 2014) to trim the excess adapter content. The following parameters were used: LEADING:3 (Low-quality bases from the start of the read were removed with a quality threshold of 3), TRAILING:3 (Low-quality bases from the end of the read were removed with a quality threshold of 3), SLIDINGWINDOW:4:15 (A sliding window approach was used to trim reads based on quality with a window size of 4 and a required average quality of 15), and MINLEN:100 (Reads shorter than 100 bases after trimming were discarded). These trimmed reads were then used for downstream analysis.

We used the Stacks 2.4.1 pipeline (Catchen et al. 2013) on the University of Washington's High Performance Computer Cluster to assemble SNPs from the adapter-cleaned dataset. We used the same protocols as in Gómez-Buckley et al. (in review; Chapter 2 in this dissertation) for demultiplexing of data via process_radtags and for parameter optimization using RADstackshelpR 0.10 (https://github.com/DevonDeRaad/RADstackshelpR). The parameters selected to assemble putative loci with the denovo_map_pl module in the program Stacks, were '-m=4' (minimum number of reads required to form a stack), '-M=8' (number of mismatches allowed between stacks within individuals), and '-n=9' (number of mismatches allowed between loci catalog building. The denovo_map.pl was run for the entire demultiplexed process_radtags output files and the resulting vcf file was run through the SNPfiltR bioinformatic pipeline (DeRaad 2021) which aided in selection of most complete samples to rerun through populations for the final csv files. The program vcf2phylip (Ortiz 2019) was used to convert the vcf file to phylip format to generate a PHYLIP alignment suitable for constructing phylogenetic trees.

Phylogenetic Analyses

To examine the effects of missing data on the support of phylogenies, we tested three values of the parameter '-R', the minimum percentage of individuals across the entire alignment required to include a locus, using the 'populations' module in Stacks. The values chosen were R=0.2, 0.3 and 0.4, corresponding to thresholds of 80%, 70%, and 60% missing data allowable, respectively. The number of loci (RADtags) and SNPs (variable sites) retained can be found in Table 3.

We generated maximum likelihood trees under the GTR+G for the three -R filtered datasets using RAxML v1.1 (Kozlov 2018) on the CIPRES (Miller et al. 2010) supercomputer portal. Support for nodes was assessed with 100 bootstrap replicates and trees were visualized using FigTree v1.4.4 (Rambaut 2012). The entire alignments including invariant sites were used for phylogenetic inference (Table 2). Heterozygote sites were coded using IUPAC ambiguity codes. For the R=0.4 dataset, we also inferred a tree using Bayesian Inference via the software MrBayes 3.2 (Ronquist et al. 2012) using the same substitution model. The MCMC analysis was run for 10 million generations, sampling every 1000 generations, discarding the first 20 percent of trees as burn-in. Convergence and mixing of MCMC chains were assessed in Tracer 1.5 (Rambaut and

Drummond 2018). A 50% majority-rule consensus tree was generated from the resulting posterior distribution of trees.

Time Calibrated Tree

A time-calibrated tree was obtained for the dataset with the least missing genetic data (R=0.4) using BEAST 2.2.3 (Bouckaert et al. 2014). We used an optimized relaxed-clock model and same parametrization for priors, partitioning, and MCMC scheme as in Tornabene et al. (2016, 2022). The three fossil calibration points used for this study are summarized in Table 3. We first set a minimum crown age of 50 Mya for Gobioidei based on the species †*Carlomonnius quasigobius* Bannikov and Carnevale 2016. Then we set the crown age of 19.1 Mya for the *Gobius* lineage (*sensu* Agorreta et al. 2013) based on the species on †*Gobius jarosi* Přikryl and Reichenbacher 2018 (Reichenbacher et al. 2018). Lastly, we set the crown age of Gobiidae + Butidae to a minimum of 28 Mya based on the species †*Pirskenius diatomaceous*, Obrhelová 1961.

Morphological character state sampling and concordance to phylogeny

To assess how morphological character states concurred with the time-calibrated phylogeny and to evaluate their phylogenetic signal, we used characters from taxonomic studies for species descriptions (Lachner and Karnella 1978, 1980) or for identification keys (Greenfield and Winterbottom 2016), and characters previously identified as being potentially for distinguishing genera or groups of genera (Tornabene et al. 2013a; Winterbottom and Hoese 1988; Vaz et al. in review). These characters (14 total) were: (1) pectoral fin rays branched or unbranched; (2-9) presence/absence of all the possible sensory pores found in this group (seven sets of pores examined, see Figure 9); (10) presence/absence of a well-developed membrane between the 5th rays of the two pelvic fins; (11) presence/absent of a frenum connecting the anterior spines of the pelvic fins, thus forming the typical pelvic disk seen in many gobies (Figure 1b); (12) total vertebrae counts; (13) the interopercle-retroarticular ligament connection; (14) and the shape of the posterior margin of the interopercle (Figure 5). We also collected information on these characters for members of Coral Goby genera (*Bryaninops*, *Pleurosicya*, *Minysicya*, *Gobiodon*, and *Paragobiodon*). Data on internal morphology were collected from specimens that were cleared and stained, x-rayed, or micro CT-scanned (see Vaz et al. in review for details) or from the literature (Larson 1987, 1990, 2002; Winterbottom and Harold 2005; Sato and Motomura 2024).

To assess the agreement between our phylogeny and each morphological character and examine their phylogenetic signal, we first used ancestral state reconstruction to map each character across the time-calibrated phylogeny. The distant outgroups from outside the Coral Gobies that were included as fossil calibration points were trimmed from for these analyses. Ancestral state was reconstructed using the ‘ace’ function in the *ape* package (Paradis and Schliep 2019) in R v. 4.3.2 (R Core Team 2021). Since some species lack data for some morphological characters, taxa with missing data were removed from the before to mapping characters onto the trees.

To determine if there was significant phylogenetic signal in each character, indicating their potentially usefulness for diagnosing species (synapomorphies), we compared the distribution of phenotypic characters on our time-calibrated tree to their distribution on a tree with all phylogenetic information removed (i.e. a complete polytomy, with characters thus being randomly distributed with respect to species relationships). We used a λ=0 transformation in the "rescale" function of the *phytools* package (Revell 2024) to remove phylogenetic information from our time-calibrated tree, and then fit the phenotypic characters to the tree using ancestral state reconstruction as we did for the original time-calibrated trees. Then we used a likelihood ratio test to compare the log-likelihood scores of the λ=0 tree to the original tree, generating a p-value to inform if there is a significant phylogenetic signal for each character. A significant p-value (alpha=0.05) indicates that a character's distribution across species is significantly driven by, or correlated with, the underlying phylogenetic relationships (versus being random with respect to phylogeny).

RESULTS

All the four phylogenetic trees obtained here, including with three maximum likelihood trees with levels of filtering (R= 0.2-0.4; Figs 1, Supplementary Files 1, 2) and the Bayesian inference tree (Supplementary File 3) produced near identical topologies, indicating the robustness of the evolutionary relationships despite missing data and the use of different analytical approaches. Among the RAxML trees, the only discrepancies appeared as slightly different placement of some taxa. For example, the two *E. cf. imitata* appeared as sister to the *E. sigillata*, *rubriceps* and *storthynx* in the R=0.4 tree (Figure 2), while in the R=0.3 tree (Supplementary Files 1) they branch of as a separate subclade sister to the previously mentioned taxa subclades and three other subclades with less bootstrap support.

All the resulting phylogenetic trees robustly established the non-monophyly of the genus *Eviota*. Two distinct and well supported (bootstrap = 100%; posterior probability = 1.0) clades containing species of *Eviota* were recovered in the RAxML tree generated with 164 taxa and 440 loci, and R=0.4 (Figure 2). One clade, which contained all *Eviota* with branched pectoral fins and all *Sueviota* samples, was sister to the other genera in the Coral Goby group (*Gobiodon, Pleurosicya*, *Minysicya*, and *Bryaninops*). This combined clade was sister to the second *Eviota* clade, which contained exclusively specimens with unbranched pectoral fin rays. Within the non-dwarfgoby Coral Gobies in our tree, there were two well supported subclades, with one containing the genera *Gobiodon* and *Paragobiodon*, and another containing *Bryaninops*, *Pleurosicya*, and *Minysicya.* The *Sueviota* specimens were deeply nested in the clade of *Eviota* with branched pectoral fin rays. The four *Sueviota lachneri* specimens clustered together in a single subclade, while the *Sueviota atrinasa* was nested within a subclade with other *Eviota* species like *E. sparsa*, *E. jewettae*, *E. cf. pinocchioi*, and *E. pseudaprica* (Figure 2).

The calibrated tree (Figure 3), which has an identical topology to the maximum likelihood trees, shows the mean crown age of the gobioids at 62.36 Mya. The Gobiidae had a mean age of 34.31 Mya, and the Coral gobies 28.25 Mya. Within the Coral Gobies, both *Eviota* clades seem to have originated at around 20.40-20.92 Mya, with the non-*Eviota/Sueviota* Coral gobies showing an earlier age of 23.47 Mya.

Our analysis of character evolution shows that most characters we examined had significant levels of phylogenetic signal (Table 4, Figures 4-10, S1-S7), as indicated by their significant better fit to our time-calibrated tree than to the $\lambda=0$ transformed tree with no phylogenetic information. For the sensory pores, the SOT and AOT pores (Figures S4 and S5) were the only characters with no significant phylogenetic signal due to both being present in all the specimens in our phylogeny except for *E. jewettae* (with all pores absent), and *E. pseudaprica and pinocchioi* (within the subclade containing the Sueviota) with SOT present and AOT absent. The IT pore (Figure 9) was absent for the entire unbranched clade and for three subclades within the branched clade. The NA pores (Figure S1) were present in most taxa except for three subclades in the unbranched clade and two other taxa in the branched clade. The POP pores (Figure 10) were present in most of the taxa except for the small subclade containing the *S. atrinasa, E. sparsa, E. pinocchioi,* and *E. jewettae*. *Pleurocysia, Minysicya,* and *Bryaninops*, had three POP instead of the usual two common in all other taxa in our tree. The AITO pore (Figure S2) was present in all but *E. jewettae* and *E. pinocchioi*. Two subclades within the unbranched clade had an enlarged AITO pore. *E. shimadai* had a unique interorbital pore (a joint AITO and PITO), and *Bryaninops* had paired AITO pores. Finally, the PITO pore (Figure S3) was absent in three subclades within the unbranched clade but present in all branched clade, except for *E. jewettae* and *pinocchioi*. Additionally, *E. shimadai* two joined pores and *Bryaninops* had a paired PITO pore.

All external characters of the pelvic fin membrane (Figure S6) and frenum (Figure 8), as well as the branching of pectoral and 5th pelvic fin rays (Figure 5) presented a strong phylogenetic signal (Table 4). The pectoral fin ray branching (Figure 7) clearly separated the two main clades of *Eviota*, with the latter clade including the non-*Eviota* Coral Gobies. The presence/absence of pelvic frenum was also very diagnostic, with this character being absent in all *Eviota* species (Figure 8). The 5th pelvic fin branching or unbranching, along with the presence or absence of a membrane between pelvic fin rays, were present/unbranched for Coral Gobies and some of the taxa in the *Sueviota* subclade (Figure S7).

All internal characters examined, such as vertebral count, presence or absence of interopercle-retroarticular ligament (Figure 6) and the wrenched shape of interopercle (Figure 5) also presented a strong phylogenetic signal (Table 4). All the taxa in the unbranched clade exclusively had 25 vertebrae while the branched clade had 26, except for *S. atrinasa* with 27 (Figure 4). All the taxa in the unbranched clade had a shallow notch in the interopercle posterior edge, while the rest of the *Eviota* in the branched clade presented a deep wrenched shape interopercle. *Gobiodon quinquestegatus* presented a shallow notch like the unbranched clade, while the *Paragobiodon echinocephalus* was the only taxa in this dataset with a straight edge or no wrench shape to the interopercle. The last internal character, the presence or absence of the ligament, was absent in all the *Eviota* taxa in both clades, with *P. echinocephalus* and the *G. quinquestrigatus* presenting this ligament.

DISCUSSION

The goals of this study were to answer: (1) If the genus *Eviota* was monophyletic with respect to other gobiid genera, namely those of the Coral Gobies or the "*Gobiodon*-lineage"; (2) The placement of the genus *Sueviota* within gobiid phylogeny; and (3) if there was concordance between clades on our tree and the distributions of potentially informative important morphological characters to assess the extent of phylogenetic signal or homoplasy in dwarfgoby morphology. Our results here provide the most detailed look at the phylogenetic relationships among members of both genera of dwarfgobies and their relatives. Information from our analyses allow us to reject the monophyly of the genus *Eviota* with respect to other gobiid genera, namely those of the Coral Gobies or the *Gobiodon*-lineage *sensu* Agorretta et al. (2013)

and confirm the placement of the genus *Sueviota* within gobiid phylogeny as being nested within *Eviota*. Finally, we assessed the extent of phylogenetic signal or homoplasy in dwarfgoby morphology by analyzing the concordance between clades on our tree and the distributions of potentially informative important internal and external morphological phenotypic traits. These data collectively allow us to revise the taxonomic classification of the group.

Phylogenetic relationships among the Eviota and close relatives, and non-monophyly of the genus

The Coral Gobies were first established as such by Thacker and Roje (2011). In that study, the only specimen of *E. afelei* in their phylogeny was sister to the *Gobiodon, Paragobiodon*, and *Bryaninops* and they were collectively named Coral Gobies. Subsequent phylogenetic studies have continued to view *Eviota* as monophyletic and sister to the other genera within the Coral Goby. The inclusion of a broad and distantly related set of outgroups and the use of a genome wide dataset for our analyses provided a robust approach for testing the non-monophyly of the *Eviota*. Our results demonstrated the non-monophyly of the genus *Eviota* and, consequently, the need to reclassify some of the group members. The two non-monophyletic clades of *Eviota* identified in this study, corresponding to the clades with branched versus unbranched pectoral fin rays, were previously recognized in the two-gene tree phylogenies of Tornabene et al. (2013a) and subsequent studies (Tornabene et al. 2015; Tornabene et al. 2016, Tornabene and Greenfield 2016; Ahmadia et al. 2018; Tornabene et al. 2021). This finding has expanded our knowledge of the breadth of diversity in the Coral Goby group.

The two clades of *Eviota* recovered here have species that differ slightly in their overall ecology. In general, the branched-pectoral *Eviota,* which are sister to the non-*Eviota* Coral Gobies we included in our phylogeny (*Gobiodon, Paragobiodon, Bryaninops, Pleurosicya,* and *Minysicya*), exhibit highly cryptic behavior. They are often strictly benthic and live deep within the reef matrix of the reef (authors pers. obs). This behavior is similar to that of their sister group and reflects their close physical ties to specific benthic habitats. *Gobiodon* and *Paragobiodon* are known to be coral-specific symbionts (Munday 2001) that use pelvic disks to perch on the corals, hiding within branches. The same can be said about some species of *Pleurosicya* and *Minysicya* (Larson 1990, 2002) which are found perching on soft corals, and some *Bryaninops* that are found in obligate association with sea whip corals (Larson 1985). In contrast, the unbranched *Eviota* clade, which includes both habitat generalists and specialists (see Tornabene et al. 2013a), tends to be less cryptic and is more conspicuous on reefs. For example, species in the *E. nigriventris* and *bifasciata groups*, as well as *E. atriventris, E. lachedeberei*, and other members of this clade have been observed hovering above the substrate, sometimes in groups. This concurrence of distinct ecological and behavioral traits in *Eviota* supports the distinctiveness of the two non-monophyletic clades.

Placement of the genus Sueviota within gobiid phylogeny

The genus *Sueviota* was erected by Winterbottom and Hoese (1988) based on having distinct morphological features, the most obvious difference being the presence of a 5th pelvic fin that was long and branched in contrast with the closely related *Eviota* not having this ray, or if present, being very short and unbranched. Along with the genus, the authors described four new species and at the time they speculated that there could be two sister groups in the genus. They also were uncertain about the relationships of *Sueviota* with the rest of the *Eviota* and if *Sueviota* could be the sister group to all the *Eviota* species or just some. The limited number of specimens of this genus available for molecular analyses has made it even more challenging to test its validity and phylogenetic placement within the Coral Gobies in general. To the best of our knowledge, our study is the first to include specimens of *Sueviota* in a phylogenetic tree. The placement of these specimens in our phylogeny as being tree nested within the branched pectoral fin ray clade makes it clear that despite the noted distinct morphological features that set them aside from the *Eviota*, they are genetically part of this clade and thus the genus should be synonymized.

Morphological character state and concordance to phylogeny

Except for the pectoral fin-ray branching trait, which had been confirmed with molecular information in Tornabene et al. (2013a), no other studies had undertaken the analysis of phylogenetic and character state reconstruction analysis for our chosen phenotypic traits and at the level of detail accomplished in the present study. Unveiling such high levels of phylogenetic signal from most of the traits analyzed was surprising, especially for the sensory pores, which were not previously considered reliable characters beyond diagnosing individual species (Tornabene et al. 2013a).

Our approach to using each of the sensory pores as separate characters, rather than mapping them as collective “pore pattern groups”, as in previous studies (Lachner and Karnella 1980; Tornabene et al. 2013a) allowed us to obtain a clearer picture of their ancestral reconstruction and phylogenetic signal. Specifically, Tornabene et al. (2013a) mapped pore pattern groups onto a calibrated phylogenetic tree with 20 taxa, finding sensory pore pattern to be highly homoplasious and with no significant phylogenetic signal. On our phylogeny, we had examples of pores (e.g. IT pore) that were absent for large clades (e.g. the unbranched clade) but were consistent present in others. Other pores, like the interorbital AITO and PITO, presented interesting distribution patterns throughout the phylogeny. For instance, AITO had five different character states with most taxa having the single pore present and very few having lost the pore altogether. Interestingly, the presence of an enlarged pore evolved several times within the larger unbranched clade. Juvenile dwarfgobies often have incompletely developed sensory canals and pores. The presence of enlarged or fused pores, or partial canals, may be due to pauses in development and neotenic retention of juvenile conditions. However, it is difficult to pinpoint specific eco-evolutionary mechanisms that would favor convergent conditions in pore patterns. Other pores were less taxonomically useful, like AOT and SOT, which were largely invariant with two exceptions (*E. pseudaprica* and *E. jewettae*). The conservation of these two pores throughout the two *Eviota* clades and the Coral Gobies outgroup points to the deep ancestral state of this character possibly shared by more distantly related gobies.

The availability of osteological characters that had never been used before in combination with our genomic information proved very useful for our phylogenetic signal analysis. The total vertebrae count, the shape of the preopercle bone, and the presence or absence of a ligament between the retroarticular and preopercle bones (Figures 4, 5, and 6) all exhibited. a strong phylogenetic signal. This demonstrates that these traits could be useful in future phylogenetic analyses and ancestral state reconstruction not only for other *Eviota* specimens not included in our study, but also for other gobiid groups. The functional significance of these characters is not well understood, and our combined morphological-phylogenomic analysis could provide a framework for further research on the eco-morphology of the group.

Taxonomic recommendations

Our study has concluded that there are currently two polyphyletic groups sharing the same generic name *Eviota*. The type species (first species assigned to the genus) is *E. abax* (Jordan and Snyder 1901). Although we did not include that species in our analyses, it has branched pectoral rays, and is therefore considered part of that clade. Consequently, the branched pectoral fin clade, which includes all members of the genus *Sueviota* (hereafter considered a synonym of *Eviota,* see below), should retain the *Eviota* genus name based on the principle of priority. Consequently, the unbranched *Eviota* clade must be renamed. Rather than using a new name, we elevate the available genus name *Eviotops* from synonymy. The genus *Eviotops* Smith 1957 was erected for *Eviotops infulatus* (currently considered valid as *Eviota infulata*) but was later considered to be a synonym of *Eviota*. Because *Eviota infulata* is part of the unbranched clade in our phylogeny, we resurrect *Eviotops* out of synonymy and assign to it all species of dwarfgobies with unbranched pectoral-fin rays. The characters that should be used from here on to recognize members of these two distinct clades are as follows: *Eviota* have branched pectoral fins, 26 total vertebrae (except for *Eviota atrinasa* (previously *Sueviota*) *atrinasa* with 27), and a deeply notched, wrench-like posterior margin of the interopercle; and *Eviotops* have unbranched pectoral fins, 25 total vertebrae, and a very shallow notched posterior margin of the interopercle

For the non-monophyletic genus *Sueviota*, which is nested within the pectoral-branched *Eviota* clade, we synonymize the genus with *Eviota*. There are several synapomorphies shared between the *Eviota* and *Sueviota* that were noted by Winterbottom and Hoese (1988) in their original description of the genera. These are: tubular nostrils (thin tube in the case of the anterior one with melanophores towards the tip); pigmented subcutaneous bars in posterior part of body; absence of scales on the head, nape, pectoral base, and in most cases, the pelvic area; enlarged and interspersed canines in the outer row of both jaws, with some in the inner row of lower-jaw teeth; and the absence of preopercular sensory pores beyond the IT pore. These synapomorphies are present in all species of *Eviota* (including *Sueviota*) and *Eviotops*. The characters previously used to differentiate some (but not all) species of *Sueviota* from *Eviota* (see Pienemann et al. in review), while apomorphic, do not diagnose a reciprocally monophyletic clade outside of the rest of *Eviota*. Collectively, the taxonomic changes proposed here are informed by the most comprehensive genomic and taxonomic sampling to date, and result in well-supported monophyletic genera that are diagnosed by combinations of shared internal and external osteological features, ultimately helping to stabilize the classification of this diverse, abundant, and ecologically important lineage of reef fishes.

ACKNOWLEDGEMENTS

We thank Valentine and Motu Vaeoso (Department of Natural Resources, and Coral Reef Advisory Group respectively) in American Samoa for assistance in the field. We thank the people of Vava’u, Tonga and their government agencies, and Karen Stone for facilitating fieldwork. Ray Buckley and Richard Ball provided technical assistance in the field. Emily McFarland, Sam Ghods, Katlyn Fuentes, Sydney Schumaker, and Junda (David) Zhao helped with DNA extractions. Many thanks to Katherine Maslenikov for assistance with cataloging specimens, as well as Katherine Rogers, Isaiah Rothschild, and Fiona. We acknowledge the

California Academy of Sciences, the South African Institute of Aquatic Biodiversity, the National Museum of Natural History: Smithsonian Institution, and the Bishop Museum for their help with specimen loans. This project was in part funded by the Hall Conservation Genetics Research Fund (University of Washington), 2018 Lerner-Gray Memorial Fund of the American Museum of Natural History, the American Academy of Underwater Sciences Foundation, the Seattle Rho Chapter of Washington Women Educators, the William and Dorothy T. Gilbert Ichthyology Research Fund, and NSF Award #1827013.

REFERENCES

Agorreta A, San Mauro D, Schliewen U, Van Tassell JL, Kovačić M, Zardoya R, Rüber L. 2013. Molecular phylogenetics of Gobioidei and phylogenetic placement of European gobies. Molecular phylogenetics and evolution. 69(3):619-33.

Ahmadia GN, Tornabene L, Smith DJ, Pezold FL. 2018. The relative importance of regional, local, and evolutionary factors structuring cryptobenthic coral-reef assemblages. Coral Reefs. 37:279-93.

Allen GR, Erdmann MV, Cahyani ND. 2016. *Sueviota bryozophila*, a new species of coral-reef goby from Indonesia (Teleostei: Gobiidae). Journal of the Ocean Science Foundation. 20:76-82.

Allen GR, Erdmann MV. 2017. *Sueviota tubicola*, a new species of coral-reef goby (Teleostei: Gobiidae) from Papua New Guinea. Journal of the Ocean Science Foundation. 25:1-7.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics. 17(2):81-92.

Atta CJ, Coker DJ, Sinclair-Taylor TH, DiBattista JD, Kattan A, Monroe AA, Berumen ML. 2019. Conspicuous and cryptic reef fishes from a unique and economically important region in the northern Red Sea. PLoS One. 14(10):e0223365.

Bannikov AF, Carnevale G. 2016. † *Carlomonnius quasigobius* gen. et sp. nov.: the first gobioid fish from the Eocene of Monte Bolca, Italy. Bulletin of Geosciences. 91:13-22.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114-20.

Brandl SJ, Goatley CHR, Bellwood DR, Tornabene L. 2018. The hidden half: ecology and evolution of cryptobenthic fishes on coral reefs. Biological Reviews. 93(4):1846–1873. doi:10.1111/brv.12423.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS computational biology. 10(4):e1003537.

Brandl SJ, Tornabene L, Goatley CHR, Casey JM, Morais RA, Côté IM, Baldwin CC, Parravicini V, Schiettekatte NMD, Bellwood DR. 2019. Demographic dynamics of the smallest marine vertebrates fuel coral reef ecosystem functioning. Science. 364(6446):1189–1192. doi:10.1126/science.aav3384.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology. 22(11):3124–3140. doi:10.1111/mec.12354. https://onlinelibrary.wiley.com/doi/10.1111/mec.12354.

Depczynski M, Bellwood DR. 2003. The role of cryptobenthic reef fishes in coral reef trophodynamics. Marine Ecology Progress Series. 256:183–191. doi:10.3354/meps256183.

Depczynski M, Bellwood DR. 2006. Extremes, plasticity, and invariance in vertebrate life history traits: insights from coral reef fishes. Ecology. 87(12):3119-27.

DeRaad D. 2021. SNPfiltR: an R package for interactive and reproducible SNP filtering. doi:10.22541/au.163976415.53888836/v1.

Erdmann MV, Greenfield DW, Tornabene L. 2023 Jun 6. *Eviota taeiae*, a new dwarfgoby (Teleostei: Gobiidae) of the *Eviota guttata* complex from Samoa. doi:10.5281/ZENODO.8011284. https://zenodo.org/record/8011284.

Fujiwara K, Suzuki T, Motomura H. 2020. Two new dwarfgobies (Gobiidae) from southern Japan: *Eviota amamiko and Eviota perspicilla*. Ichthyological Research. 67:139-54.

Gómez-Buckley MC, Gallego R, Arranz V, Halafihi T, Stone K, Erdmann M, Tornabene LM. 2023. Comparing anesthetic stations and environmental DNA sampling to determine community composition of cryptobenthic coral reef fishes of Vava'u, Kingdom of Tonga. Coral Reefs. 42(3):785-97.

Gómez-Buckley MC, Pedraza-Marrón C, Erdmann MV, Tornabene L. (in review). Species delimitation in the *Eviota sigillata* complex, a widely distributed group of cryptobenthic coral reef fishes.

Greenfield DW. 2017. An overview of the dwarfgobies, the second most speciose coral-reef fish genus (Teleostei: Gobiidae: *Eviota*). Journal of the Ocean Science Foundation. 29:32–54. doi:10.5281/zenodo.1115683.

Greenfield DW, Tornabene L. 2014. *Eviota brahmi* n. sp. from Papua New Guinea, with a redescription of *Eviota nigriventris* (Teleostei: Gobiidae). Zootaxa 3793(1):133-146. doi.org/10.11646/zootaxa.3793.1.6

Greenfield D, Winterbottom R. 2016. A key to the dwarfgoby species (Teleostei: Gobiidae: *Eviota*) described between 1871 and 2016. Journal of the Ocean Science Foundation. 24(April):35–90. doi:10.5281/zenodo.219620.

Greenfield DW, Randall JE. 2017. *Sueviota pyrios*, a new species of coral-reef dwarfgoby from the Red Sea (Teleostei: Gobiidae). Journal of the Ocean Science Foundation, 25, 8–13. doi: http://dx.doi.org/10.5281/zenodo.268651

Greenfield DW, Tornabene L, Gómez-Buckley M, Erdmann MV. 2018. *Eviota maculosa*, a new dwarfgoby from the Western Pacific Ocean (Teleostei: Gobiidae). Journal of the Ocean Science Foundation, 31, 18–31. doi: http://dx.doi.org/10.5281/zenodo.1445810.

Greenfield DW, Tornabene L, Erdmann MV, Pada DN. 2019. *Eviota gunawanae*, a new microendemic dwarfgoby from the Fakfak Peninsula, West Papua, Indonesia (Teleostei: Gobiidae). Journal of the Ocean Science Foundation, 32, 57–67. doi.org/10.5281/zenodo.2616753

Greenfield DW, Erdmann MV. 2021. *Eviota flaviarma*, a new dwarfgoby from Papua New Guinea (Teleostei: Gobiidae). Journal of the Ocean Science Foundation. 38:27-34.

Greenfield D. 2021. Addendum to the 2016 key to the dwarfgobies (Teleostei: Gobiidae: *Eviota*). Journal of the Ocean Science Foundation. 38:1–12. doi:10.5281/zenodo.4458248.

Greenfield DW, Erdmann MV, Teitelbaum A. 2024. *Eviota bacata*, a new dwarfgoby (Teleostei: Gobiidae) from New Caledonia. Journal of the Ocean Science Foundation, 41, 14–21.

Hanahara N. 2023. Cryptic diversity of *Eviota* (Teleostei: Gobiidae) and their habitat use in the shallow waters of Okinawa Island. Marine Biodiversity 53:61 https://doi.org/10.1007/s12526-023-01369-6.

Harefa T, Chen IS. 2024. Complete mitochondrial genome of larged-eye pygmy goby *Trimma macrophthalmus* (Teleostei, Gobiidae) and its phylogenetic implications. Mitochondrial DNA Part B. 1;9(2):247-51.

Jenkins OP. 1903. Report on collections of fishes made in the Hawaiian Islands, with descriptions of new species. Bulletin of the U.S. Fish Commission. 22:415–511.

Jewett SL, Lachner EA. 1983. Seven new species of the Indo-Pacific genus *Eviota* (Pisces: Gobiidae). Proceedings of the Biological Society of Washington. 96:780–806. https://www.biodiversitylibrary.org/part/45377.

Jordan DS, Snyder JO. 1901. A review of the gobioid fishes of Japan, with descriptions of twenty-one new species. Proceedings of the United States National Museum.

Lachner EA, Karnella SJ. 1978. Fishes of the genus *Eviota* of the Red Sea with descriptions of three new species (Teleostei, Gobiidae). Smithsonian Contributions to Zoology. doi:10.5479/si.00810282.286.

Lachner EA, Karnella SJ. 1980. Fishes of the Indo-Pacific Genus *Eviota* with Descriptions of Eight New Species (Teleostei: Gobiidae) Smithsonian Contributions to Zoology – Number 315.

Larson H. 1985. A revision of the gobiid genus *Bryaninops* (Pisces), with a description of six new species. Beagle: Records of the Museums and Art Galleries of the Northern Territory. 2:57-93.

Larson HK. 1987. A new species of *Bryaninops* (Pisces: Gobiidae) with notes on new records of three species of the genus. Beagle: Records of the Museums and Art Galleries of the Northern Territory. 4:139-45.

Larson HK. 1990. A revision of the commensal gobiid fish genera *Pleurosicya and Luposicya* (Gobiidae), with descriptions of eight new species of *Pleurosicya* and discussion of related genera. Beagle: Records of the Museums and Art Galleries of the Northern Territory. 7(1):1-53.

Larson HK. 2002. A new genus of small coral-reef goby (Teleostei: Gobiidae) from the Indo-west Pacific, with discussion of its relationships. Beagle: Records of the Museums and Art Galleries of the Northern Territory. 18:63-71.

Longo G, Bernardi G. 2015. The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data. Molecular Phylogenetics and Evolution, *88:*55-63.

McCraney WT, Thacker CE, Alfaro ME. 2020. Supermatrix phylogeny resolves goby lineages and reveals unstable root of Gobiaria. Molecular Phylogenetics and Evolution. 1;151:106862.

Munday PL. 2001. Fitness consequences of habitat use and competition among coral-dwelling fishes. Oecologia. doi:10.1007/s004420100690.

Orr JW, Spies I, Stevenson DE, Longo,GC, Kai Y, Ghods SAM, Hollowed M. 2019. Molecular phylogenetics of snailfishes (Cottoidei: *Liparidae*) based on MtDNA and RADseq genomic analyses, with comments on selected morphological characters. Zootaxa, 4642(1):1-79.

Ortiz EM. 2019**.** vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. DOI:10.5281/zenodo.2540861.

Paradis E, Schliep K. 2019. "ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R." *Bioinformatics*, **35**, 526-528. doi:10.1093/bioinformatics/bty633.

Peinemann V, Pombo-Ayora L, Tornabene L, Berumen ML. (in review). The grumpy dwarfgoby, a new species of *Sueviota* (Teleostei:Gobiidae) from the Red Sea.

Rambaut A. 2012. FigTree v1.4.4. Available from: http://tree.bio.ed.ac.uk/software/figtree/

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Susko E, editor. Systematic Biology. 67(5):901–904. doi:10.1093/sysbio/syy032. https://academic.oup.com/sysbio/article/67/5/901/4989127.

Reichenbacher B, Gregorová R, Holcová K, Šanda R, Vukić J, Přikryl T. 2018. Discovery of the oldest Gobius (Teleostei, Gobiiformes) from a marine ecosystem of Early Miocene age. Journal of Systematic Palaeontology. 16(6):493-513.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61:539–542.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Revell L. 2024. “phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things).” *PeerJ*, **12**, e16505. doi:10.7717/peerj.16505.

Sato MC, Motomura H. 2024. *Gobiodon spadix*, a new coral goby (Teleostei: Gobiidae) from southern Japan. Ichthyological Research. 29:1-0.

Thacker CE, Roje DM. 2011. Phylogeny of Gobiidae and identification of gobiid lineages. Systematics and Biodiversity. 9(4):329-47.

Thacker CE. 2015. Biogeography of goby lineages (Gobiiformes: Gobioidei): origin, invasions and extinction throughout the Cenozoic. Journal of Biogeography. 42(9):1615-25.

Tornabene L, Ahmadia GN, Berumen ML, Smith DJ, Jompa J, Pezold F. 2013a. Evolution of microhabitat association and morphology in a diverse group of cryptobenthic coral reef fishes (Teleostei: Gobiidae: *Eviota*). Molecular Phylogenetics and Evolution. 1;66(1):391-400.

Tornabene L, Chen Y, Pezold F. 2013b. Gobies are deeply divided: Phylogenetic evidence from nuclear DNA (Teleostei: Gobioidei: Gobiidae). Systematics and Biodiversity. 11(3):345–361. doi:10.1080/14772000.2013.818589.

Tornabene L, Valdez S, Erdmann M, Pezold F. 2015. Support for a “Center of Origin” in the Coral Triangle: Cryptic diversity, recent speciation, and local endemism in a diverse lineage of reef fishes (Gobiidae: *Eviota*). Molecular Phylogenetics and Evolution. 82(PA):200–210. doi:10.1016/j.ympev.2014.09.012. http://dx.doi.org/10.1016/j.ympev.2014.09.012.

Tornabene L, Valdez S, Erdmann MV, Pezold FL. 2016. Multi-locus sequence data reveal a new species of coral reef goby (Teleostei: Gobiidae: *Eviota*), and evidence of Pliocene vicariance across the Coral Triangle. Journal of Fish Biology. 88(5):1811–1834. doi:10.1111/jfb.12947.

Tornabene LU, Greenfield DW. 2016. *Eviota erdmanni* (Teleostei: Gobiidae), a new dwarfgoby from the Savu Sea, Flores, Indonesia. Journal of the Ocean Science Foundation. 22:1-9.

Tornabene L, Greenfield DW, Erdmann MV. 2021. A review of the *Eviota zebrina* complex, with descriptions of four new species (Teleostei, Gobiidae). ZooKeys 1057:149–184. https://doi.org/10. 3897/zookeys.1057.66675.

Tornabene L, Manning R, Robertson DR, Van Tassell JL, Baldwin CC. 2022. A new lineage of deep-reef gobies from the Caribbean, including two new species and one new genus (Teleostei: Gobiidae: *Gobiosomatini*). Zoological Journal of the Linnean Society. 1;197(2):322-43.

Vaz DFB, Goatley CHR, Tornabene L. (in Review). Osteology of dwarfgobies *Eviota* and *Sueviota* (Gobiidae: Gobiomorpharia), with comments on morphological traits of coral gobies.

Winterbottom R, Hoese DF. 1988. A new genus and four new species of fishes from the Indo-West Pacific (Pisces; Perciformes; Gobiidae), with comments on relationships. (No. 37). Royal Ontario Museum.Winterbottom R, Harold AS. 2005. *Gobiodon prolixus*, a new species of gobiid fish (Teleostei: Perciformes: Gobiidae) from the Indo-west Pacific. Proceedings of the Biological Society of Washington. 118(3):582-9.

Winterbottom R, Hanner RH, Erdmann MV, Choffe K, Zur M, Steinke C. 2020. The cryptic cornucopia revisited – an extended analysis of the COI gene in the gobiid fish genus *Trimma* (Percomorpha : Gobiiformes). 132(2020):91–132.

CHAPTER 3 FIGURES

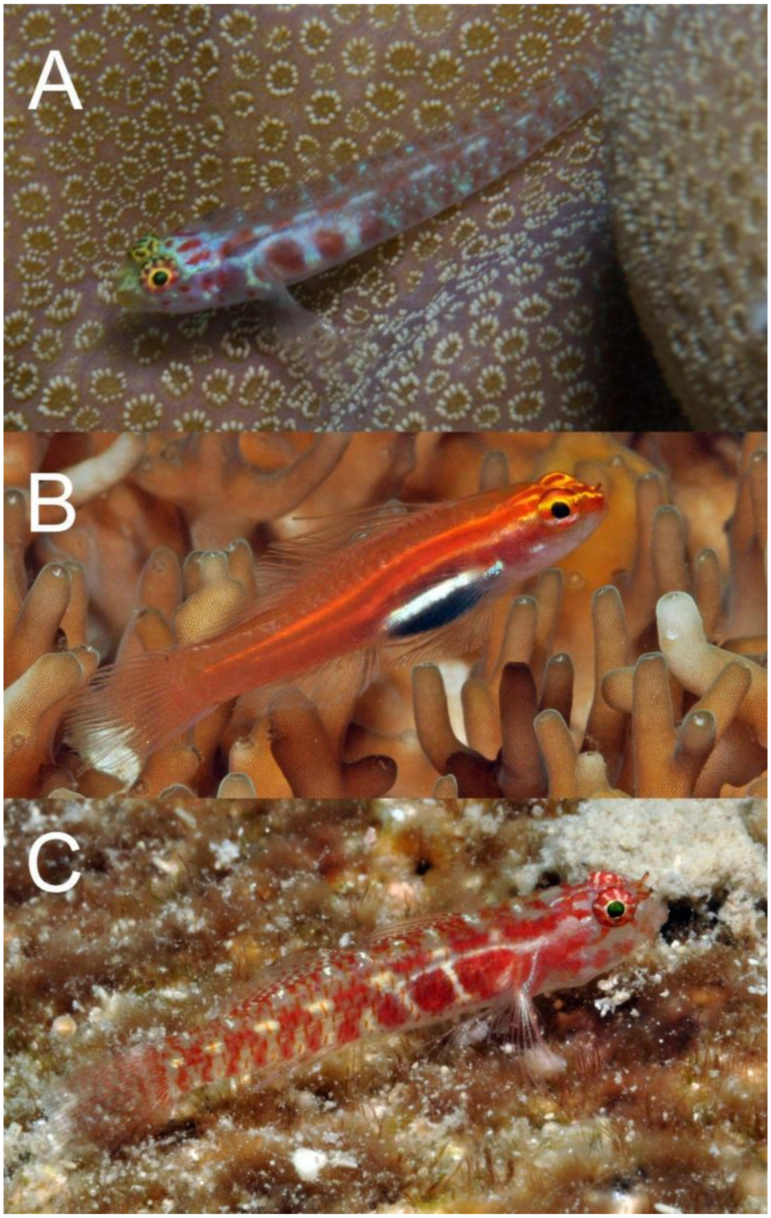


Figure 1. *In situ* photographs of: (A) representative of branched clade, *Eviota melasma;* (B) representative of unbranched clade, *Eviota cf. atriventris*, and (C) *Sueviota lachneri*. Photo credits, Mark V. Erdmann.

Figure 2. Phylogenetic maximum likelihood tree for the 164 taxa (R = 0.4). Red is the *Eviota* unbranched clade; blue is the *Eviota* branched clade. Numbers at nodes are bootstrap support values from 100 bootstrap replicates. Thickness of branches also reflect support values for the subtending node.

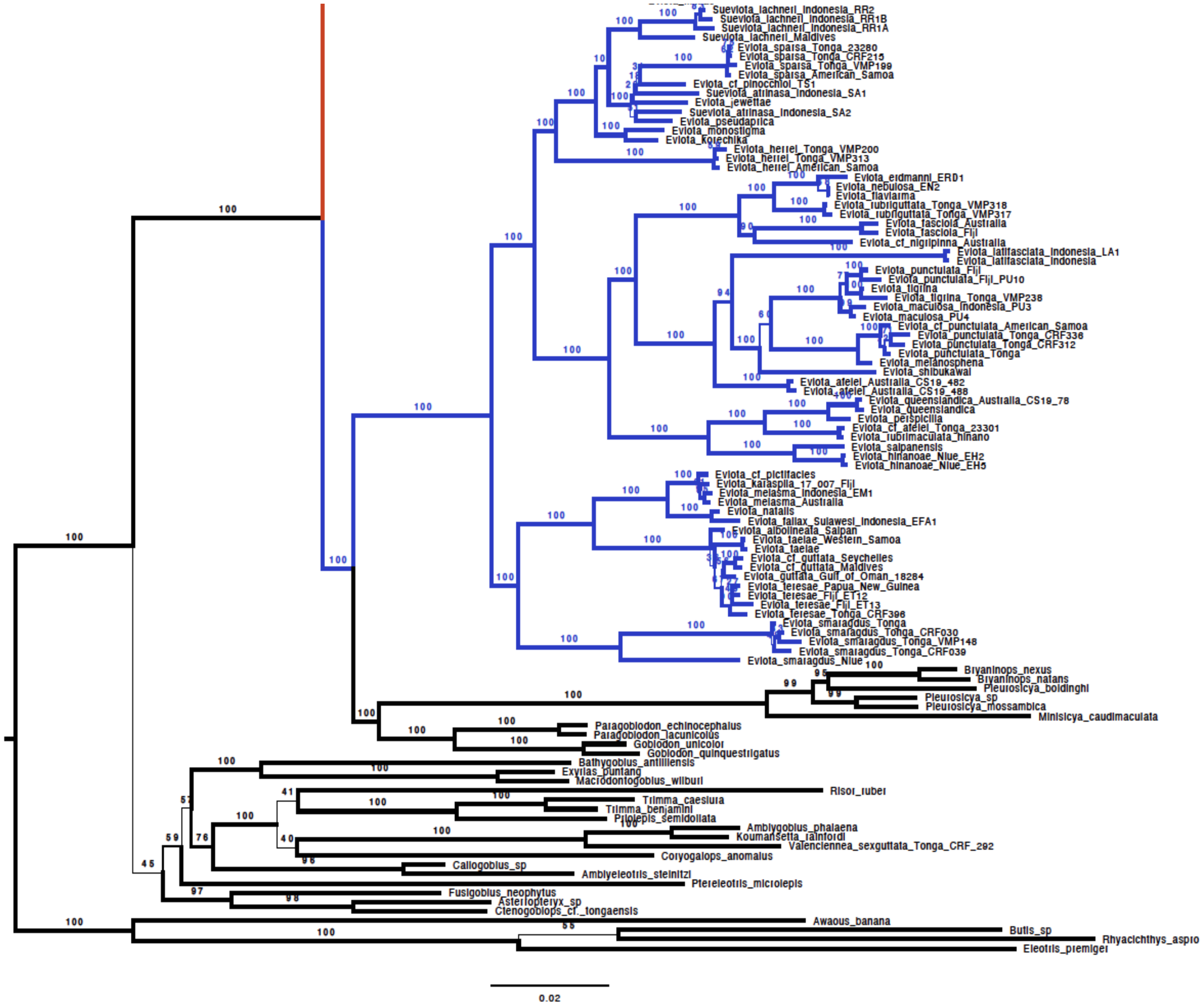


Figure 2. Continued.

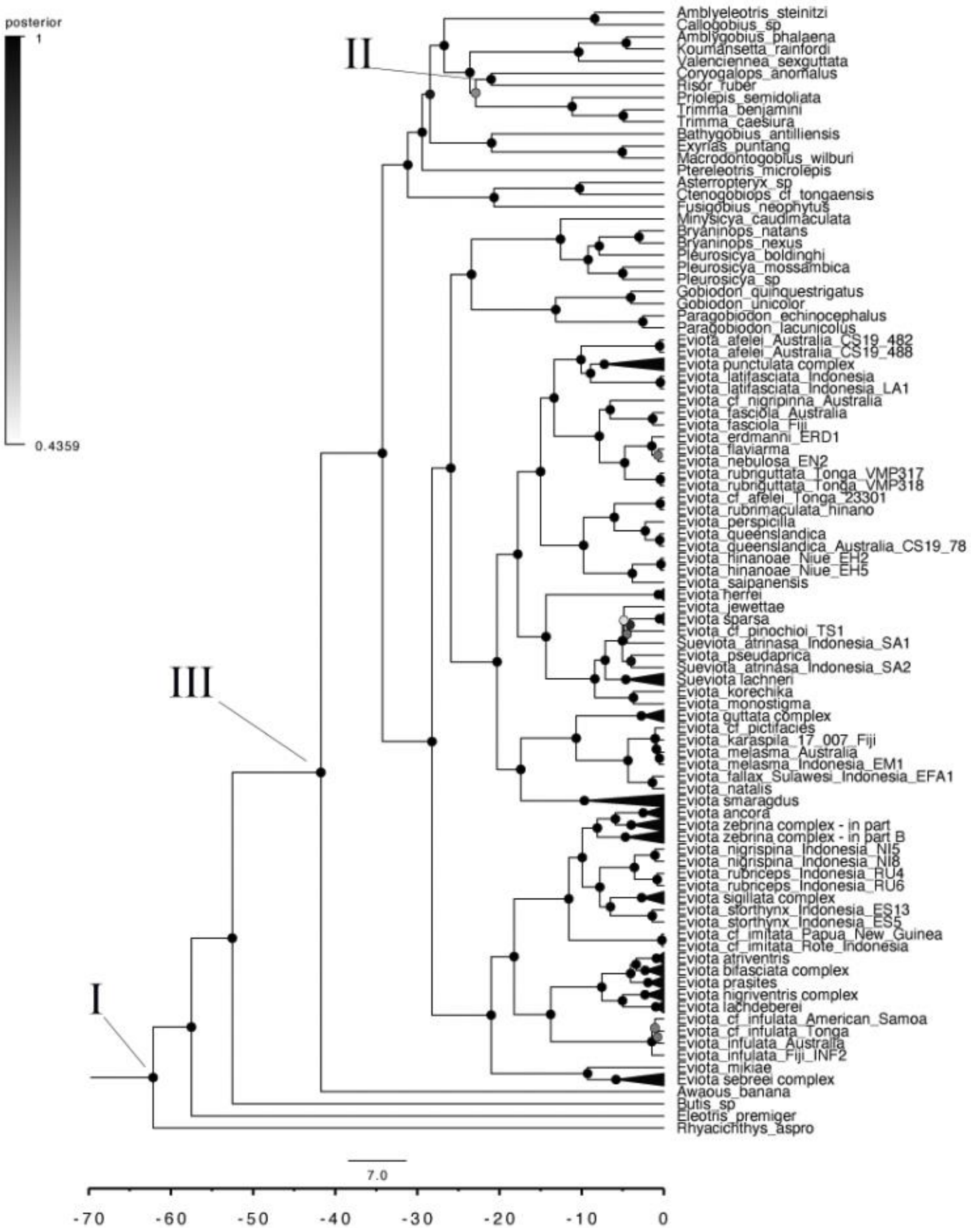


Figure 3. Time-calibrated phylogeny of Gobioidei. Roman numerals refer to placement of fossil calibration points. Colors at the nodes refer to Bayesian posterior probability support values. Clades with >3 individuals from a single species or clades that represent known species complexes previously reported in the literature are collapsed into triangles for ease of viewing.

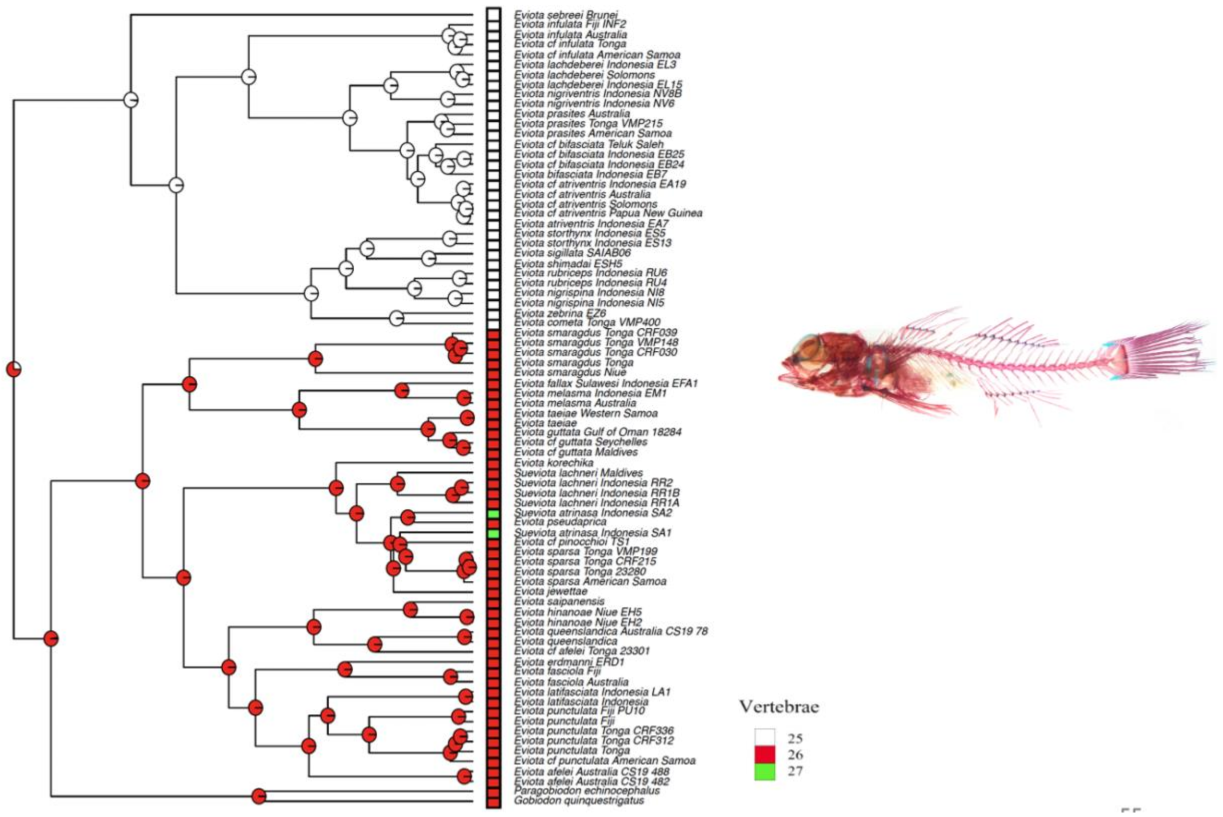


Figure 4. Ancestral character estimation for vertebral count. Pies at nodes represent posterior probabilities for ancestor's character state. Photo credit, Diego Vaz.

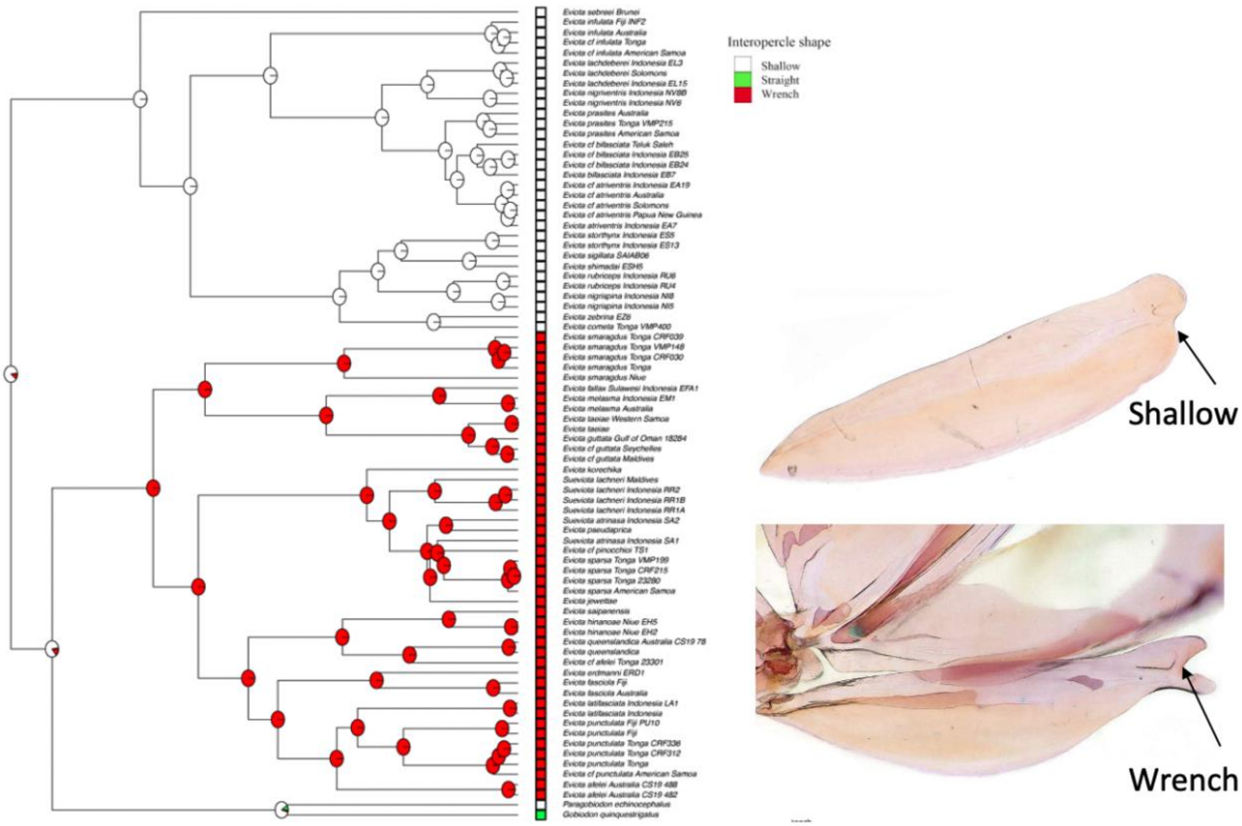


Figure 5. Ancestral state reconstruction for interopercle posterior shape. Pies at nodes represent posterior probabilities for ancestor's character state. Photo inset from Diego Vaz (in review).

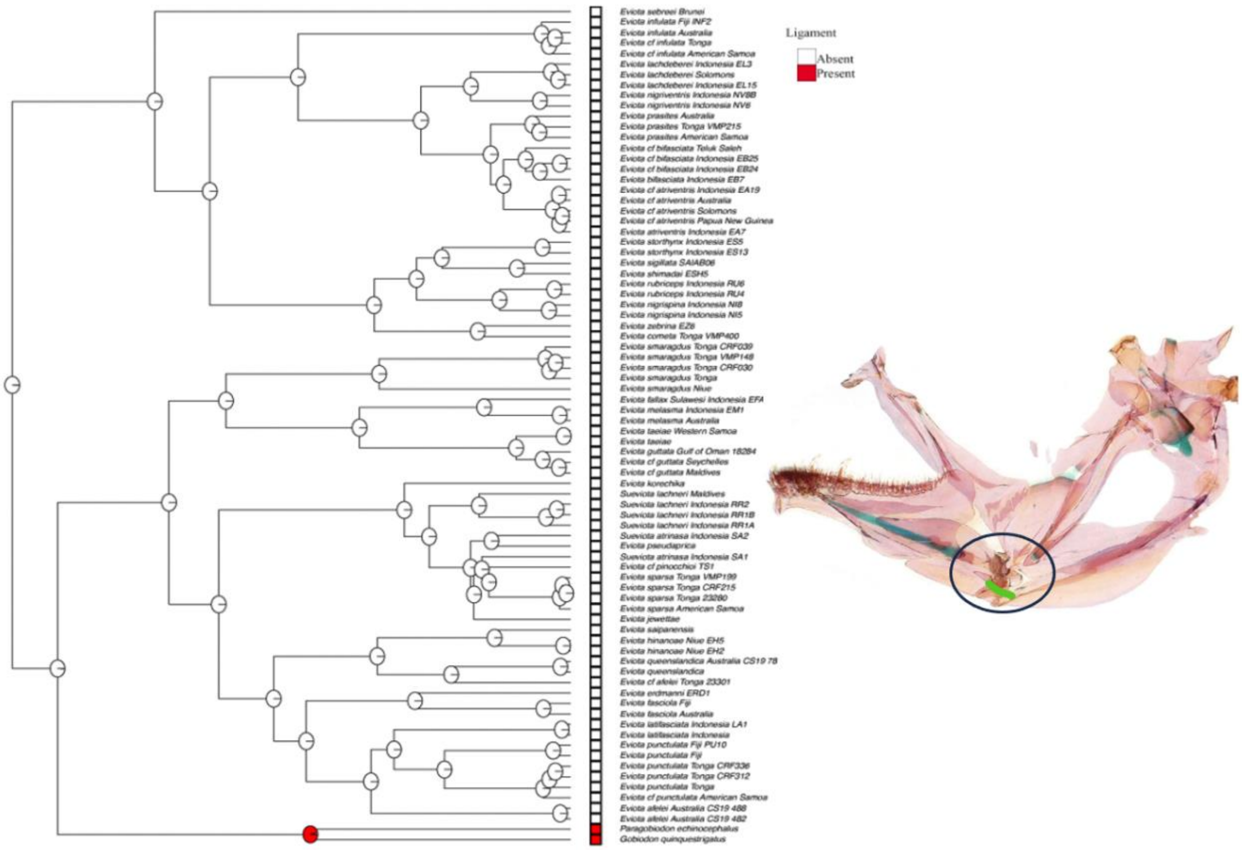


Figure 6. Ancestral state reconstruction for the ligament connecting the interopercle and retroarticular. Pies at nodes represent posterior probabilities for ancestor’s character state. Photo inset from Diego Vaz (in review).

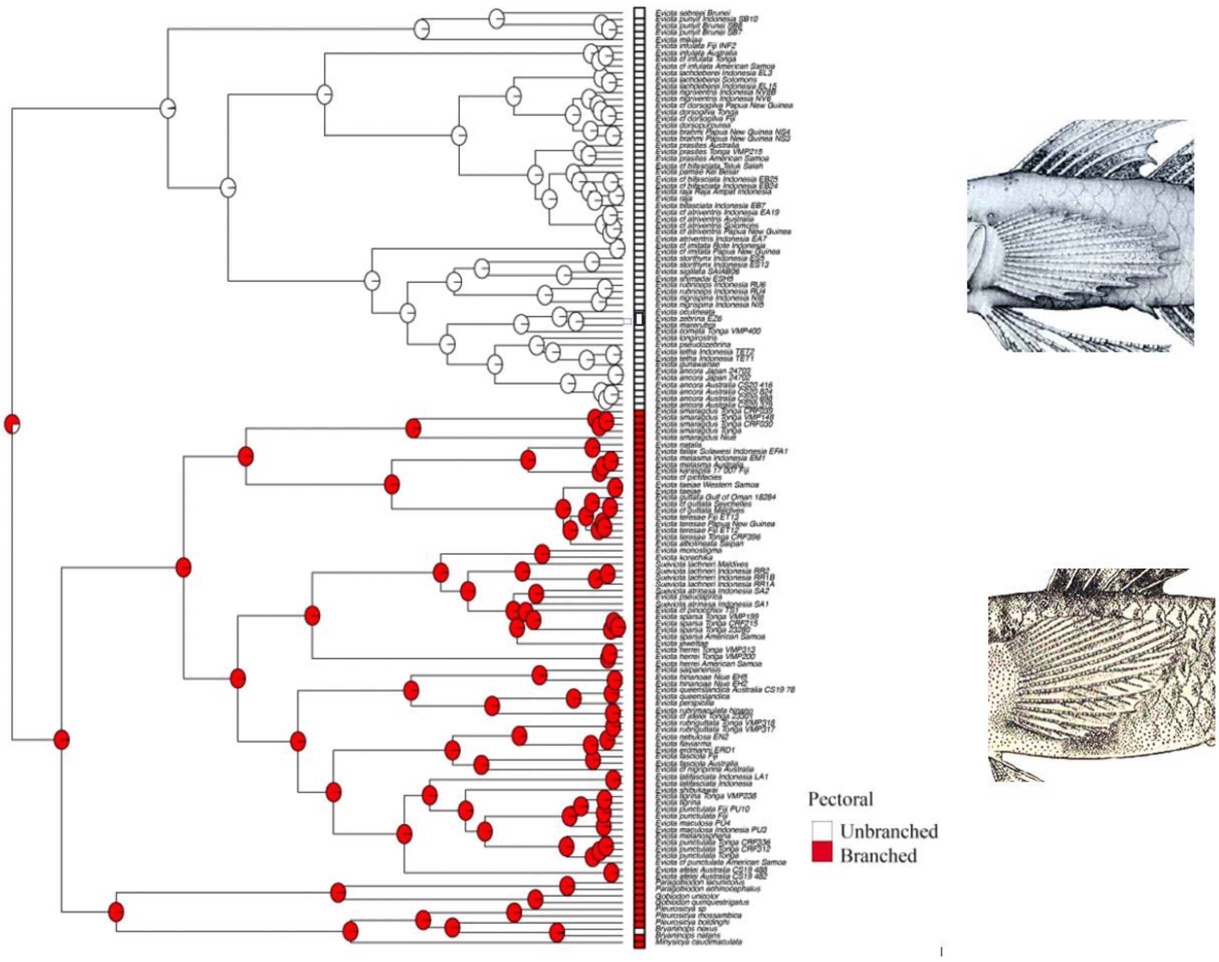


Figure 7. Ancestral state reconstruction for pectoral fin branching. Pies at nodes represent posterior probabilities for ancestor’s character state. Insets from Greenfield & Winterbottom 2016)

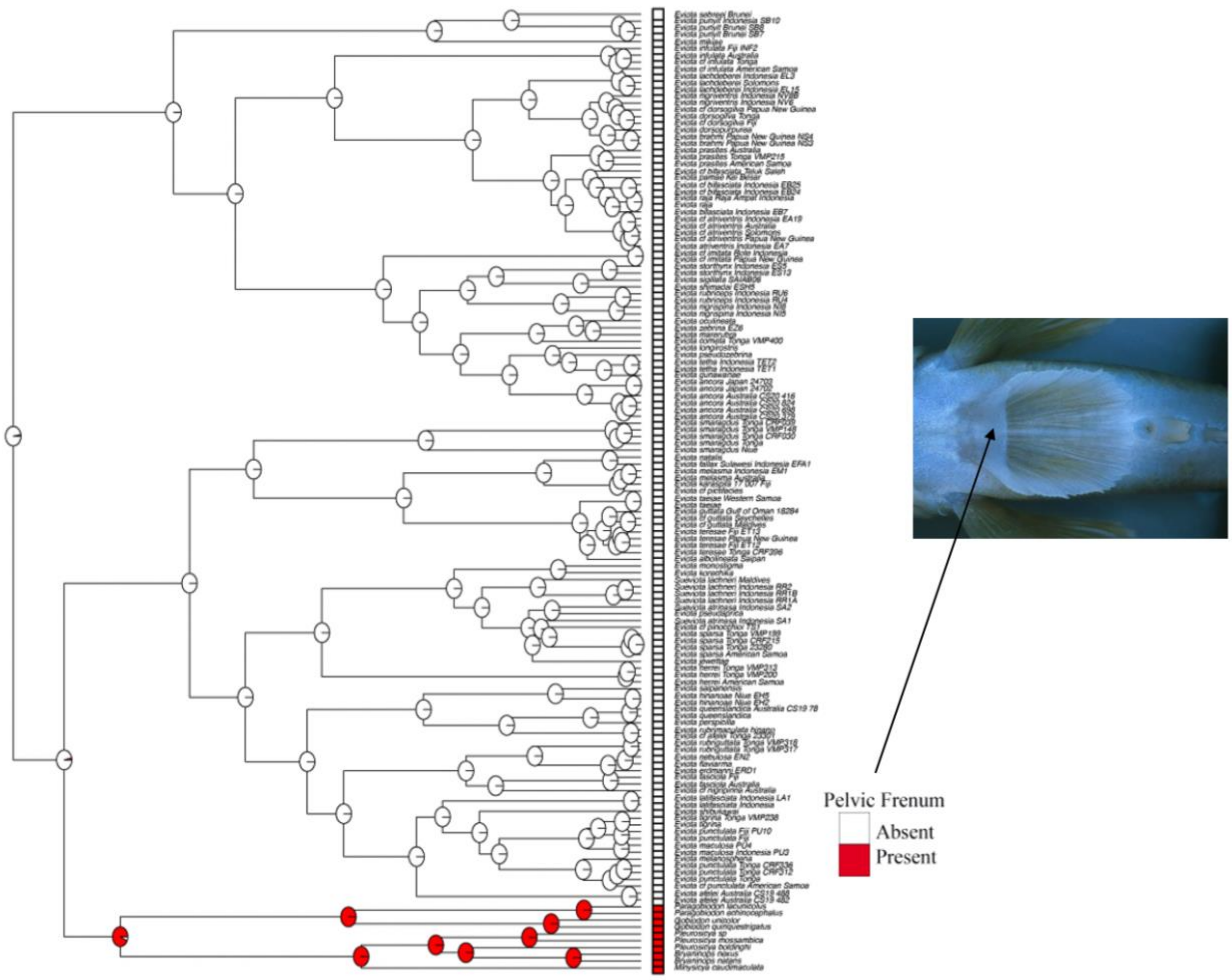


Figure 8. Ancestral state reconstruction for presence/absence of pelvic frenum. Pies at nodes represent posterior probabilities for ancestor's character state.

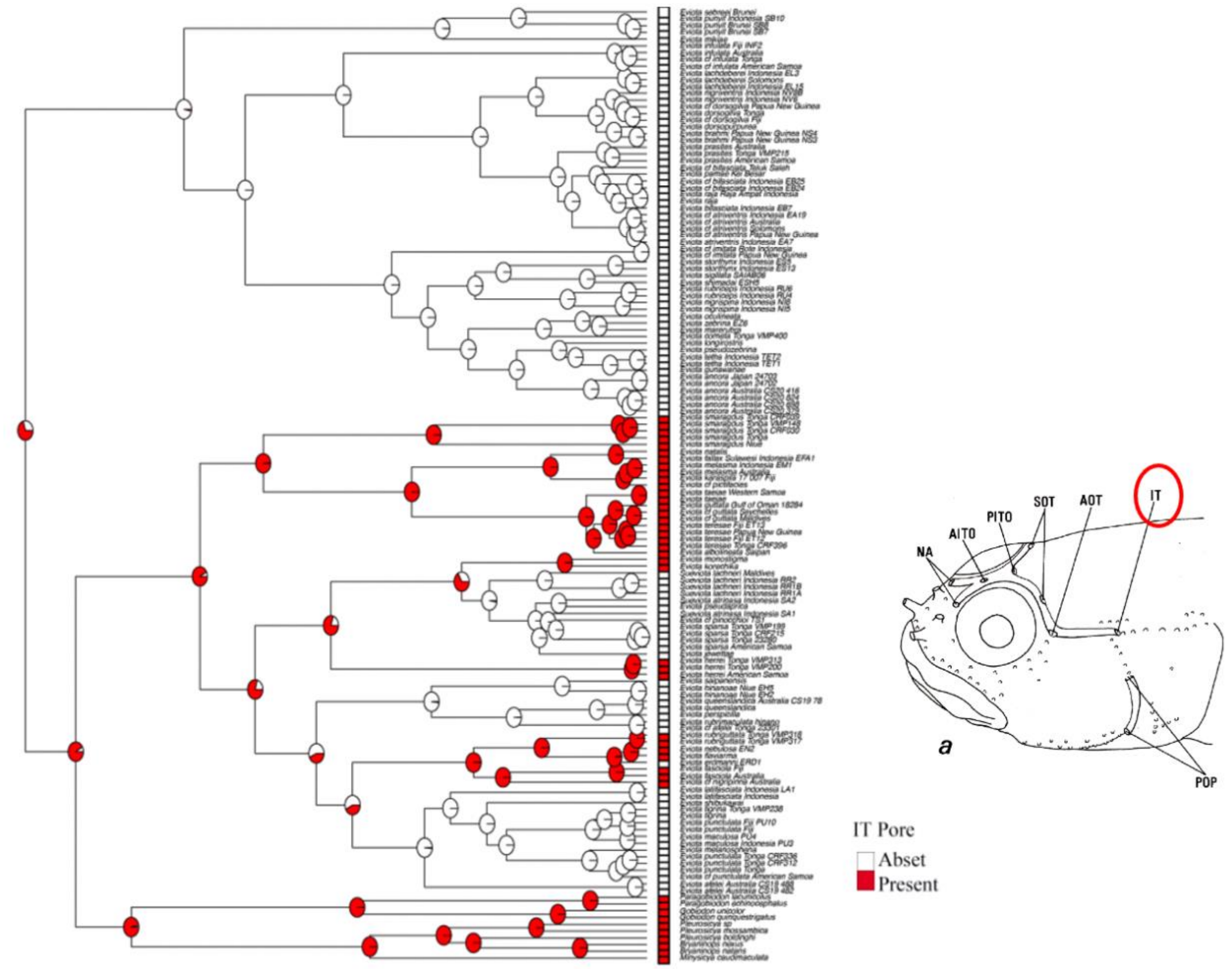


Figure 9. Ancestral state reconstruction for presence/absence of IT sensory pores. Pies at nodes represent posterior probabilities for ancestor’s character state.

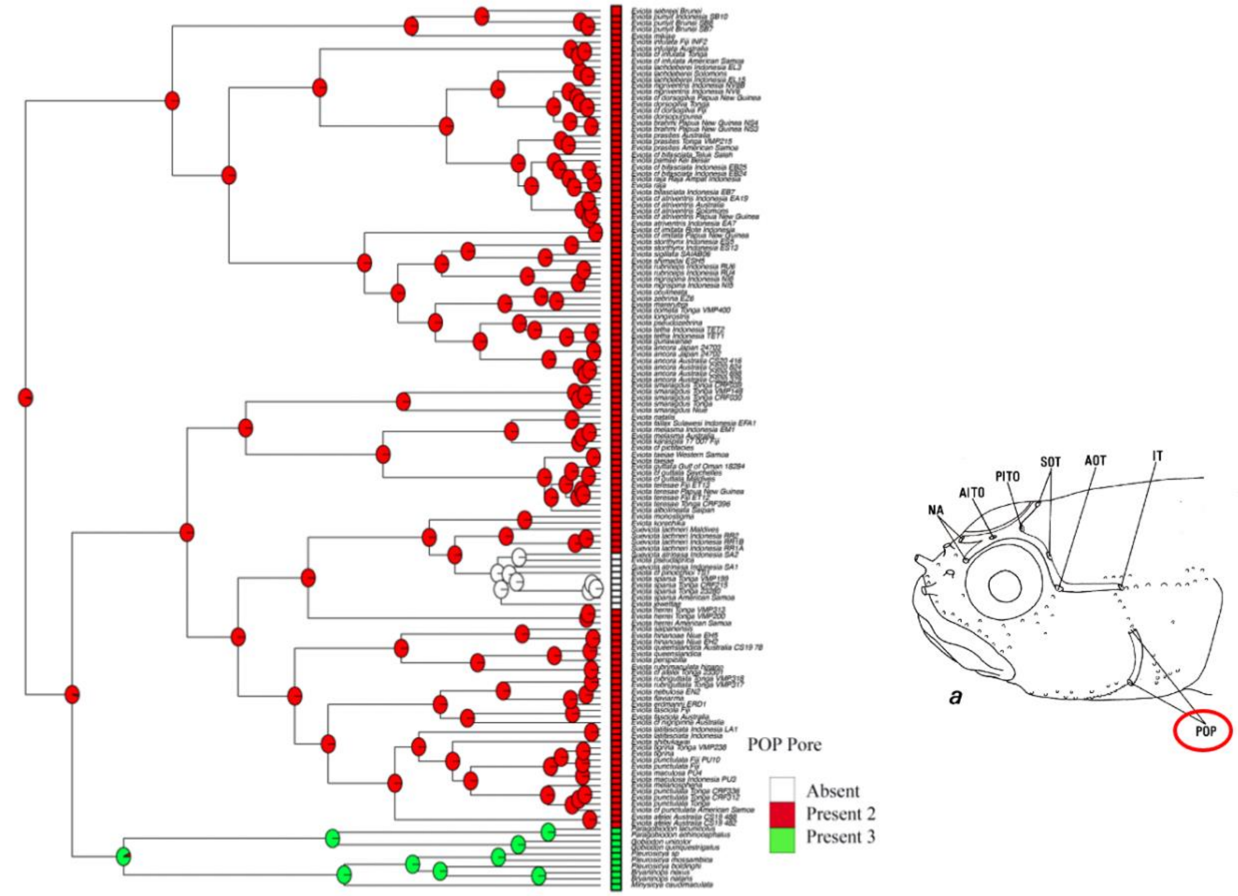


Figure 10. Ancestral state reconstruction for absence, presence (two pores), or presence (three pores) of POP sensory pores. Pies at nodes represent posterior probabilities for ancestor's character state.

CHAPTER3 TABLES

Table 1. Previous phylogenetic studies, including dwarfgobies and close relatives.

Taxa of focus	***Eviota* specimens (species)**	***Sueviota* specimens (species)**	**Mitochondrial loci**	**Nuclear loci**	**Total bp**	**Reference**
Gobiidae	1	-	3	2	5,278	Thacker and Roje 2011
Gobiidae	20(16)	-	1	1	1,696	Tornabene et al. 2013
Gobiidae	1	-	2	3	6,157	Agorreta et al. 2013
Gobiiformes	44(39)	-	-	1	18,065	McCraney et al. 2020
Coral Gobies	3(2)	-	3	-	1,199	Herler et al. 2009
E. bifasciata and E. *nigriventris* complexes	32(6)/40(6)	-	1	1	577/1,094	Tornabene et al. 2015
E. nigriventris complex	29(6)	-	1	1	1,808	Greenfield and Tornabene 2014
E. guttata complex	27(14)	-	1	1	1,842	Tornabene et al. 2023
E. zebrina complex	32(15)	-	1	1	1,787	Greenfield et al. 2019; Tornabene et al. 2021
E. sigillata complex	53(13)	-	-	1	1,142	Gómez-Buckley et al. in review
E. sigillata complex	104(11)	-	-	4,591	660,013	Gómez-Buckley et al. in review
E. punctulata complex	31(15)	-	1	1	1,787	Greenfield et al. 2018
Eviota	49(11)	-	1	-	NA	Hanahara 2023
Eviota and *Sueviota*	147(65)	7(3)	0	440	98,942	Present study

Table 2. General composition of individual samples for genomic and morphological analyses.

Taxon name	**Individuals**	**Species**
Eviota	147	65
Sueviota	7	3
Gobiodon	3	3
Paragobiodon	2	2
Other Gobiidae (non-Gobiodon lineage)	17	16
Oxudercidae	3	3
Outgroup (non-Gobiidae)	3	3
Total	182	95

Table 3. Number of sites (bp), total and variant (SNPs), loci (RADtags), and taxa retained after demultiplexing and filtering ddRAD data using Stacks program with three different values for the '-R' parameter (minimum percentage of individuals across the entire alignment required to include a locus, using the 'populations' module in Stacks).

-R	Total sites (bp)	Loci (RADtags)	Variants sites (SNPs)	Taxa
0.4	98942	440	22100	164
0.3	187744	824	41286	162
0.2	557921	2441	112555	160

Table 4. Calibration points from Tornabene et al. 2022 used for tree calibration in our study. The mean, sigma, and offset values refer to the priors used for the values of our lognormal distributions of the ages of our calibration points (Figure 3).

Calibration Points	Min Age (million years)	Crown age for	Distribution
†*Carlomonnius quasigobius* (I)	50	Gobioidei	log-normal (mean=1.8, sigma=1.0, offset=50.0)
†*Gobius jarosi* (II)	19.1	*Coryogalops* + *Risor* (Oxudercidae+ Gobiidae)	log-normal (mean=1.8, sigma=0.8, offset=19.1)
†*Pirskenius diatomaceous* (III)	28	+Butidae	log-normal (mean=1.8, sigma=1.0, offset=28.0)

Table 5. Character evolution analysis for the 15 characters with the corresponding phylogenetic p-values calculated using the likelihood ratio test.

Phenotypic trait	**Abbreviation**	**Likelihood ratio test**	**Phylogenetic signal p-value**
Vertebral count	**VERT**	0.0152	**0**
Preopercle wrench shaped	**WRENCH**	0.0303	**0**
Opercle-jaw ligament	**RAR_OP**	0.9431	**1.83E-03**
Membrane between pelvic fins	**MEMBRANE**	0.6013	**8.88E-16**
Pelvic Frenum	**FRENUM**	0.8609	**1.35E-14**
5th pelvic fin ray branching	**PELVIC**	0.0639	**1.10E-14**
Pectoral fin ray branching	**PEC**	0.3862	**0**
Sensory pores	**NA_pore**	0.6589	**5.24E-12**
	AITO	0.5294	0.3836
	PITO	0.0064	1.00
	SOT	0.3822	**0**
	AOT	0.0003	**0**
	IT	0.1821	**0**
	POP	0.7726	**0**

CHAPTER 3 SUPPLEMENT

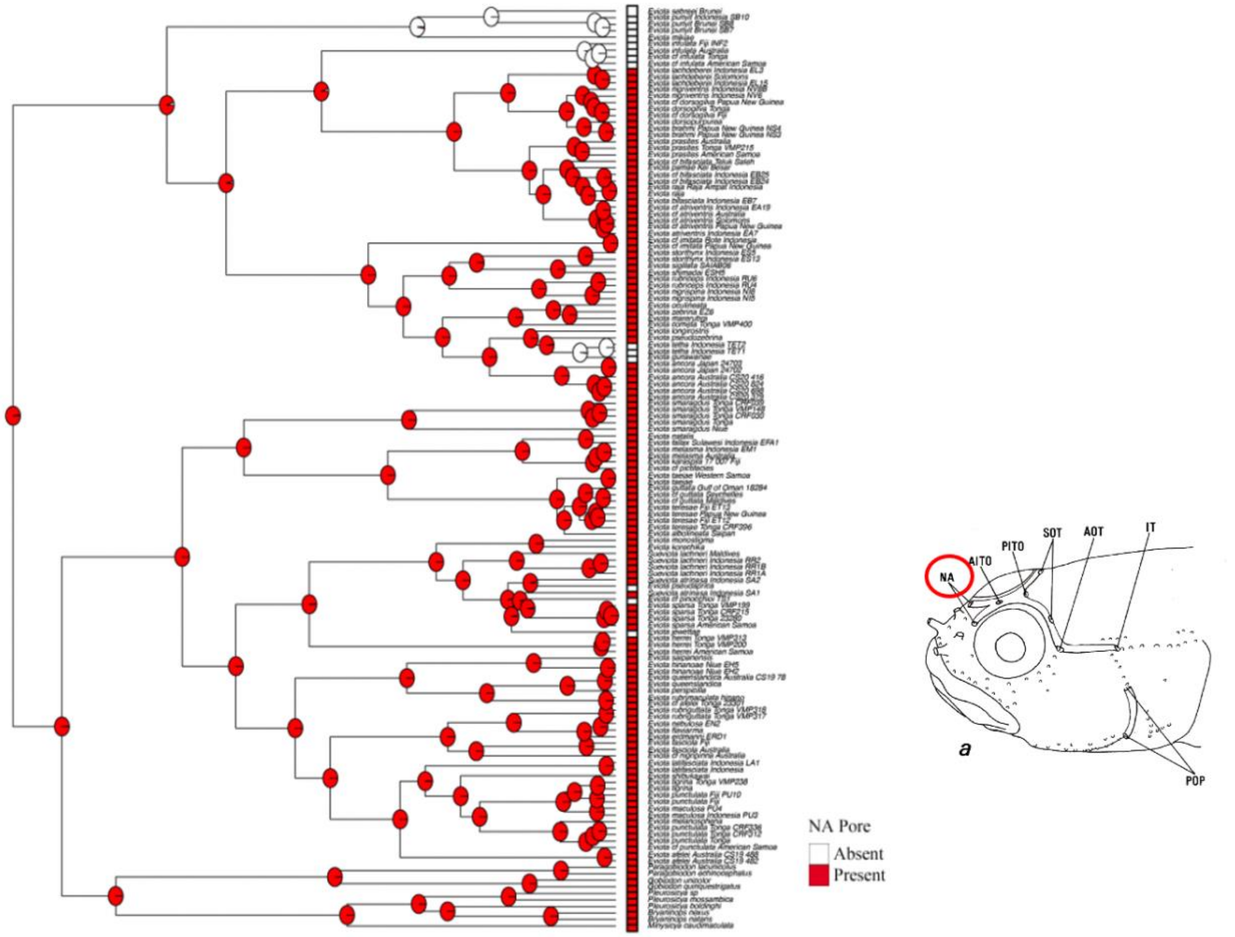


Figure S1. Ancestral state reconstruction for presence absence of NA sensory pores. Inset image from Lachner & Karnella (1980).

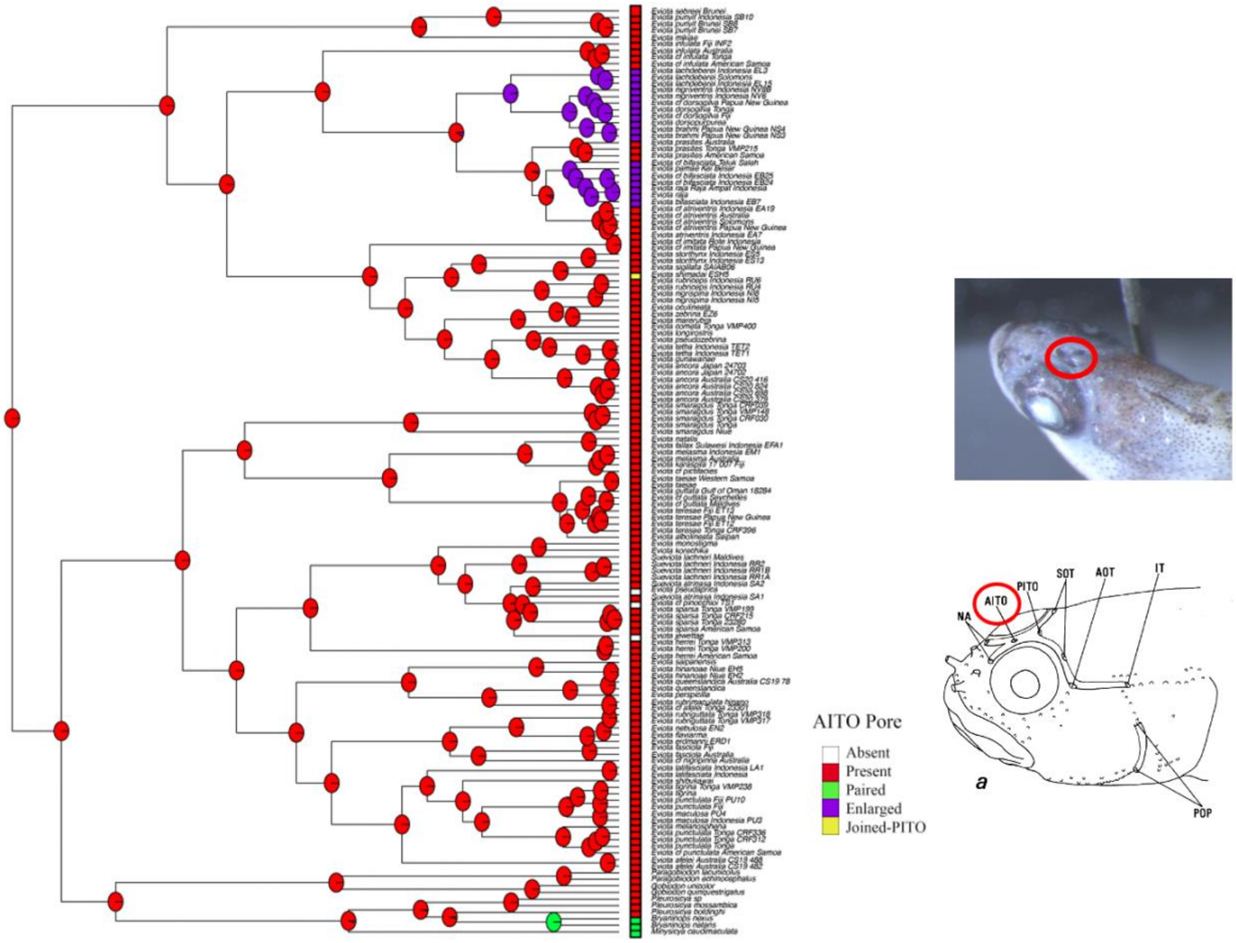


Figure S2. Ancestral state reconstruction for presence, absence, paring, enlarged, or joining of AITO and PITO sensory pores. Inset images from Lachner & Karnella (1980) or taken by main author.

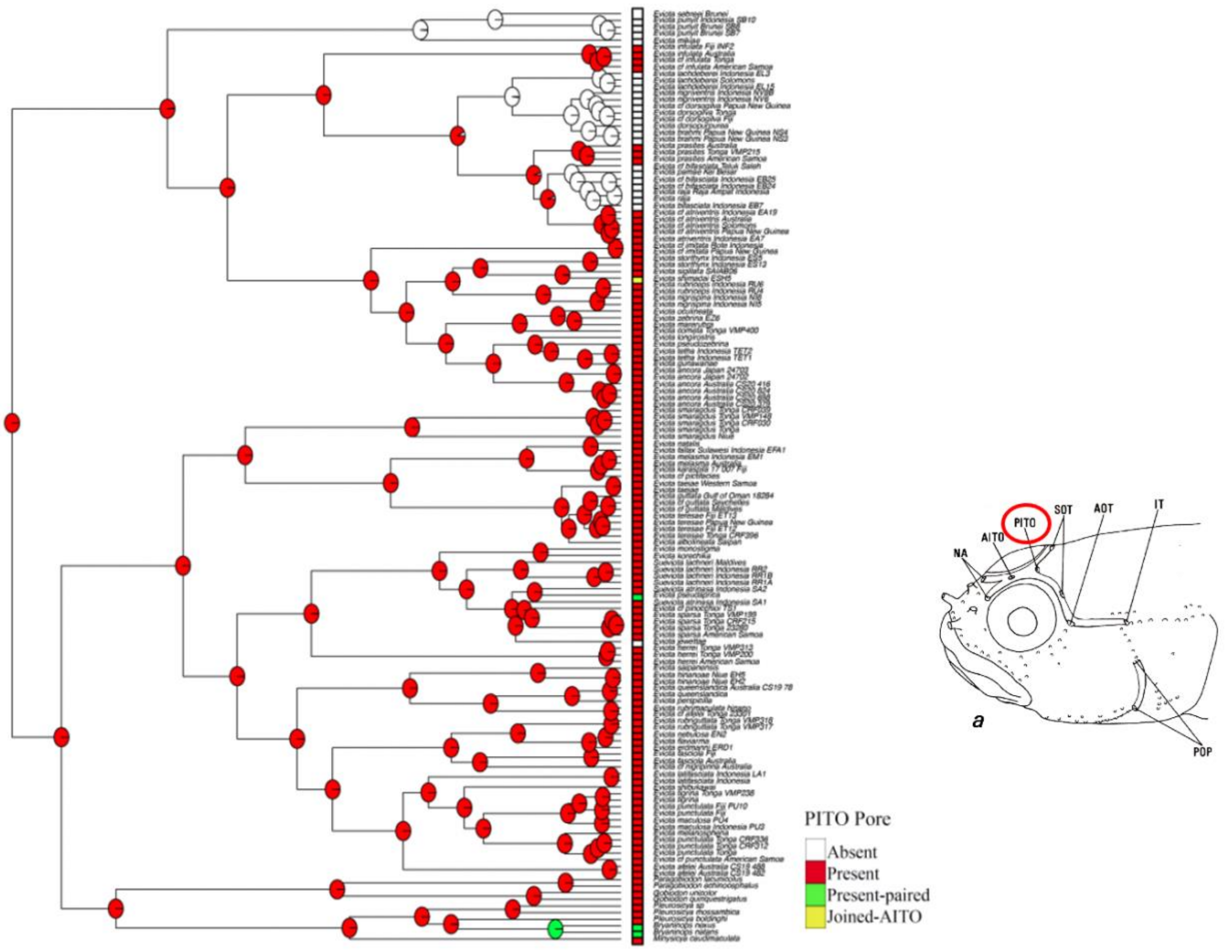


Figure S3. Ancestral state reconstruction for presence, absence, paring, joining to PITO of AITO sensory pores. Inset image from Lachner & Karnella (1980).

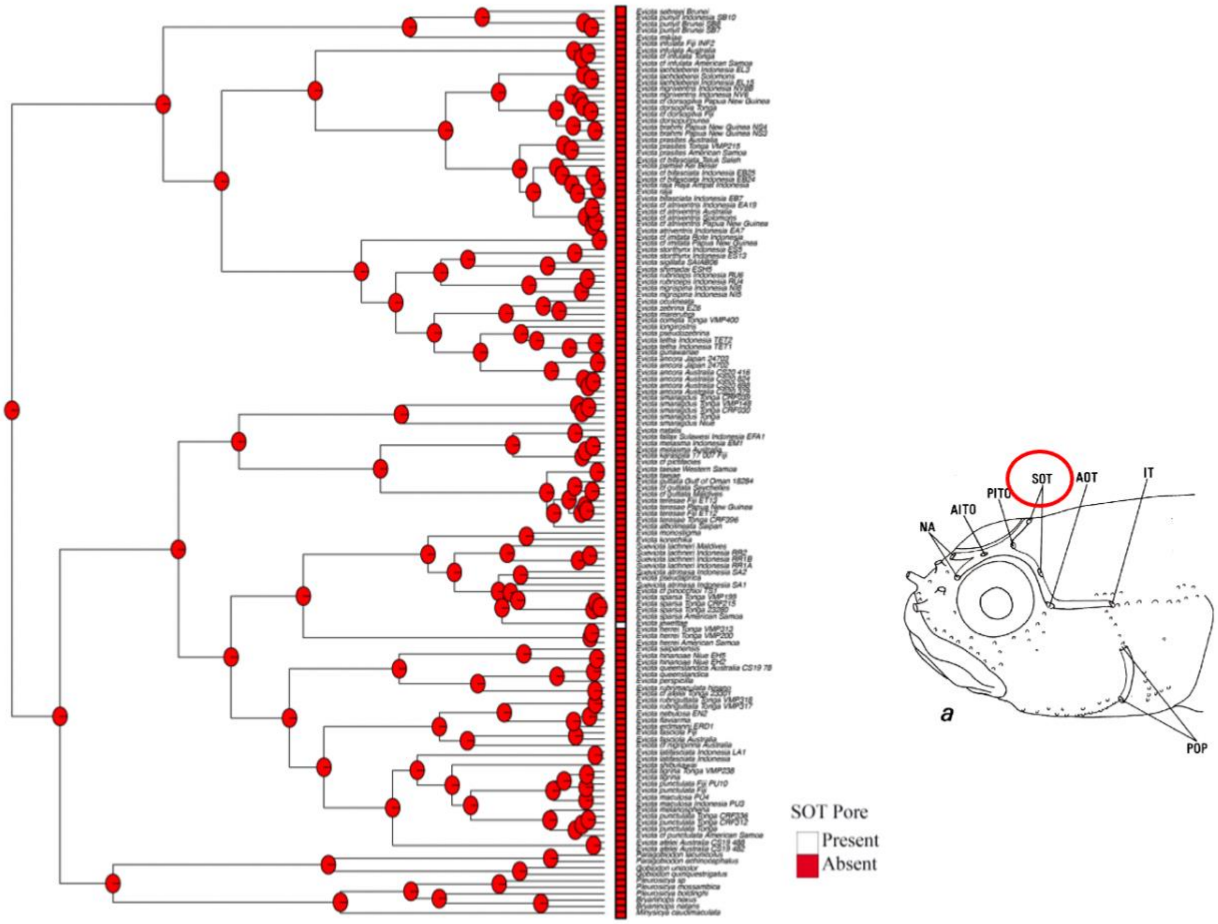


Figure S4. Ancestral state reconstruction for presence/absence of SOT sensory pores. Inset image from Lachner & Karnella (1980).

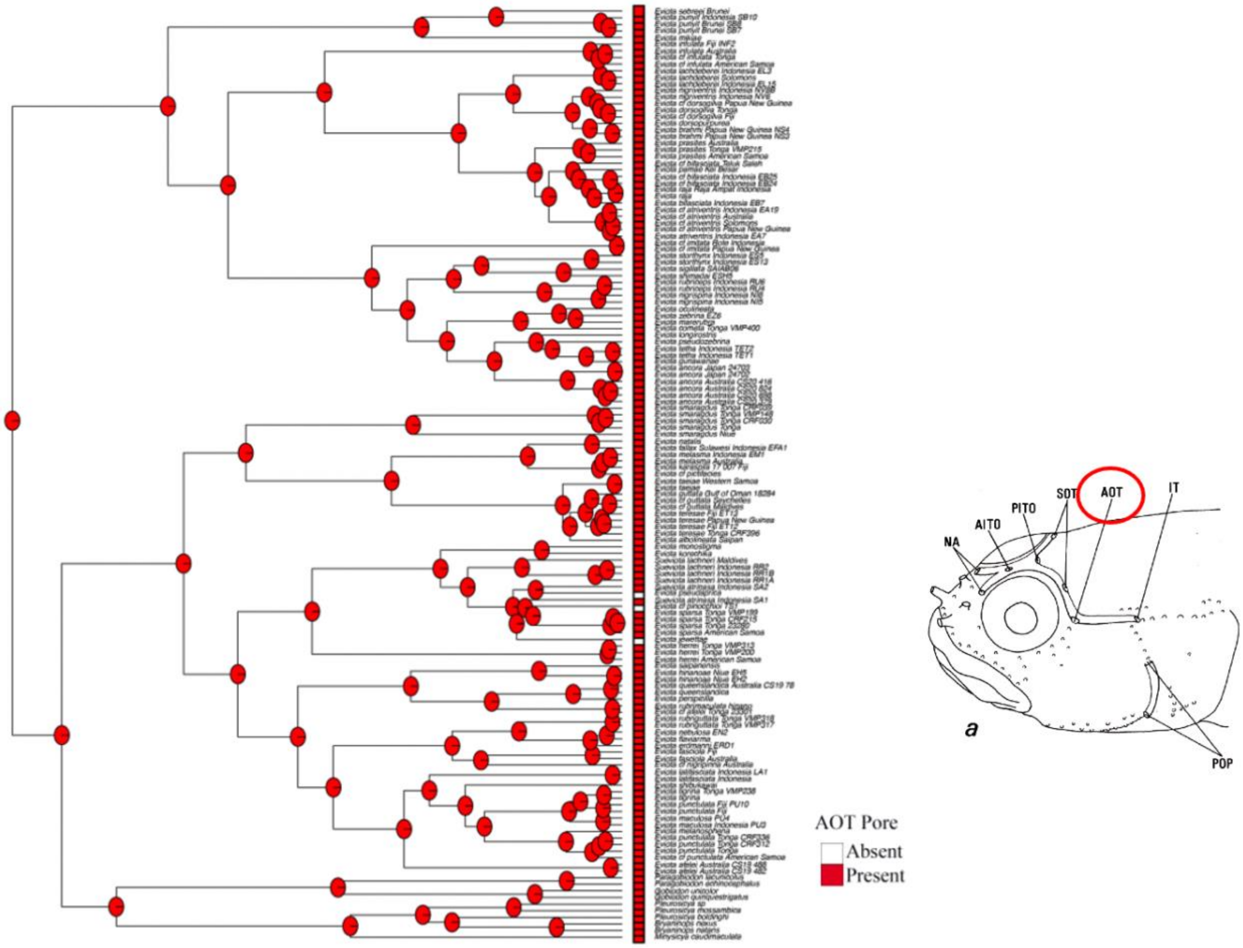


Figure S5. Ancestral state reconstruction for presence/absence of AOT sensory pores. Inset image from Lachner & Karnella (1980).

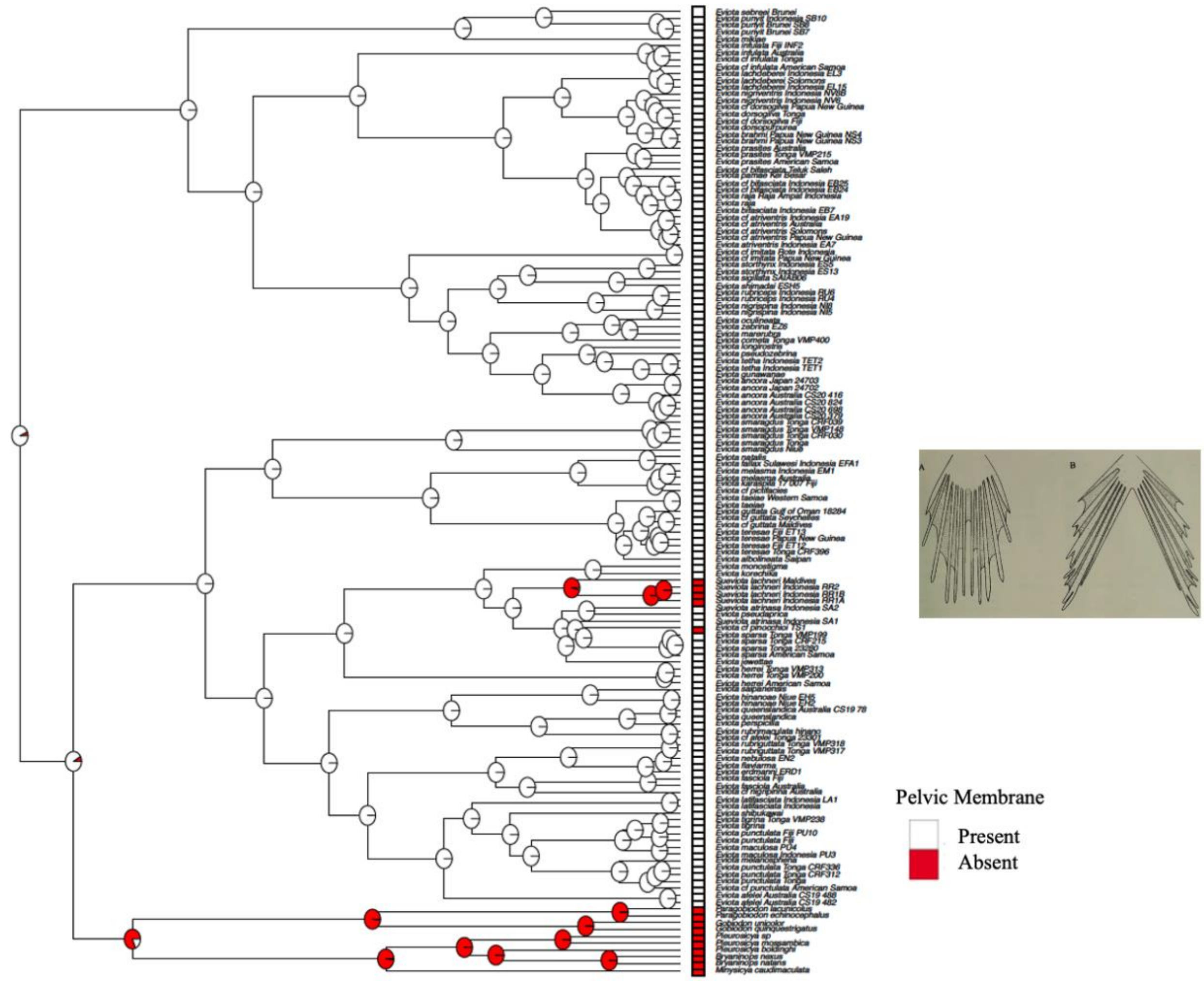


Figure S6. Ancestral state reconstruction for presence/absence of pelvic fin membrane. Inset image from Smith & Baldwin (1999).

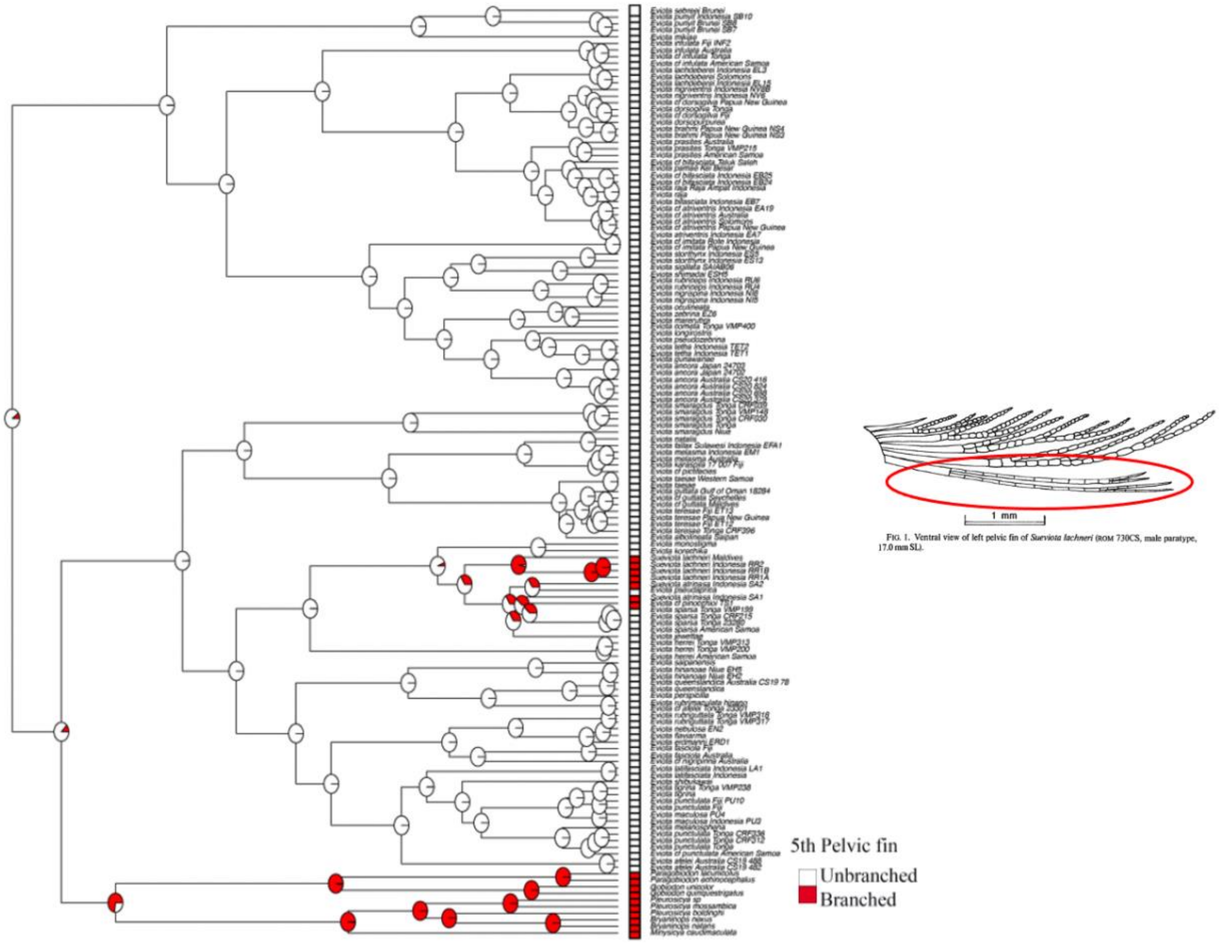


Figure S7. Ancestral state reconstruction for presence/absence of 5th pelvic fin branching. Inset image from Winterbottom & Hoese (1988).

Table S1. Samples used for this study with museum catalog numbers when available.

Family	**Genus**	**Taxa Labels**	**Catalog Num.**
Gobiidae	*Eviota*	*Eviota_afelei_Australia_CS19_482*	forthcoming
Gobiidae	*Eviota*	*Eviota_afelei_Australia_CS19_488*	forthcoming
Gobiidae	*Eviota*	*Eviota_afelei_Tonga_VMP171*	UW 204322
Gobiidae	*Eviota*	*Eviota_afelei_Tonga_VMP172*	UW 204323
Gobiidae	*Eviota*	*Eviota_afelei_Tonga_VMP178*	UW 204329
Gobiidae	*Eviota*	*Eviota_albolineata_Saipan*	forthcoming
Gobiidae	*Eviota*	*Eviota_albolineata_Tonga_CRF396*	UW 204833
Gobiidae	*Eviota*	*Eviota_ancora_Australia_CS20_379*	forthcoming
Gobiidae	*Eviota*	*Eviota_ancora_Australia_CS20_416*	forthcoming
Gobiidae	*Eviota*	*Eviota_ancora_Australia_CS20_698*	forthcoming
Gobiidae	*Eviota*	*Eviota_ancora_Australia_CS20_824*	forthcoming
Gobiidae	*Eviota*	*Eviota_ancora_Japan_24702*	forthcoming
Gobiidae	*Eviota*	*Eviota_ancora_Japan_24703*	forthcoming
Gobiidae	*Eviota*	*Eviota_atriventris_Indonesia_EA4*	CAS 237494

Gobiidae	*Eviota*	*Eviota_atriventris_Indonesia_EA7*	CAS 237494
Gobiidae	*Eviota*	*Eviota_bifasciata_Indonesia_EB2*	CAS 237498
Gobiidae	*Eviota*	*Eviota_bifasciata_Indonesia_EB7*	CAS 237497
Gobiidae	*Eviota*	*Eviota_brahmi_Papua_New_Guinea_NS3*	CAS 236622
Gobiidae	*Eviota*	*Eviota_brahmi_Papua_New_Guinea_NS4*	CAS 236622
Gobiidae	*Eviota*	*Eviota_cf_afelei_Tonga_23301*	UW 203986
Gobiidae	*Eviota*	*Eviota_cf_atriventris_Australia*	forthcoming
Gobiidae	*Eviota*	*Eviota_cf_atriventris_Indonesia_EA19*	forthcoming
Gobiidae	*Eviota*	*Eviota_cf_atriventris_Papua_New_Guinea*	forthcoming
Gobiidae	*Eviota*	*Eviota_cf_atriventris_Solomons*	forthcoming
Gobiidae	*Eviota*	*Eviota_cf_bifasciata_Indonesia_EB24*	forthcoming
Gobiidae	*Eviota*	*Eviota_cf_bifasciata_Indonesia_EB25*	forthcoming
Gobiidae	*Eviota*	*Eviota_cf_bifasciata_Teluk_Saleh*	forthcoming
Gobiidae	*Eviota*	*Eviota_cf_dorsogilva_Fiji*	forthcoming
Gobiidae	*Eviota*	*Eviota_cf_dorsogilva_Papua_New_Guinea*	forthcoming
Gobiidae	*Eviota*	*Eviota_cf_guttata_Maldives*	forthcoming
Gobiidae	*Eviota*	*Eviota_cf_guttata_Seychelles*	forthcoming
Gobiidae	*Eviota*	*Eviota_cf_imitata_Papua_New_Guinea*	forthcoming
Gobiidae	*Eviota*	*Eviota_cf_imitata_Rote_Indonesia*	forthcoming
Gobiidae	*Eviota*	*Eviota_cf_infulata_Tonga*	UW 204292
Gobiidae	*Eviota*	*Eviota_cf_infulata_American_Samoa*	UW 202617
Gobiidae	*Eviota*	*Eviota_cf_nigripinna_Australia*	forthcoming
Gobiidae	*Eviota*	*Eviota_cf_pictifacies*	CAS 237507
Gobiidae	*Eviota*	*Eviota_cf_punctulata_American_Samoa*	UW 202648
Gobiidae	*Eviota*	*Eviota_cometa_Tonga_CRF237*	UW 204750
Gobiidae	*Eviota*	*Eviota_cometa_Tonga_CRF283*	UW 204742
Gobiidae	*Eviota*	*Eviota_cometa_Tonga_VMP400*	UW 204536
Gobiidae	*Eviota*	*Eviota_dorsogilva_Tonga*	UW 203835
Gobiidae	*Eviota*	*Eviota_dorsopurpurea*	CAS 236621
Gobiidae	*Eviota*	*Eviota_erdmanni_ERD1*	forthcoming
Gobiidae	*Eviota*	*Eviota_fallax_Sulawesi_Indonesia_EFA1*	CAS 237502
Gobiidae	*Eviota*	*Eviota_fasciola_Australia*	forthcoming
Gobiidae	*Eviota*	*Eviota_fasciola_Fiji*	CAS 219782
Gobiidae	*Eviota*	*Eviota_flaviarma*	forthcoming
Gobiidae	*Eviota*	*Eviota_flebilis_FB1*	forthcoming
Gobiidae	*Eviota*	*Eviota_gunawanae*	CAS 246254
Gobiidae	*Eviota*	*Eviota_guttata_Gulf_of_Oman_18284*	forthcoming
Gobiidae	*Eviota*	*Eviota_guttata_Gulf_of_Oman_18510*	forthcoming
Gobiidae	*Eviota*	*Eviota_herrei_American_Samoa*	UW 202616
Gobiidae	*Eviota*	*Eviota_herrei_Tonga_CRF249*	UW 204720
Gobiidae	*Eviota*	*Eviota_herrei_Tonga_CRF249B*	UW 204720
Gobiidae	*Eviota*	*Eviota_herrei_Tonga_VMP200*	UW 204337
Gobiidae	*Eviota*	*Eviota_herrei_Tonga_VMP313*	UW 204449
Gobiidae	*Eviota*	*Eviota_hinanoae_Niue_EH2*	forthcoming
Gobiidae	*Eviota*	*Eviota_hinanoae_Niue_EH3*	forthcoming
Gobiidae	*Eviota*	*Eviota_hinanoae_Niue_EH5*	forthcoming
Gobiidae	*Eviota*	*Eviota_infulata_Australia*	forthcoming
Gobiidae	*Eviota*	*Eviota_infulata_Fiji_INF2*	forthcoming
Gobiidae	*Eviota*	*Eviota_jewettae*	forthcoming
Gobiidae	*Eviota*	*Eviota_karaspila_17_007_Fiji*	CAS 248064

Gobiidae	*Eviota*	*Eviota_korechika*	forthcoming
Gobiidae	*Eviota*	*Eviota_lachdeberei_Indonesia_EL15*	forthcoming
Gobiidae	*Eviota*	*Eviota_lachdeberei_Indonesia_EL2*	CAS 237508
Gobiidae	*Eviota*	*Eviota_lachdeberei_Indonesia_EL3*	CAS 237508
Gobiidae	*Eviota*	*Eviota_lachdeberei_Solomons*	forthcoming
Gobiidae	*Eviota*	*Eviota_latifasciata_Indonesia*	forthcoming
Gobiidae	*Eviota*	*Eviota_latifasciata_Indonesia_EE1*	CAS 237510
Gobiidae	*Eviota*	*Eviota_latifasciata_Indonesia_LA1*	forthcoming
Gobiidae	*Eviota*	*Eviota_longirostris*	CAS 246249
Gobiidae	*Eviota*	*Eviota_maculosa_Indonesia_PU3*	CAS 244636
Gobiidae	*Eviota*	*Eviota_maculosa_PU4*	CAS 244635
Gobiidae	*Eviota*	*Eviota_marerubra*	CAS 239043
Gobiidae	*Eviota*	*Eviota_melanosphena*	forthcoming
Gobiidae	*Eviota*	*Eviota_melasma_Australia*	forthcoming
Gobiidae	*Eviota*	*Eviota_melasma_Indonesia_EM1*	forthcoming
Gobiidae	*Eviota*	*Eviota_mikiae*	forthcoming
Gobiidae	*Eviota*	*Eviota_monostigma*	forthcoming
Gobiidae	*Eviota*	*Eviota_natalis*	forthcoming
Gobiidae	*Eviota*	*Eviota_nebulosa_EN2*	CAS 237503
Gobiidae	*Eviota*	*Eviota_nigrispina_Indonesia_NI5*	forthcoming
Gobiidae	*Eviota*	*Eviota_nigrispina_Indonesia_NI8*	forthcoming
Gobiidae	*Eviota*	*Eviota_nigriventris_Indonesia_NV6*	forthcoming
Gobiidae	*Eviota*	*Eviota_nigriventris_Indonesia_NV8B*	CAS 236620
Gobiidae	*Eviota*	*Eviota_occasa_Australia_CS20772L*	forthcoming
Gobiidae	*Eviota*	*Eviota_oculineata*	CAS 246245
Gobiidae	*Eviota*	*Eviota_pamae_Kei_Besar*	CAS 237512
Gobiidae	*Eviota*	*Eviota_perspicilla*	forthcoming
Gobiidae	*Eviota*	*Eviota_prasites_American_Samoa*	UW 202653
Gobiidae	*Eviota*	*Eviota_prasites_American_Samoa_19215*	UW 202697
Gobiidae	*Eviota*	*Eviota_prasites_Australia*	forthcoming
Gobiidae	*Eviota*	*Eviota_prasites_Tonga_CRF213*	UW 204642
Gobiidae	*Eviota*	*Eviota_prasites_Tonga_VMP215*	UW 204352
Gobiidae	*Eviota*	*Eviota_pseudaprica*	forthcoming
Gobiidae	*Eviota*	*Eviota_pseudozebrina*	CAS 246250
Gobiidae	*Eviota*	*Eviota_punctulata_Fiji*	CAS 244638
Gobiidae	*Eviota*	*Eviota_punctulata_Fiji_PU10*	CAS 244639
Gobiidae	*Eviota*	*Eviota_punctulata_Tonga*	UW 203791
Gobiidae	*Eviota*	*Eviota_sebteei_Tonga_VMP397*	UW 204533
Gobiidae	*Eviota*	*Eviota_smaragdus_Tonga_CRF039*	UW 205024
Gobiidae	*Eviota*	*Eviota_punctulata_Tonga_CRF312*	UW 205125
Gobiidae	*Eviota*	*Eviota_punctulata_Tonga_CRF336*	UW 205037
Gobiidae	*Eviota*	*Eviota_punctulata_Tonga_VMP142*	UW 204293
Gobiidae	*Eviota*	*Eviota_punyit_Brunei_SB7*	CAS 238168
Gobiidae	*Eviota*	*Eviota_punyit_Brunei_SB8*	CAS 238168
Gobiidae	*Eviota*	*Eviota_punyit_Indonesia_SB10*	CAS 238170
Gobiidae	*Eviota*	*Eviota_queenslandica*	forthcoming
Gobiidae	*Eviota*	*Eviota_queenslandica_Australia_CS19_78*	forthcoming
Gobiidae	*Eviota*	*Eviota_raja*	CAS 237513
Gobiidae	*Eviota*	*Eviota_raja_Raja_Ampat_Indonesia*	CAS 237513
Gobiidae	*Eviota*	*Eviota_rubriceps_Indonesia_RU4*	forthcoming

Gobiidae	*Eviota*	*Eviota_rubriceps_Indonesia_RU6*	forthcoming
Gobiidae	*Eviota*	*Eviota_rubriguttata_Tonga_VMP317*	UW 204455
Gobiidae	*Eviota*	*Eviota_rubriguttata_Tonga_VMP318*	UW 204456
Gobiidae	*Eviota*	*Eviota_rubrimaculata_hinano*	forthcoming
Gobiidae	*Eviota*	*Eviota_saipanensis*	forthcoming
Gobiidae	*Eviota*	*Eviota_sebreei_Brunei*	CAS 238172
Gobiidae	*Eviota*	*Eviota_sebreei_Indonesia_SB1*	CAS 238172
Gobiidae	*Eviota*	*Eviota_shibukawai*	forthcoming
Gobiidae	*Eviota*	*Eviota_shimadai_ESH5*	forthcoming
Gobiidae	*Eviota*	*Eviota_sigillata_SAIAB06*	SAIAB_77969_6
Gobiidae	*Eviota*	*Eviota_smaragdus_Niue*	forthcoming
Gobiidae	*Eviota*	*Eviota_smaragdus_Tonga_CRF030*	UW 205026
Gobiidae	*Eviota*	*Eviota_smaragdus_Tonga_VMP148*	UW 204299
Gobiidae	*Eviota*	*Eviota_smaragdus_Tonga*	UW 203713
Gobiidae	*Eviota*	*Eviota_sparsa_American_Samoa*	UW 202655
Gobiidae	*Eviota*	*Eviota_sparsa_Tonga_23280*	UW 203965
Gobiidae	*Eviota*	*Eviota_sparsa_Tonga_CRF215*	UW 204734
Gobiidae	*Eviota*	*Eviota_sparsa_Tonga_VMP199*	UW 204336
Gobiidae	*Eviota*	*Eviota_storthynx_Indonesia_ES13*	forthcoming
Gobiidae	*Eviota*	*Eviota_storthynx_Indonesia_ES5*	forthcoming
Gobiidae	*Eviota*	*Eviota_taeiae*	UW 202694
Gobiidae	*Eviota*	*Eviota_taeiae_Western_Samoa*	CAS 248040
Gobiidae	*Eviota*	*Eviota_teresae_Fiji_ET12*	CAS 248061
Gobiidae	*Eviota*	*Eviota_teresae_Fiji_ET13*	CAS 248062
Gobiidae	*Eviota*	*Eviota_teresae_Papua_New_Guinea*	forthcoming
Gobiidae	*Eviota*	*Eviota_tetha_Indonesia_EF2*	CAS 237517
Gobiidae	*Eviota*	*Eviota_tetha_Indonesia_TET1*	CAS 246252
Gobiidae	*Eviota*	*Eviota_tetha_Indonesia_TET2*	CAS 246252
Gobiidae	*Eviota*	*Eviota_tigrina*	UW 203777
Gobiidae	*Eviota*	*Eviota_tigrina_Tonga_CRF463*	UW 204760
Gobiidae	*Eviota*	*Eviota_tigrina_Tonga_VMP238*	UW 204375
Gobiidae	*Eviota*	*Eviota_zebrina_EZ6*	CAS 246251
Gobiidae	*Sueviota*	*Sueviota_atrinasa_Indonesia_SA1*	forthcoming
Gobiidae	*Sueviota*	*Sueviota_atrinasa_Indonesia_SA2*	forthcoming
Gobiidae	*Sueviota*	*Sueviota_lachneri_Indonesia_RR2*	forthcoming
Gobiidae	*Sueviota*	*Sueviota_lachneri_Indonesia_RR1B*	forthcoming
Gobiidae	*Sueviota*	*Sueviota_lachneri_Maldives*	forthcoming
Gobiidae	*Sueviota*	*Sueviota_lachneri_Indonesia_RR1A*	forthcoming
Gobiidae	*Gobiodon*	*Gobiodon_prolixus_American_Samoa_19205*	UW 202700
Gobiidae	*Gobiodon*	*Gobiodon_unicolor*	UW 203801
Gobiidae	*Gobiodon*	*Gobiodon_quinquestrigatus*	UW 203946
Gobiidae	*Paragobiodon*	*Paragobiodon_echinocephalus*	UW 203945
Gobiidae	*Paragobiodon*	*Paragobiodon_lacunicolus*	UW 202679
Gobiidae	*Bryaninops*	*Bryaninops_natans*	UW 203756
Gobiidae	*Minysicya*	*Minysicya_caudimaculata*	UW 203841
Gobiidae	*Pleurosicya*	*Bryaninops_nexus*	UW 203842
Gobiidae	*Pleurosicya*	*Pleurosicya_boldinghi*	UW 203832
Gobiidae	*Pleurosicya*	*Pleurosicya_mossambica*	UW 203873
Gobiidae	*Pleurosicya*	*Pleurosicya_sp*	UW 203872
Gobiidae	*Awaous*	*Awaous_banana*	UW203233

Gobiidae	*Bathygobius*	*Bathygobius_antilliensis*	UW203216
Gobiidae	*Exyrias*	*Exyrias_puntang*	UW156629
Gobiidae	*Coryogalops*	*Coryogalops_anomalus*	UW 158493
Gobiidae	*Risor*	*Risor_ruber*	UW158140
Gobiidae	*Lophogobius*	*Lophogobius_cyprinoides*	UW203247
Gobiidae	*Callogobius*	*Callogobius_sp*	UW 203973
Gobiidae	*Callogobius*	*Callogobius_sp*	UW 203694
Gobiidae	*Asterropteryx*	*Asterropteryx_sp*	UW 203693
Gobiidae	*Fusigobius*	*Fusigobius_neophytus*	UW 203814
Gobiidae	*Valenciennea*	*Valenciennea_sexguttata*	UW 203995
Gobiidae	*Amblygobius*	*Amblygobius_phalaena*	UW 202630
Gobiidae	*Trimma*	*Trimma_benjamini*	UW 203695
Gobiidae	*Trimma*	*Trimma_caesiura*	UW 203991
Gobiidae	*Priolepis*	*Priolepis_semidoliata*	UW 203904
Gobiidae	*Ctenogobiops*	*Ctenogobiops_cf._tongaensis*	UW 203796
Gobiidae	*Macrodontogobius*	*Macrodontogobius_wilburi*	UW 203760
Gobiidae	*Koumansetta*	*Koumansetta_rainfordi*	UW 203699
Gobiidae	*Amblyeleotris*	*Amblyeleotris_steinitzi*	UW 203692
Microdesmidae	*Ptereleotris*	*Ptereleotris_microlepis*	UW 203969
Rhyacichthydae	*Rhyacichthys*	*Rhyacichthys_aspro*	UW202097
Elotridae	*Eleotris*	*Eleotris_premiger*	UW203228
Butidae	*Butis*	*Butis_sp*	UW202152

Chapter 3 Supplementary Data - Available on Dryad (https://doi.org/10.5061/dryad.cnp5hqcdt)

Supplementary File 1. Maximum likelihood tree for the R = 0.3 dataset. Numbers at nodes are bootstrap support values from 100 bootstrap replicates.

Supplementary File 2. Maximum likelihood tree for the R = 0.2 dataset. Numbers at nodes are bootstrap support values from 100 bootstrap replicates.

Supplementary File 3. Bayesian inference tree from the R= 0.4 dataset. Numbers at nodes are posterior probability support values.