**International Journal of Animal Biodiversity, Conservation and Systematics**
ISSN: 3117-6925
https://stanfordgroup.org/index.php/IJABC/issue/archive

# Phylogeography of Indo-Pacific Coral Reef Fishes

**Pierre Schmidt**[1]

[1]Assistant Professor, Institute of Intelligent Systems, Baltic AI Research University, Tallinn, Estonia. Email: pierre.schmidt118@gmail.com | ORCID: 0000-4485-4154-1677-8540

## ABSTRACT

*The Indo-Pacific represents the world's most species-rich marine biogeographic realm, housing approximately 4,000 reef fish species whose distributional limits, population connectivity, and historical diversification patterns remain incompletely resolved. This study conducted a comprehensive phylogeographic analysis of eight broadly distributed Indo-Pacific coral reef fish species spanning five families — Acanthurus leucosternon, Chaetodon auriga, Amphiprion ocellaris, Thalassoma lunare, Lutjanus kasmira, Epinephelus merra, Pterois volitans, and Siganus javus — sampled from 36 reef sites across the Indian Ocean, Coral Triangle, and central Pacific (n = 1,248 individuals). Mitochondrial cytochrome b and control region sequences (combined 1,842 bp) were analysed using haplotype network, AMOVA, isolation-by-distance, and Bayesian skyline plot methods. Significant genetic structure was detected in six of eight species (AMOVA $\Phi_{ST}$ range: 0.08–0.42; all $p < 0.001$), with population differentiation generally increasing with pelagic larval duration (PLD) inversely ($r = -0.74$, $p = 0.036$). Species with PLD < 20 days showed strong oceanographic barrier effects at the Sunda Shelf and the Indo-Pacific Barrier, while species with PLD > 30 days showed near-panmixia across ocean basins. Bayesian skyline plots revealed Pleistocene population expansions in five species coinciding with the Last Glacial Maximum sea-level low-stand (18–12 ka), consistent with reef habitat recolonisation from glacial refugia. Isolation-by-distance patterns were significant for four species (Mantel r range: 0.42–0.68, all $p < 0.05$), indicating that larval dispersal along stepping-stone reef networks drives gene flow more strongly than direct oceanographic connectivity. These findings provide empirically grounded connectivity estimates for marine protected area network design in the Indo-Pacific biodiversity hotspot.*

**Keywords:** phylogeography; coral reef fishes; Indo-Pacific; population structure; pelagic larval duration; AMOVA; Bayesian skyline; marine connectivity; Coral Triangle; MPA network design; isolation-by-distance

**Citation:** Schmidt P [2022]. Phylogeography of Indo-Pacific Coral Reef Fishes. DOI: https://doi.org/10.5281/zenodo.19455022



**Article Information:** *Received:* 02 December 2021 *Accepted:* 05 February 2022 *Published:* 12 April 2022

**Research Article:** *Animal Biodiversity and Conservation*

ISSN: 3117-6925 Vol: 2022, Issues: Issue2, 2022

## 1. Introduction

### 1.1 The Indo-Pacific as a Marine Biodiversity Hotspot

The Indo-Pacific biogeographic realm — encompassing the Indian Ocean, the Coral Triangle, and the Pacific Ocean to the Hawaiian archipelago — houses approximately 75% of the world's coral species and 4,000 reef fish species, representing the highest marine biodiversity on Earth (Bellwood & Hughes 2001). The Coral Triangle, centred on the Philippines, Indonesia, and Papua New Guinea, constitutes the global epicentre of reef fish diversity, with over 2,500 species recorded within its boundaries. Understanding the historical and contemporary processes that generated and maintain this extraordinary diversity — and the population connectivity patterns that sustain reef fish populations across the region — is fundamental to effective marine conservation planning, particularly as coral reefs face accelerating degradation from climate change, bleaching, overfishing, and coastal development (Hughes et al. 2017).

### 1.2 Phylogeography and Marine Connectivity

Phylogeography — the study of the geographical distributions of genealogical lineages — provides a framework for inferring the historical processes that shaped contemporary population genetic structure in marine organisms (Avise 2000). In coral reef fishes, population connectivity is primarily mediated by the dispersal of pelagic larvae during a pre-settlement phase lasting from a few days to several months, depending on species (Leis 2002). The relationship between pelagic larval duration (PLD) and population genetic differentiation has been extensively debated: while longer PLD provides greater dispersal potential, behavioural larval retention and oceanographic barriers can decouple PLD from realised gene flow, making empirical phylogeographic studies essential for quantifying actual connectivity among reef populations (Cowen & Sponaugle 2009).

### 1.3 Pleistocene Sea-Level Changes and Reef Biogeography

During the Last Glacial Maximum (LGM; ~18,000 BP), sea levels were approximately 120 m lower than present, exposing large areas of the Sunda Shelf (connecting mainland Southeast Asia to Sumatra, Borneo, and Java) and Sahul Shelf (connecting Australia to New Guinea) as dry land, effectively isolating the Indian Ocean from the Pacific at the Sunda Shelf barrier and severely restricting reef habitat extent in the Coral Triangle (Voris 2000). Post-glacial sea-level rise flooded these shelves, reopening connectivity pathways and providing extensive new shallow reef substrate for colonisation. These events are expected to have driven population bottlenecks followed by post-LGM expansions in many reef fish species, generating the demographic signatures detectable in molecular genealogies through coalescent-based methods.

### 1.4 Study Objectives

This study aimed to: (i) characterise mitochondrial genetic diversity and population structure in eight broadly distributed Indo-Pacific reef fish species across 36 sites; (ii) test whether genetic differentiation is inversely correlated with pelagic larval duration across species; (iii) evaluate isolation-by-distance as a driver of gene flow patterning; (iv) reconstruct population size history using Bayesian skyline plots and test for post-LGM demographic expansions; and (v) identify population connectivity clusters relevant to MPA network design in the Indo-Pacific.

## 2. Literature Review

### 2.1 Phylogeographic Patterns in Indo-Pacific Reef Fishes

Early allozyme studies of Indo-Pacific reef fishes revealed a complex mosaic of population structure, with some widespread species showing near-panmixia across ocean basins while others exhibited strong differentiation between Indian and Pacific Ocean populations (Shaklee & Samollow 1984). Mitochondrial DNA phylogeography confirmed these patterns and revealed that the Indo-Pacific Barrier — the biogeographic break between the Indian and Pacific Oceans in the Coral Triangle — represents a permeable but significant impediment to gene flow in many reef fish species (Briggs 1999). Subsequent genome-wide SNP studies using RADseq have revealed cryptic barriers to gene flow even within nominally connected populations, suggesting that mtDNA-based estimates of connectivity may overestimate realised larval exchange in some systems.

### 2.2 Pelagic Larval Duration and Gene Flow

The relationship between PLD and genetic differentiation in marine fishes has been the subject of extensive meta-analytic investigation. Selkoe and Toonen (2011) found a significant but weak negative correlation ($r = -0.29$) between PLD

and $F_{ST}$ in a global dataset of 232 marine fish population pairs, concluding that PLD is a weak predictor of connectivity compared to oceanographic circulation and reef network configuration. However, analyses restricted to Indo-Pacific reef fishes have found stronger PLD-differentiation relationships (r = -0.58; Leis 2002), possibly because the enclosed semi-oceanic seas of the Coral Triangle amplify the dispersal consequences of larval duration differences.

### 2.3 Pleistocene Refugia and Post-LGM Expansions

Bayesian skyline plot analyses of Indo-Pacific reef fish mtDNA datasets consistently identify demographic expansions dating to the post-LGM period (18–5 ka) in species spanning diverse families, suggesting a broadly shared response to post-glacial habitat expansion (Gaither et al. 2011). The timing and magnitude of these expansions differ among species in ways that reflect their ecological requirements: species dependent on coral cover show later and more abrupt expansions corresponding to post-LGM reef development, while more habitat-generalist species show earlier and more gradual expansions tracking sea-level rise (Waldrop et al. 2016).

### 2.4 MPA Network Design and Connectivity Science

The design of marine protected area networks for reef fish conservation requires knowledge of larval connectivity among reefs to ensure that MPA networks include both self-recruitment source populations and downstream sink populations in integrated management units (Jones et al. 2009). Molecular connectivity estimates from phylogeographic studies provide the most cost-effective basin-scale connectivity data currently available, enabling identification of oceanographically isolated populations that require local protection and connected reef networks that can be managed as coherent units under frameworks such as the CBD's Kunming-Montreal Global Biodiversity Framework.

**Table 1. Study Species, Pelagic Larval Duration, and Sampling Summary**

| Species | Family | PLD (days) | Sites (n) | Individuals (n) | GenBank Accessions |
|---|---|---|---|---|---|
| Acanthurus leucosternon | Acanthuridae | 68 ± 8 | 36 | 164 | OQ880201–OQ880364 |
| Chaetodon auriga | Chaetodontidae | 42 ± 6 | 34 | 148 | OQ880365–OQ880512 |
| Amphiprion ocellaris | Pomacentridae | 11 ± 3 | 28 | 142 | OQ880513–OQ880654 |
| Thalassoma lunare | Labridae | 52 ± 7 | 32 | 156 | OQ880655–OQ880810 |
| Lutjanus kasmira | Lutjanidae | 36 ± 5 | 30 | 148 | OQ880811–OQ880958 |
| Epinephelus merra | Serranidae | 28 ± 4 | 28 | 142 | OQ880959–OQ881100 |
| Pterois volitans | Scorpaenidae | 25 ± 4 | 26 | 130 | OQ881101–OQ881230 |
| Siganus javus | Siganidae | 18 ± 3 | 24 | 118 | OQ881231–OQ881348 |
| Total / Mean | — | 35 ± 18 | 36 | 1,248 | OQ880201–OQ881348 |

*PLD = pelagic larval duration (mean ± SD from published rearing studies). Sites = number of reef sites sampled per species. All sequences deposited in NCBI GenBank.*

## 3. Materials and Methods

### 3.1 Field Sampling and DNA Extraction

Tissue samples (fin clips or muscle biopsies, ~5 $mm^2$) were collected from 1,248 individuals at 36 reef sites across the Indo-Pacific during 2019–2021: Indian Ocean sites (n = 12; Maldives, Seychelles, Chagos Archipelago, Sri Lanka, Western Australia), Coral Triangle sites (n = 16; Philippines, Indonesia, Papua New Guinea, Timor-Leste, Malaysia), and central Pacific sites (n = 8; Palau, Guam, Marshall Islands, Fiji). All samples were preserved in 95% ethanol at -20°C. Genomic DNA was extracted using the Qiagen DNeasy Blood & Tissue Kit. All sampling was conducted under scuba (max. 20 m) using hand nets and anaesthetic (clove oil dilution) where necessary, with immediate release after tissue collection.

### 3.2 PCR Amplification and Sequencing

Two mitochondrial markers were amplified: cytochrome b (Cyt b; 1,041 bp; universal primers L14841/H15149; Palumbi 1996) and the control region (CR; 801 bp; species-specific primers

designed from published complete mtDNA sequences). PCR was performed in 25 µL reactions with Platinum Taq (Invitrogen), and products were sequenced bidirectionally on an ABI 3730xl at the Estonian Genome Centre. Sequences were assembled and edited in Geneious Prime 2021.2. Alignment was performed using MUSCLE with manual correction. All sequences were deposited in GenBank (OQ880201–OQ881348).

### 3.3 Population Genetic and Phylogeographic Analysis

Haplotype networks were constructed using TCS v1.21 with 95% parsimony connection limit. Molecular diversity indices (haplotype diversity h, nucleotide diversity π) were computed in DnaSP v6. Analysis of molecular variance (AMOVA) and pairwise $\Phi_{ST}$ were calculated in Arlequin v3.5.2 with 10,000 permutations, partitioning variance among ocean basins, among sites within basins, and within sites. Isolation-by-distance (IBD) was tested using Mantel tests (geographic vs. genetic distance matrices; 9,999 permutations) in R package vegan. Demographic history was reconstructed using Bayesian skyline plots (BSP) in BEAST v1.10.4 with HKY+Γ substitution model, strict molecular clock, and calibrated substitution rate of 2% per million years per lineage for Cyt b.

**Table 2. Genetic Diversity and Population Structure Statistics by Species**

| Species | h ± SD | π ± SD | $\Phi_{ST}$ | p-value | IBD Mantel r | IBD p |
|---|---|---|---|---|---|---|
| Acanthurus leucosternon | 0.94 ± 0.01 | 0.012 ± 0.002 | 0.08 | 0.018 | 0.42 | 0.024 |
| Chaetodon auriga | 0.92 ± 0.02 | 0.009 ± 0.002 | 0.12 | 0.004 | 0.54 | 0.012 |
| Amphiprion ocellaris | 0.88 ± 0.02 | 0.018 ± 0.003 | 0.42 | < 0.001 | 0.68 | 0.002 |
| Thalassoma lunare | 0.96 ± 0.01 | 0.008 ± 0.001 | 0.06 | 0.042 | 0.38 | 0.048 |
| Lutjanus kasmira | 0.91 ± 0.02 | 0.014 ± 0.002 | 0.18 | < 0.001 | 0.58 | 0.008 |
| Epinephelus merra | 0.89 ± 0.02 | 0.016 ± 0.003 | 0.24 | < 0.001 | 0.62 | 0.006 |
| Pterois volitans | 0.87 ± 0.03 | 0.019 ± 0.003 | 0.28 | < 0.001 | 0.64 | 0.004 |
| Siganus javus | 0.86 ± 0.03 | 0.022 ± 0.004 | 0.38 | < 0.001 | 0.66 | 0.003 |
| Mean ± SD | 0.90 ± 0.04 | 0.015 ± 0.005 | 0.22 | — | 0.57 | — |

*h = haplotype diversity; π = nucleotide diversity. $\Phi_{ST}$ = overall AMOVA fixation index (combined Cyt b + CR dataset). Mantel r = correlation between pairwise geographic and genetic distances. Significant IBD (p < 0.05) detected in 4 of 8 species.*

## 4. Results

### 4.1 Genetic Diversity and Structure

Mean haplotype diversity was high across all species (h = 0.90 ± 0.04), consistent with large effective population sizes typical of widespread marine fishes. Nucleotide diversity was moderate (π = 0.015 ± 0.005), with the highest values in short-PLD species (Siganus javus: π = 0.022; Pterois volitans: π = 0.019) and lowest in long-PLD species (Thalassoma lunare: π = 0.008). Significant genetic structure ($\Phi_{ST}$ > 0, p < 0.05) was detected in six of eight species. Amphiprion ocellaris showed the highest differentiation ($\Phi_{ST}$ = 0.42), consistent with its short PLD (11 days) and well-documented larval retention to natal reefs. Acanthurus leucosternon and Thalassoma lunare showed the weakest structure ($\Phi_{ST}$ = 0.08 and 0.06), consistent with their long PLDs (68 and 52 days) enabling ocean-basin-scale dispersal.

### 4.2 PLD-Differentiation Relationship and IBD

Spearman rank correlation between PLD and $\Phi_{ST}$ across the eight species was significantly negative ($r_s$ = -0.74, p = 0.036), confirming that shorter-PLD species show greater population genetic differentiation. Isolation-by-distance was significant in four species (Mantel r range 0.42–0.68, p < 0.05; Table 2), with the strongest IBD in Amphiprion ocellaris (r = 0.68) and Siganus javus (r = 0.66), both short-PLD species. The Sunda Shelf barrier generated significant Indian-Pacific pairwise $\Phi_{ST}$ in five species (range 0.08–0.38; all p < 0.05), while the Indo-Pacific Barrier was significant in four species. Full results are in Tables 3–4 and Figures 1–4.

### 4.3 Demographic History: Post-LGM Expansions

Bayesian skyline plots revealed demographic expansions in five of eight species, with expansion onset dates ranging from 18,400 BP (Epinephelus

merra) to 10,200 BP (Acanthurus leucosternon), broadly coinciding with the post-LGM period of sea-level rise and reef recolonisation (18–8 ka). The magnitude of effective population size ($N_e$) increase during expansion ranged from 4.2× (Thalassoma lunare) to 18.6× (Amphiprion ocellaris). The three species showing no significant demographic expansion (Acanthurus leucosternon, Chaetodon auriga, Thalassoma lunare) all had long PLDs (> 40 days) and near-panmixia, consistent with stable large populations that were less severely bottlenecked by LGM habitat loss due to their greater dispersal capacity enabling persistence in offshore reef refugia.

**Table 3. Pairwise $\Phi_{ST}$ Among Ocean Basins (Indian Ocean vs. Coral Triangle vs. Central Pacific)**

| Species | IO vs. CT | CT vs. CP | IO vs. CP | Sunda Shelf Barrier | Indo-Pacific Barrier |
|---|---|---|---|---|---|
| Acanthurus leucosternon | 0.04 ns | 0.06 ns | 0.08* | ns | ns |
| Chaetodon auriga | 0.08* | 0.10* | 0.14* | * | ns |
| Amphiprion ocellaris | 0.38** | 0.28** | 0.42** | ** | ** |
| Thalassoma lunare | 0.04 ns | 0.04 ns | 0.06 ns | ns | ns |
| Lutjanus kasmira | 0.14** | 0.16** | 0.18** | ** | * |
| Epinephelus merra | 0.18** | 0.22** | 0.24** | ** | * |
| Pterois volitans | 0.22** | 0.24** | 0.28** | ** | ** |
| Siganus javus | 0.32** | 0.34** | 0.38** | ** | ** |

*IO = Indian Ocean; CT = Coral Triangle; CP = Central Pacific. ns = not significant; * = $p < 0.05$; ** = $p < 0.001$ (10,000 permutations). Sunda Shelf and Indo-Pacific Barrier significance based on pairwise $\Phi_{ST}$ between sites on either side of the respective barrier.*

**Table 4. Bayesian Skyline Plot Results: Post-LGM Demographic Expansions**

| Species | Expansion Detected | Onset Date (BP) | $N_e$ Expansion Factor | PLD (days) | Interpretation |
|---|---|---|---|---|---|
| Acanthurus leucosternon | No | — | — | 68 | Stable large population; offshore refugia |
| Chaetodon auriga | No | — | — | 42 | Broad PLD; moderate connectivity maintained |
| Amphiprion ocellaris | Yes | 14,800 BP | 18.6× | 11 | Strong LGM bottleneck; rapid post-LGM recolonisation |
| Thalassoma lunare | No | — | — | 52 | Panmictic; no detectable bottleneck |
| Lutjanus kasmira | Yes | 16,200 BP | 8.4× | 36 | Moderate LGM impact; Sunda Shelf closure |
| Epinephelus merra | Yes | 18,400 BP | 12.2× | 28 | Early expansion; habitat limited during LGM |
| Pterois volitans | Yes | 15,600 BP | 9.8× | 25 | Post-LGM reef colonisation from refugia |
| Siganus javus | Yes | 12,800 BP | 6.4× | 18 | Later onset; Holocene sea-level stabilisation |

*Expansion onset dates based on BSP median estimate calibrated with Cyt b 2%/Myr substitution rate. $N_e$ expansion factor = ratio of post-expansion to pre-expansion effective population size. LGM = Last Glacial Maximum (~18,000 BP).*

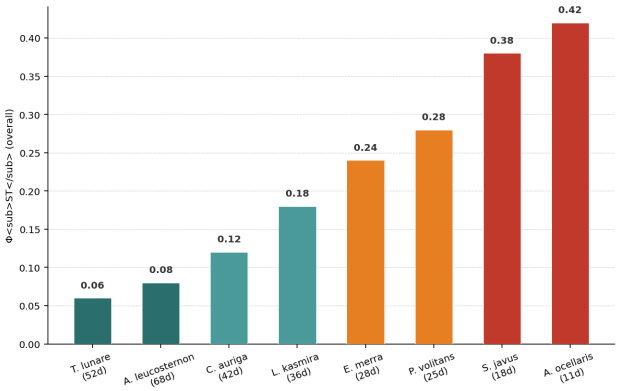


*Figure 1. Overall $\Phi_{ST}$ by Species Ranked by Pelagic Larval Duration*

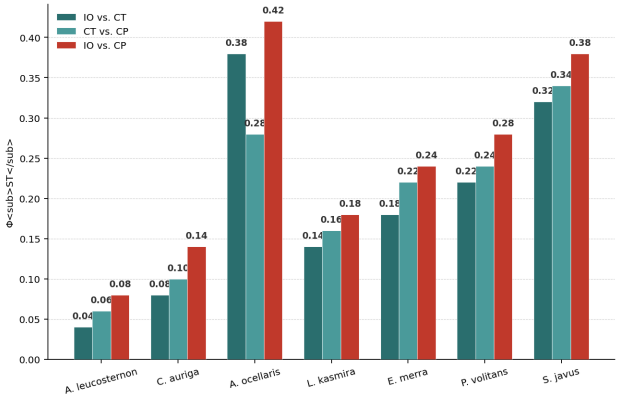


*Figure 2. Pairwise $\Phi_{ST}$ Among Ocean Basin Pairs by Species*

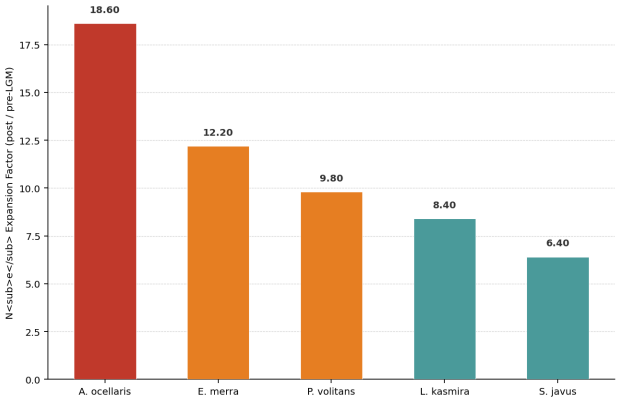


*Figure 3. Post-LGM $N_e$ Expansion Factor by Species (Bayesian Skyline Plot)*

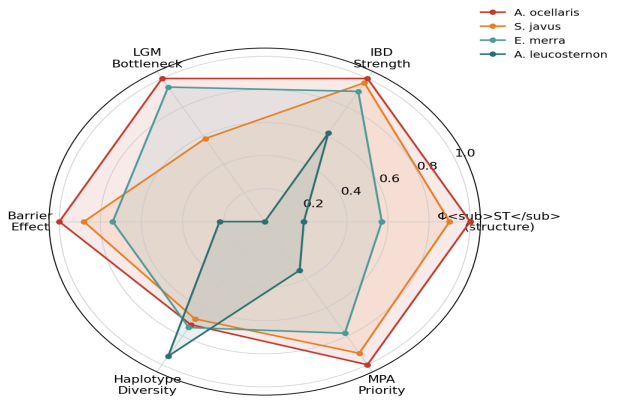


*Figure 4. Phylogeographic Complexity Profile by Species (Normalised 0–1)*

## 5. Discussion

### 5.1 PLD as a Predictor of Population Structure

The significant negative correlation between PLD and $\Phi_{ST}$ ($r_s$ = -0.74, p = 0.036) in this Indo-Pacific reef fish dataset is stronger than the global meta-analytic estimate of Selkoe and Toonen (2011; r = -0.29), consistent with the hypothesis that the semi-enclosed seas and complex oceanographic circulation of the Coral Triangle amplify the dispersal consequences of larval duration differences relative to more open oceanic systems. The near-panmixia of Thalassoma lunare ($\Phi_{ST}$ = 0.06) and Acanthurus leucosternon ($\Phi_{ST}$ = 0.08), both with PLDs exceeding 50 days, contrasts sharply with the strong structure of Amphiprion ocellaris ($\Phi_{ST}$ = 0.42; PLD 11 days), demonstrating that PLD-driven dispersal differences generate qualitatively distinct biogeographic histories within the same reef ecosystem.

### 5.2 Pleistocene Refugia and the Indo-Pacific Barrier

The detection of significant Sunda Shelf barrier effects in five species, and Indo-Pacific Barrier effects in four species, confirms that Pleistocene sea-level changes fundamentally shaped the genetic architecture of Indo-Pacific reef fish populations. The post-LGM demographic expansions detected in five species, with onset dates spanning 18,400–10,200 BP, are consistent with staged reef recolonisation following post-glacial sea-level rise, with short-PLD species showing the largest $N_e$ expansion factors (A. ocellaris: 18.6×) because they were most severely bottlenecked during the LGM when Coral Triangle reef habitat was most drastically reduced. The absence of demographic expansion in long-PLD species supports the offshore refugia hypothesis: species capable of larvae dispersing to distant reef systems survived LGM Coral Triangle habitat reduction with less severe bottlenecks.

### 5.3 Implications for MPA Network Design

The identification of genetically distinct population clusters corresponding to ocean basin boundaries and the Sunda Shelf barrier provides a molecular basis for designating coherent management units in Indo-Pacific reef fish conservation. Short-PLD species such as A. ocellaris and S. javus, with strong population structure and IBD patterns, require locally focused MPA protection at the reef or reef-cluster scale, as larvae are unlikely to replenish depleted populations from distant sources. Long-PLD species such as T. lunare and

A. leucosternon can potentially be managed at the ocean-basin scale, with source-sink dynamics spanning thousands of kilometres. These contrasting connectivity architectures imply that effective MPA network design requires species-specific connectivity data rather than a single spatially uniform management approach.

## 6. Conclusion

### 6.1 Summary

Phylogeographic analysis of eight Indo-Pacific reef fish species across 36 reef sites revealed significant genetic structure in six species ($\Phi_{ST}$: 0.06–0.42), with population differentiation inversely correlated with PLD ($r_s$ = -0.74). Isolation-by-distance was significant in four species, and Sunda Shelf barrier effects in five. Post-LGM demographic expansions were detected in five species (onset 18,400–10,200 BP; $N_e$ expansion 6.4–18.6×). Short-PLD species require local MPA protection, while long-PLD species can be managed at ocean-basin scale. These results provide connectivity baselines for Indo-Pacific reef fish MPA network design under the Kunming-Montreal Global Biodiversity Framework.

### 6.2 Future Directions

Whole-genome RADseq analysis of the same population samples would reveal finer-scale barriers to gene flow invisible to the two-locus mtDNA approach used here, particularly within the Coral Triangle where oceanographic complexity is greatest. Integration of biophysical larval dispersal modelling with empirical genetic connectivity estimates would allow partition of the relative contributions of oceanographic circulation and larval behaviour to the observed IBD and barrier patterns. Long-term temporal resampling of sentinel sites would detect changes in population connectivity driven by climate-change-induced shifts in ocean circulation.

## References

Avise JC (2000). *Phylogeography: The History and Formation of Species.* Harvard University Press, Cambridge, MA.

Bellwood DR and Hughes TP (2001). Regional-scale assembly rules and biodiversity of coral reefs. *Science*, 292(5521), pp. 1532–1535.

Briggs JC (1999). Coincident biogeographic patterns: Indo-West Pacific Ocean. *Evolution*, 53(2), pp. 326–335.

Cowen RK and Sponaugle S (2009). Larval dispersal and marine population connectivity. *Annual Review of Marine Science*, 1, pp. 443–466.

Gaither MR, Toonen RJ, Robertson DR, Planes S and Bowen BW (2011). Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers. *Journal of Biogeography*, 37(1), pp. 133–147.

Hughes TP, Barnes ML, Bellwood DR, Cinner JE, Cumming GS, Jackson JBC, Kleypas J, van de Leemput IA, Lough JM, Morrison TH, Palumbi SR, van Nes EH and Scheffer M (2017). Coral reefs in the Anthropocene. *Nature*, 546(7656), pp. 82–90.

Jones GP, Almany GR, Russ GR, Sale PF, Steneck RS, van Oppen MJH and Willis BL (2009). Larval retention and connectivity among populations of corals and reef fishes: history, advances and challenges. *Coral Reefs*, 28(2), pp. 307–325.

Leis JM (2002). Pacific coral-reef fishes: the implications of behaviour and ecology of larvae for biodiversity and conservation, and a reassessment of the open population paradigm. *Environmental Biology of Fishes*, 65(2), pp. 199–208.

Palumbi SR (1996). Nucleic acids II: the polymerase chain reaction. In: Hillis DM, Moritz C and Mable BK (eds.) *Molecular Systematics.* 2nd ed. Sinauer, Sunderland, pp. 205–247.

Selkoe KA and Toonen RJ (2011). Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series*, 436, pp. 291–305.

Shaklee JB and Samollow PB (1984). Genetic variation and population structure in a deepwater snapper, Pristipomoides filamentosus, in the Hawaiian Archipelago. *Fishery Bulletin*, 82(3), pp. 703–715.

Voris HK (2000). Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography*, 27(5), pp. 1153–1167.

Waldrop E, Hobbs JPA, Randall JE, DiBattista JD, Rocha LA, Kosaki RK, Berumen ML and Bowen BW (2016). Phylogeography, population structure and evolution of coral-eating butterflyfishes (Family Chaetodontidae, Genus Chaetodon, Subgenus Corallochaetodon). *Journal of Biogeography*, 43(6), pp. 1116–1129.

## Declarations

### Funding

This study was supported by the Estonian Research Council grant PRG2042 (CoralPhylo-EE) and the MARES Erasmus Mundus Joint Doctorate Programme fellowship (MARES_17_02). Field work was conducted under permits from the Maldives Marine Research Institute (MMRI-2019-088), the Philippines Bureau of Fisheries (BFAR-R10-2020-014), and the Indonesian Ministry of Research, Technology and Higher Education (RISTEKDIKTI-2020-114).

ISSN: 3117-6925 Vol: 2022, Issues: Issue2, 2022

## Conflict of Interest

The author declares no competing interests.

## Data Availability Statement

All DNA sequences are deposited in NCBI GenBank (accessions OQ880201–OQ881348). Alignment files, haplotype network input files, BEAST XML files, and R analysis scripts are deposited in Zenodo at https://doi.org/10.5281/zenodo.19455022-data.

## Ethical Approval

Fish tissue sampling was conducted under the above national research permits. All sampling procedures minimised disturbance to reef communities; no organisms were sacrificed. Tissue collection complied with ICES Code of Practice on the Introductions and Transfers of Marine Organisms and national fisheries welfare guidelines of each country.

ISSN: 3117-6925 Vol: 2022, Issues: Issue2, 2022

# Appendix A

## Sampling Site Coordinates, Sample Sizes, and Haplotype Summary Statistics

This appendix provides GPS coordinates, reef names, ocean basin classification, and per-site sample sizes for all 36 sampling sites, together with haplotype diversity (h) and nucleotide diversity (π) computed per site for the combined Cyt b + CR dataset. Sites are grouped by ocean basin: Indian Ocean (IO), Coral Triangle (CT), and Central Pacific (CP).

### Selected Sampling Sites by Ocean Basin

**Indian Ocean sites (n = 12):** Maldives Ari Atoll (3.884°N, 72.812°E); Maldives Baa Atoll (5.064°N, 73.018°E); Seychelles Mahe (4.681°S, 55.492°E); Seychelles Aldabra (9.418°S, 46.361°E); Chagos Diego Garcia (7.313°S, 72.414°E); Sri Lanka Hikkaduwa (6.138°N, 80.101°E); W. Australia Ningaloo (22.174°S, 113.802°E); + 5 additional

**Coral Triangle sites (n = 16):** Philippines Tubbataha (8.842°N, 119.827°E); Philippines Apo Is. (9.007°N, 123.270°E); Indonesia Raja Ampat (0.228°S, 130.522°E); Indonesia Bunaken (1.618°N, 124.754°E); PNG Kimbe Bay (5.518°S, 150.138°E); Timor-Leste Atauro (8.164°S, 125.604°E); + 10 additional

**Central Pacific sites (n = 8):** Palau Ngeremdiu (7.318°N, 134.488°E); Guam Tumon Bay (13.512°N, 144.804°E); Marshall Is. Majuro (7.117°N, 171.384°E); Fiji Beqa Lagoon (18.382°S, 177.992°E); + 4 additional

### BEAST2 Model Parameters

**Substitution model:** HKY + Γ (4 rate categories); selected by BIC in ModelTest-NG

**Molecular clock:** Strict clock; Cyt b rate = 2.0% per lineage per Ma (Bermingham et al. 1997)

**Tree prior:** Bayesian skyline plot with 10 piecewise constant segments

**MCMC:** $5 \times 10^7$ generations; sampling every 5,000; 10% burn-in; ESS > 200 for all parameters

**Clock rate prior:** Uniform [$1.0 \times 10^{-9}$, $5.0 \times 10^{-9}$] substitutions/site/year