THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Mapping within-species biodiversity: Patterns and predictors of genetic diversity on Indo-Pacific reefs.

Jude Keyse
BSc

A thesis submitted for the degree of Doctor of Philosophy at
The University of Queensland in 2014
School of Biological Sciences

ABSTRACT

The rate of biodiversity loss across the globe is alarming, suggesting a need to better understand the conditions favourable for generating biodiversity, especially in hyperdiverse regions. The shallow waters of the Indo-Pacific contain the highest concentration of tropical marine species on the planet; not surprisingly, the region has been the focus of extensive study by biogeographers since the 19th century. The Indian and Pacific Oceans border over 65 nations, of which 18 are classified by the UN as Least Developed Countries, with burgeoning human populations along their coastlines. These pressures make understanding the patterns and processes underlying the generation and maintenance of all levels of biodiversity a pressing need.

Compared to species diversity, genetic diversity is rarely considered in conservation planning. Genetic diversity data are commonly published as studies of a single or few species, particularly phylogeographic or population genetic studies focusing on spatial patterns within species. Once published, however, these public data can be used to answer questions on a larger spatial scale than the scale of their constituent parts. In a quantitative, comparative framework, these data can be synthesised to address questions about the bewildering diversity of the Indo-Pacific by treating species as 'replicates'. Yet, use of these data carries certain caveats. In particular, data extraction requires significant time and often returns summary statistics rather than raw genetic data. This thesis explores the availability of such data for tropical marine fauna in the Indo-Pacific and makes use of it to examine patterns of genetic diversity at an oceanic scale.

The first two chapters of this thesis use publicly available data from 108 studies for 116 species of marine fauna in the tropical Indo-Pacific. In Chapter 2, I discuss the extent and scope of currently published genetic data for Indo-Pacific marine fauna, highlighting its strengths and omissions. I show that there is a distinct bias towards studies reporting genetic diversity for marine fishes compared to marine invertebrates; that most studies focus on a single species and that there is little coherence across regions in terms of species studied. I find that, generally, regions are not consistently co-sampled with neighbouring regions. For example, the same species is not often sampled from the Great Barrier Reef as well as from Indonesia, immediately to the North. I identify locations from where a disproportionately high number of species have been the focus of genetic studies, which may serve as useful 'anchor' locations for researchers to build a network of sampling locations. In an appropriate collaborative framework this could allow the linking of data across very broad spatial extents to explore commonalities across multiple species.

The pattern of decreasing species richness of tropical marine taxa with distance from the Coral Triangle is well known. A concordant pattern has been suggested for genetic diversity, however this concordance has not been rigorously tested. In the third chapter I assess the correlation between genetic diversity and species richness of reef fauna and hard corals in the Indo-Pacific. I expected a positive correlation, given the similar processes that govern the spatial distribution of both levels

of diversity: migration, extinction/drift and speciation/mutation. By using so many species to assess this question I was able to see past some of the interspecific variation in genetic diversity that has clouded the waters in previous attempts. I find a positive, albeit weak, correlation between species richness of reef fishes or hard corals and within-species genetic diversity of shallow-water marine species.

Finally, in Chapter 4, I focus on a discrete species group to illustrate the advantages of combining data from multiple research groups with focused sampling to unite disconnected geographic regions in a single analysis. Here I chose three codistributed species of giant clams (*Tridacna maxima, T. crocea* and a cryptic species here referred to as *T.* sp.) to investigate the factors underlying the spatial genetic diversity within species in the Indo-Pacific. (The paper reporting our discovery of this cryptic species in included as an Appendix to this thesis). By combining new data from the West Pacific with existing data I can reveal the relative strength of the Torres Strait landbridge as a barrier to gene flow within giant clams, when compared to other barriers in the region. My analyses of population structure show that there is strong population structure among regions in *Tridacna*. This has significant consequences for the management of populations but also provides further evidence for the central Indo-Pacific being a region of sympatry among divergent clades.

In summary, this thesis provides a statement on the problems inherent in data synthesis and provides guidance on best practice to facilitate ease of open data access; makes use of published genetic data to test one of the big questions at the interface of ecology and evolution: the species-genetic diversity correlation; and demonstrates the utility of combining raw sequence data across lab groups to allow robust assessment of both broad and fine scale patterns of genetic diversity within a group of conservation concern, the giant clams.

DECLARATION BY AUTHOR

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

PUBLICATIONS DURING CANDIDATURE

Peer-reviewed papers

Keyse J, Crandall ED, Toonen RJ, Meyer CP, Treml EA, Riginos C. 2014. The scope of published population genetic data for Indo-Pacific marine fauna and future research opportunities in the region. Bulletin of Marine Science 90(1):47-78.

Huelsken T*, Keyse J*, Liggins L, Penny S, Treml EA, Riginos C. 2013. A Novel Widespread Cryptic Species and Phylogeographic Patterns within Several Giant Clam Species (Cardiidae: Tridacna) from the Indo-Pacific Ocean. PLoS ONE 8(11):e80858.
* indicates joint first authors

Fadli N, Campbell S, Ferguson K, Keyse J, Rudi E, Riedel A, Baird A. 2012. Quantifying change in the community structure of an artificial reef. Oryx 46(4):501 507.

Pope LC, Liggins L, Keyse J, Carvalho SB, Riginos C (2015) Not the time or the place: the missing spatio-temporal link in publicly available genetic data. Molecular Ecology. DOI 10.1111/mec.13254

Conference abstracts

Keyse J, Riginos C, Treml EA, Rhodes JR. 2013. WHAT ARE WE MISSING? Patterns and predictors of genetic diversity on Indo-Pacific reefs. International Biogeography Society Early Career Conference. Canberra, Australia. (Poster and 5 minute talk)

Keyse J, Treml EA, Rhodes JR, Riginos C. 2013. Patterns and predictors of multi-species genetic diversity in the marine Indo-Pacific. Evolution 2013. Snowbird, Utah, USA. (12 minute talk)

Keyse J, Riginos C, Rhodes JR, Treml EA. 2013. Taking stock: the scope and utility of published population genetic data for marine Indo-Pacific fauna. The Association of Tropical Biology and Conservation, Asia Pacific Chapter Meeting: Linking biodiversity science to policy and conservation action. Banda Aceh, Indonesia. (12 minute talk)

Keyse J, Crandall ED, Toonen RJ, Meyer CP, Treml EA, Riginos C. 2013. WHAT ARE WE MISSING? Beyond species richness... Student Conference on Conservation Science. Brisbane, Australia. (Poster)

Keyse J, Rhodes JR, Treml EA, Riginos C. 2012. Beyond species richness: testing concordant patterns of species and genetic diversity in Indo-Pacific marine fauna. Ecological Society of Australia 2012 Annual Conference: Ecology: Fundamental Science of the Biosphere. Melbourne, Australia. (12 minute talk)

PUBLICATIONS INCLUDED IN THIS THESIS

Keyse J, Crandall ED, Toonen RJ, Meyer CP, Treml EA, Riginos C. 2014. The scope of published population genetic data for Indo-Pacific marine fauna and future research opportunities in the region. Bulletin of Marine Science 90(1):47-78.

Incorporated as Chapter 2.

Contributor	*Statement of Contribution*
Keyse, J	Designed the study (70%), collected data, analysed data, wrote paper
Crandall, ED	Contributed to discussion of concepts (20%), advised JK on analysis
Toonen RJ	Contributed to discussion of concepts (20%), edited paper significantly
Meyer CP	Contributed to discussion of concepts (20%), edited paper significantly
Treml EA	Contributed to discussion of concepts (20%)
Riginos C	Contributed to discussion of concepts (20%), designed the study (30%)

CONTRIBUTIONS BY OTHERS TO THE THESIS

Jonathan Rhodes wrote the R code to perform the analysis for Chapter 3.

Eric Treml supplied predictions from a biophysical model of larval dispersal for Chapter 4.

Cynthia Riginos wrote a Perl script for editing sequences for Chapter 4 and R code for some analyses in Chapter 2.

STATEMENT OF PARTS OF THE THESIS SUBMITTED TO QUALIFY FOR THE AWARD OF ANOTHER DEGREE

None

// ACKNOWLEDGEMENTS

Several individuals and organisations supported me through the process of completing my PhD. Below I describe some of the reasons for which I owe them all gratitude and respect.

I am extremely grateful to Dr Cynthia Riginos, my primary advisor, for her continued support and confidence in me. Dr Riginos never failed to highlight the positive aspects of each stage of the process while keeping me focussed on the key goals. Dr Jonathan Rhodes and Dr Eric Treml, my secondary advisors, provided excellent support and guidance in spatial and statistical aspects of my work.

I am grateful to Dr Arthur Riedel for many years of encouragement and support, not to mention his excellent guidance with print design.

The Riginos Lab members, past and present, were always supportive and kind in their criticisms of my work. Dr Tom Huelsken trained me in molecular lab work, Dean Blower talked me through a number of coding issues, Carly Martin assisted me in the lab.

Members of the UQ Kung Fu Club helped me unwind after long days at my desk and always inspired me to keep up the forward motion.

Monetary support from the University of Queensland (UQRS and UQI scholarships and GSITA), the Ecological Society of Australia, Australian Geographic, The Linneaen Society of New South Wales and The Malacological Society of Australasia was invaluable in allowing me to carry out the work contained in this thesis.

Finally, I thank David Colbert who has changed the colour of the world and proven his worth by his patience and understanding over the past five months.

KEYWORDS
genetic diversity, biogeography, phylogeography, multi-species, Bayesian, open-access, data synthesis

AUSTRALIAN AND NEW ZEALAND STANDARD RESEARCH CLASSIFICATIONS (ANZSRC)

ANZSRC code: 050104 Landscape Ecology, 30%
ANZSRC code: 060302 Biogeography and Phylogeography, 40%
ANZSRC code: 060411 Population, Ecological and Evolutionary Genetics, 30%

FIELDS OF RESEARCH (FOR) CLASSIFICATION

FoR code: 0501 ECOLOGICAL APPLICATIONS, 30%
FoR code: 0603 EVOLUTIONARY BIOLOGY, 40%
FoR code: 0604 GENETICS, 30%

TABLE OF CONTENTS

Abstract	2-3
Declaration by author	4
Publications during candidature	5
Publications included in this thesis	6
Contributions by others to the thesis	6
Statement of parts of the thesis submitted to qualify for the award of another degree	6
Acknowledgements	7
Keywords	8
Australian And New Zealand Standard Research Classifications (ANZSRC)	8
Fields Of Research (FoR) Classification	8
List of Figures and Tables	11-17
List of Abbeviations used in the thesis	18-19

Chapter 1. **20-41**

General Introduction

Chapter 2. **41-81**

The scope of published population genetic data for Indo-Pacific marine fauna and future research opportunities in the region.

Abstract	42
Introduction	43-44
Methods	44-45
Results & Discussion	45-55
Recommendations for future work	55-58
Acknowledgements	58
Literature Cited and Appendix	58-81

Chapter 3. **82-114**

Multi-species approach to testing the species-genetic diversity correlation in Indo-Pacific reef animals.

Abstract 82
Introduction 82-85
Methods 85-90
Results 90-91
Discussion 91-97
Literature Cited and Supplementary Material 98-103

Chapter 4. **115-155**

Strong signature of the Torres Strait landbridge on population structure in three codistributed species of *Tridacna* giant clams.

Abstract 115
Introduction 115-121
Methods 121-126
Results 127-140
Discussion 136-139
Literature Cited and Supplementary Material 139-160

Chapter 5. **161-170**

General Discussion

Appendices **171-189**

Huelsken T*, Keyse J*, Liggins L, Penny S, Treml EA, Riginos C. 2013. A Novel Widespread Cryptic Species and Phylogeographic Patterns within Several Giant Clam Species (Cardiidae: Tridacna) from the Indo-Pacific Ocean. PLoS ONE 8(11):e80858.
* indicates joint first authors

Pope LC, Liggins L, Keyse J, Carvalho SB, Riginos C (2015) Not the time or the place: the missing spatio-temporal link in publicly available genetic data. Molecular Ecology. DOI 10.1111/mec.13254

LIST OF FIGURES & TABLES

Chapter 1.

General Introduction

Figure 1. 28
Map of the Indo-Pacific showing the Coral Triangle. Grey outline shows the approximate coastline during Pleistocene lower sea levels 250,000 - 18,000 years ago.

Chapter 2.

The scope of published population genetic data for Indo-Pacific marine fauna and future research opportunities in the region

Figure 1. 46
Histogram showing the number of taxa studied by the four categories of molecular marker type.

Figure 2. 48
Summary of sampling for genetic surveys included in the present study. Total area surveyed (km2) and the number sites survey are indicated per species

Text Box. 49-50
The top five species with the greatest sampling extent and the top five species with the most locations sampled.

Figure 3. 52
(A) Sampling intensity for the 116 species surveyed. A heatmap coloured by the proportion of studied species is shown per site with a correction for species range. For example, from the main Hawaiian Islands, 22 species have been surveyed and 39 species from the 116 in the data set have species ranges that encompass this location, which gives a percentage of 56.4. (B) Sampling intensity for the 116 species corrected for the area of the study locality. As the locality polygons are of different area depending on the proximity of sampling locations, this correction allows us to see intensity of sampling per unit area. For example, the Main Hawaiian Islands locality has an area of 69,063 km2, so the corrected sampling intensity is 56.4 / 69,063 or 0.0008.

Figure 4. 53
Analysis of species co-sampling. Cluster dendrogram is based on squared Euclidean distances among sampling localities, derived from the composition of species that have been co-sampled in each locality. Localities with a higher number of co-sampled species have a lower Euclidean distance between them. Colors show

the geographic spread of clusters of co-sampled localities across the Indo-Pacific region. Only localities where more than five species have been surveyed are shown on the map.

Appendix 1. 77-81
All species included in the dataset, with their citations and the genetic marker used in the study. Names in black are those species for which species authority was derived from the Integrated Taxonomic Information System, # represents those derived from the World Register of Marine Species and ‡ represents those derived from Fishbase.

Chapter 3.

Multi-species approach to testing the species-genetic diversity correlation in Indo-Pacific reef animals.

Figure 1. 87
The study region boundary is shown with sampling intensity of genetic diversity data points.

Figure 2. 87
Estimated species richness within six families of marine taxa (Acanthuridae, Labridae, Chaetodontidae, Epinephelidae, Elapidae and Pomacanthidae) represented as a heatmap. Red indicates the highest and green is the lowest species richness.

Table 1. 90
Posterior estimates from each SGDC beta regression model. Mean values are averaged across all species. % ssdiff >0 indicates good model fit at ≥50%. Support for the full model is derived from DIC scores lower than the null model where the slope is fixed at zero. MC Error is a measure of the accuracy of the Markov Chain used to infer posterior estimates.

Figure 3. 91
Graphical plots of the predicted genetic diversity across all species (μ) as a function of species richness. The solid line shows the mean posterior estimate of the relationship between haplotype diversity (A), nucleotide diversity (B) and heterozygosity (C); and standardised species richness. 95% credible limits are shown as dashed lines.

Figure 4. 92
Individual species mean SGDC slopes for sequence-based diversity indices, haplotype diversity (A and B) and nucleotide diversity (C and D). Species are divided into reef fishes (A and C) and invertebrates (B and D) for ease of viewing. Colours are the same for species across the upper and lower graphs.

Figure 5. 93
Individual species mean SGDC slopes for heterozygosity calculated from microsatellites. All

species are shown on a single axis. Colours for species also occurring in Fig. 4 are the same.

Figure 6. 94
Correlograms of the correlation among model residuals at 100 km distance classes (horizontal axis). Correlation among residuals for models predicting A. haplotype diversity, B. nucleotide diversity and C. expected heterozygosity is displayed as correlation coefficients on the vertical axis. Filled circles indicate significant correlations at $p < 0.05$ (Moran's I statistic using 1000 permutations).

Supplementary Table 1. 104
Species used in haplotype diversity analyses with number of populations sampled (N) and details of the spatial autocorrelation among haplotype diversity within species (Moran's I, p value). Citation numbers refer to Supp. Table 7.

Supplementary Table 2. 105
Species used in nucleotide diversity analyses with number of populations sampled (N) and details of the spatial autocorrelation among haplotype diversity within species (Moran's I, p value). Citation numbers refer to Supp. Table 7.

Supplementary Table 3. 106
Species used in expected heterozygosity analyses with number of populations sampled (N) and details of the spatial autocorrelation among heterozygosity within species (Moran's I, p value). Citation reference numbers refer to Supp. Table 7.

Supplementary Table 4. 107
Parameter values (mean, median, upper and lower credible limits and standard deviation) for 61 species from the SGDC beta regression models predicting haplotype diversity. Shaded rows indicate species analysed in Messmer et al (2012).

Supplementary Table 5. 108
Parameter values (mean, median, upper and lower credible limits and standard deviation) for 57 species from the SGDC beta regression models predicting nucleotide diversity. Shaded rows indicate species analysed in Messmer et al (2012).

Supplementary Table 6. 109
Parameter values (mean, median, upper and lower credible limits and standard deviation) for 22 species from the SGDC beta regression models predicting expected heterozygosity from microsatellites with species richness.

Supplementary Table 7. 110-114
Full bibliography of studies from which genetic diversity data were extracted for use in this study. Ref. No. refers to study number in Supp. Tables 1-3.

Chapter 4.

Strong signature of the Torres Strait landbridge on population structure in three codistributed species of *Tridacna* giant clams.

Figure 1. 116

The Indo-Australian Archipelago and central Pacific; dominant currents are in blue, (adapted from Benzie and Williams 1997): NECC (North Equatorial Counter Current), SECC (South Equatorial Counter Current), NGCC (New Guinea Coastal Current), ITF (Indonesian Throughflow). Putative barriers to gene flow are in black lines: HE (Halmahera Eddy) and TS (Torres Strait). The light grey outline shows the 120 metre isobath: the approximate coastline during the lowest Pleistocene sea levels. The dashed square marks the extent of the biophysical model used to estimate larval dispersal.

Figure 2. 117

Sampling of *Tridacna maxima* for this study is shown in red and existing data are shown in black dots. The species range (light blue) is adapted from (Lucas 1988). Lighter grey outline indicates the 120m isobath: the approximate coastline during the lowest Pleistocene sea levels. Sampling location names can be found in Supp. Table 1, along with sources for existing data.

Figure 3. 118

Sampling of *Tridacna crocea* for this study is shown in red and existing data are shown in black dots. The species range (light blue) is adapted from (Lucas 1988). Lighter grey outline indicates the 120 m. isobath: the approximate coastline during the lowest Pleistocene sea levels.Sampling location names can be found in Supplementary Table 1, along with sources for existing data.

Figure 4. 120

Sampling of *Tridacna* sp. for this study is shown in red and populations where fewer than 6 individuals were sampled are shown as small black dots. Lighter grey outline indicates the 120 m. isobath: the approximate coastline during the lowest Pleistocene sea levels. Sampling location names can be found in Supplementary Table 1, along with sources for existing data.

Figure 5a. 128

Map showing *Tridacna maxima* sampling locations, coloured by clade identity with reference to (5b). Lighter grey outline indicates the 120m isobath: the approximate coastline during the lowest Pleistocene sea levels.

Figure 5b. 129

Median joining haplotype network from a 385bp fragment of COI for *T. maxima.* Nodes represent unique haplotypes, sized according to their frequency. Singleton haplotypes were removed to ease viewing. Perpendicular marks on network edges represent single nucleotide polymorphisms separating haplotypes from each other.

Figure 6. 130

Map showing sampling locations for *T. crocea*, coloured by clade identity of individuals collected there. Lighter grey outline is the coastline during the last glacial maximum. Below, median joining haplotype network from a 332 bp fragment of COI. Unique haplotypes are represented by nodes, the size of which is proportional to haplotype frequency. Singleton haplotypes were removed to ease viewing. Perpendicular marks on network edges indicate the number of mutational steps separating haplotypes.

Figure 7. 131

Map shows sampling locations for *Tridacna* sp., coloured to denote clade identity of individuals found at that location. Lighter grey outline is the coastline during the last glacial maximum. Below, median joining haplotype network from a 353 bp fragment of COI for *T*. sp. Unique haplotypes are represented by nodes, the size of which is proportional to haplotype frequency. Perpendicular marks on network edges indicate the number of mutational steps separating haplotypes..

Table 1. 132

Marginal tests of the effects of continuous predictors within *T. crocea*, *T. maxima* and *T*. sp. in the context of other, competing predictors. Degrees of freedom indicate the number of significant axes in the PCNM of the distance matrix. Marg. Var. is the variance in Φ_{ST} explained by the predictor after all others have been fitted.

Table 2. 133

Marginal tests on best models of the effects of continuous distances on Φ_{ST} in *T. crocea* and *T. maxima*. Degrees of freedom indicate the number of significant axes in the PCNM of the distance matrix. Marg. Var. is the variance in Φ_{ST} explained by the predictor after others in the model have been fitted.

Table 3. 134

Marginal tests of the effects of continuous predictors within clade 2 of *T. crocea* and clade 3 of *T. maxima* in the context of other, competing predictors. Degrees of freedom indicate the number of significant axes in the PCNM of the distance matrix. Marg. Var. is the variance in Φ_{ST} explained by the predictor after all others have been fitted.

Table 4. 135

Marginal tests of the relative effects of the Torres Strait and Halmahera barriers on Φ_{ST} in *T. crocea* and *T. maxima* when conditioned on (considered in the context of) each continuous distance measure separately. Degrees of freedom indicate the number of significant axes in the PCNM of the distance matrix. Marg. Var. is the variance in Φ_{ST} explained by the barrier after others in the model have been fitted.

Supplementary Figure 1. 147
Isolation by distance plots of ΦST predicted by modern overwater distance (A, C, E) and LGM overwater distance (B, D, F) for *T. maxima* (A, B); *T. crocea* (C, D) and *T. sp*. (E, F).

Supplementary Figure 2. 148
Isolation by distance plots of Φ_{ST} within clade 3 of *Tridacna maxima* (A, B) and clade 3 of *Tridacna crocea* (C, D) predicted by modern overwater distance (A, C) and overwater distance at the last glacial maximum (B, D).

Supplementary Table 1(a) 149
Summary genetic statistics for populations of *Tridacna maxima.* Bold values indicate significant values after Bonferroni correction to $p = 0.0011$.

Supplementary Table 1(b) 150
Summary genetic statistics for populations of *Tridacna crocea.* Bold values indicate significant values after Bonferroni correction to $p = 0.0011$.

Supplementary Table 1(c) 151
Summary genetic statistics for populations of *Tridacna sp.* Bold values indicate significant values after Bonferroni correction to $p = 0.0100$.

Supplementary Table 1(d) 152
Summary genetic statistics for populations of *Tridacna maxima* before Bonferroni correction.

Supplementary Table 1(e) 153
Summary genetic statistics for populations of *Tridacna crocea* before Bonferroni correction.

Supplementary Table 1(f) 154
Summary genetic statistics for populations of *Tridacna sp.* before Bonferroni correction.

Supplementary Table 2 (a) 155-156
Inter-population genetic differentiation at COI for *Tridacna maxima.* Φ_{ST} shown above diagonal, F_{ST} below diagonal. Bold values indicate significance at Bonferroni corrected $p = 0.01$

Supplementary Table 2 (b) 157-158
Inter-population genetic differentiation at COI for *Tridacna crocea.* Φ_{ST} shown above diagonal, F_{ST} below diagonal. Bold values indicate significance at Bonferroni corrected $p = 0.01$

Supplementary Table 2 (c) 159
Inter-population genetic differentiation at COI for *Tridacna sp.* Φ_{ST} shown above diagonal, F_{ST} below diagonal. Bold values indicate significance at Bonferroni corrected $p = 0.01$

Supplementary Table 3 160

Results of univariate dbRDA models predicting Φ_{ST} in *T. crocea, T. maxima* and *T.*sp using all principle coordinates of continuous predictor distance matrices. The degrees of freedom (df) give information on the number of axes of variation in the variable. Significant predictors from these tests were used in comparative models to assess the relative influence of each predictor (Table 3).

LIST OF ABBREVIATIONS USED IN THE THESIS

π	Nucleotide diversity
BC	Before Christ
BOLD	Barcode of Life Data Systems
CBD	Convention on Biological Diversity
CITES	Convention on International Trade in Endangered Species
COI	Cytochrome oxidase subunit one
CI	Credible interval
CnidTOL	Cnidarian Tree of Life Project
Cyt. B	Cytochrome B.
dbRDA	Distance Based Redundancy Analysis
DDBJ	DNA Data Bank of Japan
DIC	Deviance Information Criterion
DNA	Deoxyribonucleic acid
EMBL	European Molecular Biology Laboratory
FishBOL	Fish Barcode of Life Initiative
GBIF	Global Biodiversity Information Facility
GBR	Great Barrier Reef
h	Haplotype diversity
H_E	Expected heterozygosity

H_O	Observed heterozygosity
ITF	Indonesian Through Flow
IUCN	International Union for Conservation of Nature
LGM	Last Glacial Maximum
MarBOL	Marine Barcode of Life Initiative
MCMC	Markov Chain Monte Carlo
MRDM	Multiple Regression on Distance Matrices
mtDNA	Mitochondrial DNA
NCBI	National Centre for Biotechnology Information
NESCent	National Evolutionary Synthesis Centre
OECD	Organisation for Economic Co-operation and Development
PCNM	Principle Coordinates of Neighbourhood Matrix
PCR	Polymerase Chain Reaction
pers. obs.	Personal Observation
PLD	Pelagic larval duration
PNG	Papua New Guinea
SGDC	Species-Genetic Diversity Correlation
SNP	Single nucleotide polymorphism
UN	United Nations
WoRMS	World Register of Marine Species
XML	Extensible Markup Language

CHAPTER 1

General Introduction

Life is not evenly distributed across the planet. Habitats such as rainforests and coral reefs are repositories of numerous varieties of life, while deserts and the vast sediments of the sea floor contain a comparatively depauperate fauna and flora. Globally, a pattern of decreasing species richness with increasing latitude has been the subject of numerous studies, generating many hypotheses to explain its near ubiquity across taxonomic groups (Pianka 1966, Rohde 1992, Willig et al. 2003). In the marine environment two centres of species diversity have been identified: the Caribbean and the Coral Triangle (Roberts et al. 2002b). These strong patterns beg several questions: Why do we see such strong spatial gradients in the diversity of life? What processes are generating and maintaining such diversity? Can we predict where we will find high levels of diversity? These questions date back to early botanists and collectors who noted the similarities and differences in the biodiversity between continents as they sailed between continents for the first time (Von Humboldt 1849, Darwin 1859, Wallace 1876). These early scientists laid the groundwork for the field of biogeography, the study of the geography of life. Similar questions concern biogeographers today, despite much of our modern exploration being *in silico* rather than at sea. Today, however, our questions are more pressing as we begin to recognise the declining quality of our environment (Laurance 2010) and an accelerating extinction rate in many taxonomic groups (Barnosky et al. 2011). Many modern biogeographers seek answers to these questions predominantly with a view to halting the loss of what inspired our forebears.

Despite its recent rapid loss, biodiversity (the variety of life) can still be overwhelming in the sheer volume of matter it encompasses. In an effort to simplify our thoughts about this mass of biotic complexity, Norse et al. (1986) categorised biodiversity into hierarchical levels; from the most fundamental level of variation between members of the same species, through the differences between species to deeper levels of the taxonomic hierarchy, and finally to the variation among collections of species in the form of ecosystems. This system of organisation mimics the taxonomical hierarchy, with the exception of ecosystem-level biodiversity, and formed the basis for the Convention on Biological Diversity (CBD, Rio de Janeiro, 1992), a document setting goals for halting biodiversity loss. While this approach has, arguably, led to a focus on the conservation of discrete units rather than the processes underlying variation (Bowen 1999), the CBD was a powerful signal in conservation. The signing of this document marked the first time that the value of biodiversity, and the need to safeguard its future, was recognised in the political sphere (Magurran 2004). The CBD aimed to make conservation of biodiversity, in the units of genes, species and ecosystems, a legal requirement for signatories. This meant that reliable and meaningful quantification of biodiversity became important to assess compliance.

Species-centric measurement of biodiversity

Overwhelmingly, we measure biodiversity in the units of species (Magurran 2004). Species are generally the types we humans most readily recognise (Mayr 1963), though this generality depends upon the species concept. Species concepts are many and varied in their assumptions and applicability (De Queiroz 2007). The choice of which to use is commonly driven by necessity and familiarity, i.e. field biologists may use the Phenetic Species Concept (Michener 1970) or the Ecological Species Concept (Van Valen 1976) while evolutionary biologists might use a Phylogenetic Species Concept (Donoghue 1985). Whatever species concept is used, species are the most common currency for measurement of biodiversity. Most biodiversity management plans under the CBD are targeted at the species level (Laikre et al. 2009).

Diversity can be divided into *alpha*, *beta* and *gamma* components; *alpha* describing the diversity within a site, *beta* measuring the differences in composition among sites, and *gamma* referring to the broad spatial differences among biogeographic regions (Whittaker 1972, Brown 1989, Whittaker et al. 2001). Of the different metrics to measure species *alpha* diversity, species richness, the number of different species found at a locality (McIntosh 1967), is the most commonly used. Species evenness (Simpson 1949), which takes into account the relative abundances of species in a sample, requires more effort to determine and is less commonly reported. As a measure of the relative rarity of evenness as compared to richness in scientific articles, a Web of Science search on 2 July 2014 for “species richness” returned 167,104 articles versus “evenness” with 27,824 or “species evenness” with 1,397. A recent paper, however, shows strong dissimilarities between the placement of biodiversity hotspots in the oceans as mapped by these two diversity indices (Stuart-Smith et al. 2013). Clearly, the measurement units we choose can have a significant impact on the maps we build of biodiversity.

Diversity within species

Genetic diversity, variation between individual members of a species, is not commonly used as a metric for mapping biodiversity. Diversity within species can be measured using neutral genes (not directly affected by selection, these can be mitochondrial, chloroplast or nuclear genes) or non-neutral genes (those directly influencing the phenotype of individuals, generally only measured using select nuclear genes or variance among phenotypes). Both types of genetic diversity are affected by the neutral processes of genetic drift and immigration, but non-neutral variation is additionally altered by selection on individual phenotypes. Some authors have found diversity of these two types to covary (Romiguier et al. 2014, Mittell et al. 2015), while others disagree (Reed and Frankham 2001, Whitlock 2014). Whether neutral genetic diversity is a good predictor of non-neutral diversity is not yet determined, but it seems rational to expect that there is some correlation since both are affected by neutral dynamics. The field of conservation genetics takes this assumption as the basis for making reasonable inferences about population persistence under low population sizes. Typically, neutral diversity has been the target of measurement to date, with most existing data being of this kind, however the relatively recent development of whole genome

and transcriptome sequencing eases investigation of non-neutral genetic diversity. These ‘next generation’ techniques allow efficient and high resolution assessment of both types of genetic diversity, improving the power of analyses by assessing large numbers of genes.

Despite being recognised in the Convention on Biological Diversity (1992) as an important facet of biodiversity, very few signatories of the Convention acted upon the requirement of monitoring and conserving genetic diversity set out in the agreement (Laikre et al. 2009). This inaction flies in the face of evidence of the importance of within-species diversity for maintaining healthy, productive and resilient ecosystems. For example, genetic diversity has been shown to enhance the ecosystem function of communities, particularly if genetic diversity is measured within an ecosystem engineer. These effects are evidenced by multiple agricultural examples (Allard et al. 1978, Di Falco and Perrings 2003, Crutsinger et al. 2006, Fridley et al. 2007, Cadotte et al. 2008, Hajjar et al. 2008) (for a review, see Hughes et al. 2008). Genetic diversity in plants has been shown to increase productivity (Smithson and Lenne 1996, Hughes et al. 2010, Reynolds et al. 2012)) and resilience to disturbance (Davy et al. 1990, Hughes and Stachowicz 2004, Reusch et al. 2005, Crutsinger et al. 2008). These positive effects of genetic diversity have also been shown in insect colonies (Oldroyd and Fewell 2007), specifically honey bee (Tarpy 2003, Mattila and Seeley 2007) and ant populations (Hughes and Boomsma 2004). A diverse range of genotypes in a population is likely to mean a diversity of phenotypes, which can boost productivity by reducing competition and maximising the efficiency of resource use. The expectation is that diverse assemblages will partition the available resources efficiently, promoting coexistence rather than competitive exclusion (Stachowicz et al. 2007). Arguably, the benefits of genetic diversity surpass the abovementioned effects it can have on ecosystem services to humans (Ehrlich 1988, Crozier 1997). Genetic diversity is the fundamental unit of diversity and forms the raw material upon which evolution acts to generate species and to allow species to adapt to changes in their environment. As such, mapping the distribution of genetic diversity and identifying the factors underlying its accumulation, should be a priority.

Potential drivers of genetic diversity

Similar to species diversity, *alpha* and *beta* components exist within genetic diversity. Genetic differentiation among populations is analogous to *beta* diversity among communities of species. Genetic variation within populations, or standing genetic variation, is analogous to *alpha* diversity in communities of species. Here I consider the factors known to influence levels of diversity among and within populations of species.

Factors affecting genetic diversity within populations

Spatial and environmental factors, such as the area and isolation of a habitat patch or the productivity and heterogeneity of an environment, are known to affect genetic diversity within populations, analogously to their effects on species richness (Antonovics 1976). These factors influence genetic diversity through their effect on effective population size (N_E). Large populations can contain a greater variety of alleles (versions of a gene) than small populations, and are less likely to lose alleles through the process of genetic drift (Wright 1931). The genetic diversity of

populations can be facilitated by immigration of new individuals if they bring new alleles to a population. The variety of niches created by heterogeneous habitats also promotes genetic diversity by allowing different genotypes to exploit resources in different ways, as shown experimentally in bacteria (Rainey and Travisano 1998). Ecological interactions such as competition (Schutz 1969) and predation (Allen 1976) will also influence within-species diversity in much the same way as they affect species diversity. Extinction affects both species and alleles in the same way, in that individuals of a certain type are lost from a population or community (Palumbi 1997).

Factors affecting genetic diversity among populations

Assuming equal population sizes and in the absence of selection, the similarity (in terms of shared alleles) among populations of a species is expected to scale proportionally to the level of migration among them, a pattern termed Isolation By Distance (Wright 1943). Genetic drift, migration and time-in-isolation will all affect the differentiation among populations (Slatkin 1993, Hey and Nielsen 2004). Populations that have been isolated from each other for long periods are likely to diverge due to genetic drift and/or the differential effects of selection, potentially forming new species (Templeton 1989, Coyne and Orr 1998). Even when discrete barriers do not exist, ease of passage through the environment is not always homogenous (McRae and Beier 2007) so simple geographic distance may not always clearly explain genetic differentiation. For this reason, heterogenous environments are likely to increase genetic diversity among populations by allowing micro-allopatry to occur (Smith 1965). Species are likely to differ in their abilities to disperse, or may have different populations sizes, so the factors of migration and genetic drift may differ also. The use of multiple, codistributed species with shared evolutionary history can help to isolate the mechanisms underlying population differentiation (Dawson 2012).

Concordance across hierarchical levels of diversity

Given that the processes affecting diversity are similar (regardless of the hierarchical level considered) one would expect spatial patterns of genetic and species diversity to be relatively similar. This expectation was formalised by Vellend (2003) as the Species-Genetic Diversity Correlation (SGDC) which takes its central tenets from the Island Theory of Biogeography (MacArthur and Wilson 1967) and the Island Model of Population Genetics (Wright 1940). Both are neutral theories, giving more weight to the processes of migration and drift than to speciation or selection. Vellend (2003) points out that the two theories are analogous to one another, one concerned with the diversity of species and the other of alleles within species. If neutral dynamics are not the driving forces in communities or populations, perhaps because ecological factors have a stronger effect, we might not expect a positive SGDC. A correlation between the two levels of diversity would be expected if both are driven by similar processes, but also if one level of diversity effects an increase in another; i.e., there is a causative relationship between one level of diversity and another. For example, genetic diversity of a dominant, habitat-forming organism may foster species diversity of the community it supports if different genotypes produce different morphologies which increase the heterogeneity of habitat and thus create more niches for species. The effect of species diversity on genetic diversity may work in a similar way. Typically, these effects are

considered in relation to habitat-forming organisms (Bangert et al. 2005, Whitham et al. 2006), but it should be possible that the genetic or species diversity of any dominant focal group could cause heterogeneity in interactions between this focal group and the rest of the assemblage. For example, dominant reef herbivores may graze in slightly different ways depending on their genetic or species identity; this, in turn, may drive algal dynamics on reefs and affect the entire reef assemblage. However, a causal link between species and genetic diversity is difficult to establish without manipulative experimental techniques. Alternatively, a habitat patch with a fixed carrying capacity for the number of individuals it can support could drive a negative correlation between species and genetic diversity. As species number increases the population size of each or some species will decline, resulting in loss of alleles through genetic drift (Wright 1940).

Evidence for the SGDC

Tests of the SGDC have been strongly terrestrially and freshwater focused with very few marine examples (two examples are discussed below, Robinson et al. 2010, Messmer et al. 2012). Modelling of terrestrial plant communities (Vellend 2005) conformed to the expectation that species and genetic diversity should positively correlate. Field-based studies have generally shown a positive correlation between species and genetic diversity, particularly where studies sampled within discrete habitat patches such as islands, forest patches or ponds rather than in arbitrary units (Vellend et al. 2014). Exceptions to a positive SGDC show marked effects of variability among species (Puscas et al. 2008, Derry et al. 2009, Struebig et al. 2011). Derry (2009) found that species diversity in freshwater crustacean zooplankton did not correlate with haplotype diversity of a common copepod within boreal lakes with various histories of human acid disturbance. The authors concluded that this discord was due to isolation being the main driver of genetic diversity in their focal species while environmental factors were the strongest drivers of species diversity. Struebig et al. (2011) report a concordant decline between species diversity and genetic diversity of forest bats only in the species with the lowest capacity for dispersal. These results argue for a important role of life history and for caution against extrapolating results to other systems or organisms.

In the marine environment, where propagules are often carried in a fluid medium and species range sizes can be substantially larger than their terrestrial counterparts (Rapoport 1994), one might expect strong ecological forcing, such as that assumed to underpin the correlation between species and genetic diversity, to be less common. It is surprising, however, that few explicit tests of the SGDC exist in the marine environment; these ideas have been discussed in the field previously (Palumbi 1997, Briggs 2004) and shallow water marine systems fit the structure of discrete habitat units suggested by Vellend et al. (2014) as being appropriate for study. Marine studies have been characterised by substantial variation in the slopes of the SGDC among species. Robinson et al. (2010) found that genetic diversity within invertebrate species inhabiting salt marshes significantly correlated with species diversity of marsh plants. The eight invertebrate species sampled were marine or euryhaline species with a planktonic dispersal phase, but the positive effect was only found when genetic diversity was combined across the species, individual species slopes were not significantly positive. Messmer et al. (2012) found similar mixed results in coral reef fishes;

genetic diversity within 11 species of coral reef fish did not consistently correlate with the species richness of fishes across three island sites in the Pacific ocean; some species showed a positive correlation while others did not. Clearly, support for a positive relationship between species and genetic diversity can depend heavily on the species chosen for genetic analyses. This strong signal of interspecific variation argues for maximising the number of species with a view to reducing the effect of variance among species on the results of analyses, if the goal is to determine the generality of the SGDC.

Multi-species approaches

Spatial studies of within-species genetic diversity tend to focus on just one or a few taxa, so their scope of inference is limited. That single species approaches are the norm in studies of genetic variation (Chapter 2, Keyse et al. 2014) is possibly due to the significant outlay of resources, both financial and of expertise, required to produce genetic data (Féral 2002). Additionally, to make a thorough and robust assessment of genetic patterns within a species, one must be familiar with its biology and ecology (Bowen et al. 2014). This level of detailed knowledge of a taxonomic group does not come easily, so we are likely to continue to see single species studies persisting into the future, despite the easy access of large volumes of genetic data. This is no bad thing: single species studies provide a level of detail often not achieved in studies of multiple species where detailed knowledge of the taxa is not prioritised (Hubert et al. 2012). Provided that these studies sample in such a way that their efforts can be usefully combined with those of others in the field, single species studies can greatly improve progress of mapping the distribution of genetic diversity.

By combining genetic data from multiple species we can begin to see general patterns that often remain unclear when considering one species at a time. Concordant patterns can emerge when one considers multiple species, provided among-species variation does not obscure these general patterns (Hickerson et al. 2003). The idea of searching across multiple species for concordant patterns of genetic structure is not new; it forms the basis of the field of comparative phylogeography (Avise 2000). This field makes use of biogeographic patterns within species (phylogeographic patterns or population structure) to infer processes generating and maintaining diversity. It does so by drawing upon genetic information from multiple, codistributed species. This idea of using species as natural experiments within which we can gather data about the historical and contemporary factors contributing to the spatial arrangement of genetic variation is analogous to treating species as replicates (Hickerson et al. 2010). Where we find similar patterns across multiple species we can infer similar processes with increasing confidence as we include more species.

Where the aim is to combine species in analyses of among-population variation, some studies take a qualitative approach by calculating the proportion of species sharing concordant genetic breaks (Carpenter et al. 2011, Toonen et al. 2011). A more statistically robust approach involves accounting for the separate demographic and evolutionary history of each species (Hickerson et al. 2006).

Where the aim is to determine the distribution of within-species standing variation, combining data is complicated by inter-specific differences in mutation rates, population sizes and other life history characteristics. Life history has emerged as an important determinant of within-species genetic diversity (Charlesworth and Wright 2001, Leffler et al. 2012, Romiguier et al. 2014). While the links between species-specific traits and genetic diversity is an interesting field of study, if one is interested in testing for the generality of such patterns as the SGDC, these factors are a source of noise in one's data.

Inter-specific variation causes genetic diversity within different species to be fundamentally non-comparable, unless one controls for these sources of variation. One way of controlling this among-species variance is to choose closely-related species on the assumption that their phylogenetic relatedness will correspond to similar life history and demography (Grafen 1989). This assumption does not always hold, however, as closely related species can show marked differences in their genetic structure (Palumbi 1996, Crandall et al. 2008a). An effective method, and the one employed in Chapter 3 of this thesis, of accounting for among-species variation is to include species grouping factors (random effects) into quantitative analyses of genetic diversity data. Mixed effects models were developed to account for non-independence of data in longitudinal studies (Laird and Ware 1982) and have since gained ground in the fields of ecology and evolution (Bolker et al. 2009). Essentially, mixed effects models allow examination of relationships between variables while allowing for non-independence among data points. For example, we can combine genetic diversity data for many species and regress them on some forcing factor (such as species richness) while accounting for the fact that data from Species A are more similar to each other than data from Species B. This allowance for the non-independence of data allows us to partition out the genetic variation associated with species identity from the genetic variation introduced by the forcing factor. Patterns within a few species are unlikely to represent entire communities (Toonen et al. 2011) so multi-species studies are an essential tool for tackling questions posed above: what processes are important for generating and maintaining genetic diversity? How can we predict genetic diversity distribution? Answering these questions requires an appropriate choice of study system that encompasses diversity gradients amongst discrete habitat patches and has been the target of molecular genetic diversity surveys of multiple, codistributed taxa.

The Tropical Indo-Pacific

Housing the greatest concentration of tropical marine life on the planet (Ekman 1935, Briggs 1974, Veron 1995), the Indo-Pacific spans two thirds of the globe and borders over 65 nations, providing sustenance and livelihoods for billions of people. The sheer diversity of life in this region sparked the minds of notable biogeographers in centuries past (Forbes 1856, Wallace 1876) and continues to do so today (Briggs 2005, Carpenter and Springer 2005, Bellwood et al. 2012). Species diversity for many tropical marine taxa peaks in the centre of this region, declining longitudinally outward into the Indian and Pacific Oceans (Roberts et al. 2002a, Hoeksema 2007, Tittensor et al. 2010). Efforts to understand and disentangle the processes underlying such complex systems are ongoing,

requiring input from such diverse fields as biogeography, molecular genetics, oceanography and ecology. Researchers working in the Indo-Pacific have made significant progress in terms of mapping regions of genetic disjunction within multiple species (Palumbi 1997, Carpenter et al. 2011, Toonen et al. 2011). Molecular genetics have been instrumental in moving away from the paradigm of marine systems as open networks of populations connected by high levels of homogenising dispersal (Scheltema 1988). Early evidence from allozyme studies supported this idea of high levels of gene flow among marine populations (Campbell et al. 1975, Winans 1980, Nishida and Lucas 1988, Williams and Benzie 1996), but, with newer genetic markers (mtDNA sequence data and genotyping) we increasingly see evidence of strong genetic structure (Springer and Williams 1990, Benzie 1998, Benzie 1999, Ovenden et al. 2004, DeBoer et al. 2014). This is not to say that modern studies always find evidence of restricted gene flow among marine populations, patterns consistent with high connectivity in marine species continue to be published (Lessios and Robertson 2006, Crandall et al. 2014). Regions of genetic discontinuity across multiple species in the Indo-Pacific have been identified by bringing together data from multiple single-species studies (Carpenter et al. 2011, Toonen et al. 2011) however these studies generally take a qualitative approach to summarising genetic patterns across species. Broadly, evidence has accumulated for the Coral Triangle, at the centre of the Indo-Pacific, being a region where multiple differentiated clades of marine species coexist (Barber et al. 2002, Crandall et al. 2008a, Drew and Barber 2009, Knittweis et al. 2009, Gaither et al. 2011, Ackiss et al. 2013).

The Coral Triangle – a hotspot of genetic diversity?

Within the Coral Triangle, evidence from phylogeographic studies has highlighted the action of restricted gene flow in determining patterns of within-species diversity. The evidence described here focuses on genetic diversity among, rather than within, populations: analogous to *beta* diversity in a community ecology context. The region, located at the confluence of the Indian and Pacific Oceans, is characterised by extensive shallow shelf habitat and swept by strong oceanic currents. During the Pleistocene Epoch, approximately 2.5 million to 12 thousand years ago, multiple glacial cycles caused sea levels to fluctuate, reducing sea levels in this region by up to 130 metres below present (Voris 2000). As a consequence, the Sunda and Sahul shelves were exposed and the area of shallow water habitat was significantly reduced (Fig. 1). Sea level change caused the closure of both the Sunda and Torres Straits and a significant shallowing of the region in the centre of the Coral Triangle, potentially significantly reducing the flow between the Indian and Pacific Oceans. Freshwater inputs into this region are thought to have increased during this period, along with cold water upwellings (Fleminger 1986) reducing habitat suitability for many shallow water fauna. Several marine species show population differentiation concordant with these sea level changes, in terms of population differentiation (Elliott 1996, Begg et al. 1998, Chenoweth et al. 1998, Dethmers et al. 2006, Imron et al. 2007, Lukoschek et al. 2007, Van Herwerden et al. 2009, Mirams 2011) and population expansion following the resubmersion of the continental shelves (Lukoschek et al. 2007, Crandall et al. 2012).

Other physical factors causing reductions to gene flow are more subtle. Rapidly flowing or bifurcating ocean currents can have significant impact on gene flow within species wherein dispersal is accomplished via a planktonic larval phase. The Coral Triangle is affected from the north and east by strong oceanic currents (Wyrtki 1961) and experiences seasonal fluctuations in the strength and direction of currents. The Indonesian Throughflow passes through the Coral Triangle, carrying water rapidly from the Pacific to the Indian Ocean through a deep channel (Gordon et al. 2003) and has been shown to correlate with population divergence of a shark species (Giles et al. 2014). The presence of the seasonally strong Mindanao and Halmahera eddies, to the southeast of the Philippines, have been associated with patterns of genetic discontinuity in multiple marine species inhabiting the northeast of the Coral Triangle (Barber et al. 2006, Crandall et al. 2008b, DeBoer et al. 2008, Liu et al. 2008, DeBoer et al. 2014, Jackson et al. 2014). Despite these many examples of positive correlations between putative barriers to gene flow and significant interpopulation genetic structure, there are many examples of species not showing these disjunctions (Williams and Benzie 1996, Horne et al. 2008, Reece et al. 2010, Mirams 2011).

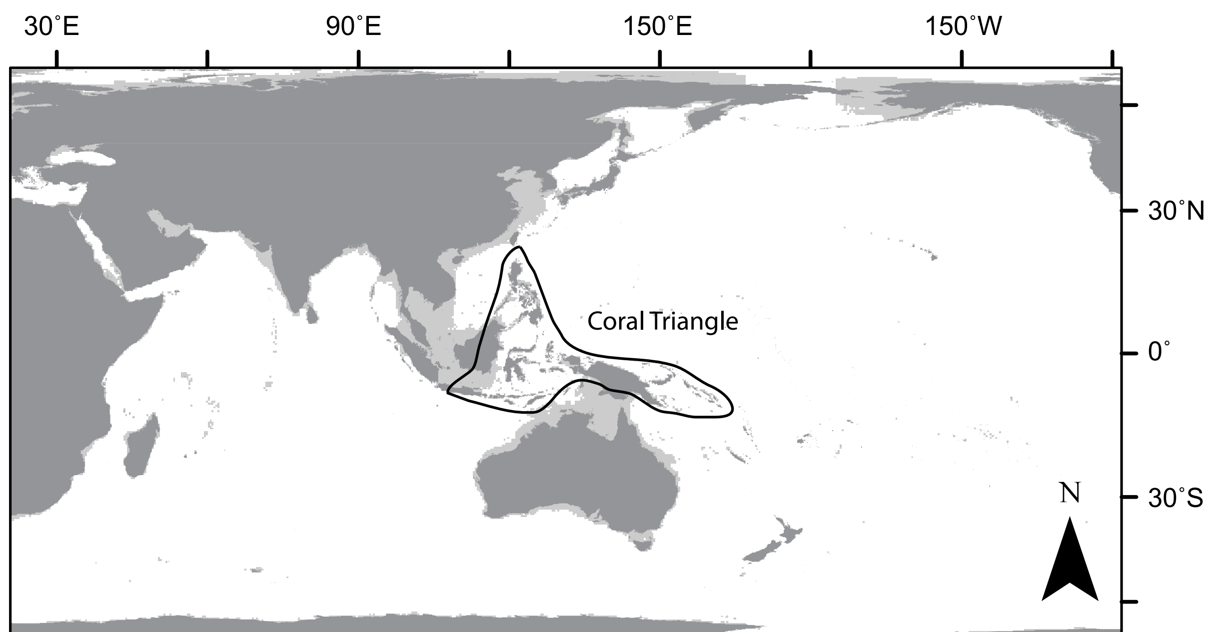


Figure 1. Map of the Indo-Pacific showing the Coral Triangle. Grey outline shows the approximate coastline during Pleistocene lower sea levels 250,000 - 18,000 years ago.

Aims of thesis

As described above, the Indo-Pacific region has been the focus of molecular genetic studies by many different research groups. This great repository of genetic data on individual species has great potential to allow us to answer questions across broad spatial and taxonomic extents. In the second chapter of this thesis I aim to assess the geographic and taxonomic scope of these data, focusing on marine fauna and data presented as summary statistics in the literature. My decision to collate published indices of genetic diversity, rather than raw genetic data, stemmed from a desire to capture data pre-dating the general cooperation of researchers with online databases such as NCBI Genbank. Secondly, I found that many database entries provided insufficient geographic information, rendering the data useless for my purposes. This lack formed the focus of a recent publication (Pope et al 2015) included as an Appendix to this thesis. Finally, I found that many

data sets consisted of only unique DNA sequences, meaning that each data point would need cross-referencing to the publication to derive the full data set. As such, the benefits of using raw genetic data (quality control, reassignment of populations, possibility of further analyses such as inferring the direction of gene flow) were outweighed by the sheer number of species for which I could derive geographic information. When these shortcomings are solved and large volumes of georeferenced raw data become available we will be able to make great progress towards synthesising the products of the extensive efforts made to date in the molecular ecology of the Indo-Pacific.

In the third chapter, I aim to make the first use of this multi-species data set to assess the relevance of the Species-Genetic Diversity Correlation (SGDC) in marine systems at this broad spatial extent. In this chapter I introduce mixed effect modelling into the SGDC to deal with non-independent data from multiple species. Furthermore, I demonstrate a robust approach to modelling published indices of genetic diversity that accounts for the proportional and skewed nature of these indices. By combining these data from 75 species in a mixed beta regression framework I am able to partition out the variation in genetic diversity that can be explained by species richness rather than by species life history.

In the fourth chapter I focus in on a group of sessile marine invertebrates, giant clams in the genus *Tridacna*, with the aim to assess the spatial drivers of genetic diversity in the region. Several species in this group have been studied in the region before, both intensively in the Coral Triangle and more diffusely in the Indian and Pacific Oceans. By linking these disparate sampling regimes together and filling a significant gap in the west Pacific I am able to maximise the value of previous data and test the relative strength of historical versus contemporary factors influencing gene flow.

Overall, this thesis aims to demonstrate the utility of combining published genetic data from multiple studies in order to answer questions at a broad spatial and taxonomic scale. I aim to show the volume of data available for such synthetic studies and highlight ways we can improve data collection, collation and ensure continued utility. Ideally, as genetic data become vastly more numerous and detailed in their scope, we will be able to leverage these data to make quantitative assessment of global patterns of diversity at the fundamental level of organisation, within species. I aim to demonstrate two different ways in which the existing data can help us understand the generation and maintenance of diversity. Firstly, I aim to develop a statistical framework for analysing summary statistic indices of genetic diversity from multiple species. Using these methods I will make the first assessment of the species-genetic correlation to account for variation among multiple species. Finally, I will demonstrate the continued utility of mtDNA sequence data when accompanied by detailed geographic information to improve the scope of inference for modern phylogeographic studies assessing the influence of historical and contemporary factors. In the current climate of rapid biodiversity loss, with limited funding for biodiversity research, the ability to maximise the value of genetic data will allow us to maintain forward motion in mapping diversity within species in the Indo-Pacific. The Indo-Pacific region contains the greatest accumulation of

tropical marine biodiversity on the planet but is also home to high densities of human populations, very few of whom have scientific expertise (Barber et al. 2014). With the intention of building biodiversity research capacity in the region I will present an argument for open access publication of georeferenced genetic data as standard, and for clear guidelines to foster effective collaboration among molecular ecologists across national boundaries.

Literature Cited

Ackiss, A. S., S. Pardede, E. D. Crandall, M. C. A. Ablan-Lagman, Ambariyanto, N. Romena, P. H. Barber, and K. E. Carpenter. 2013. Pronounced genetic structure in a highly mobile coral reef fish, *Caesio cuning*, in the Coral Triangle. Marine Ecology Progress Series 480:185-197.

Allard, R., R. Miller, and A. Kahler. 1978. The relationship between degree of environmental heterogeneity and genetic polymorphism. Pages 49–73 in A.H.J. Freysen and J. W. Woldendorp, editors. Structure and Functioning of Plant Populations. North-Holland Publishing Co, Amsterdam.

Allen, J. A. 1976. Further evidence for apostatic selection by wild passerine birds experiments. Heredity 36:173-180.

Antonovics, J. 1976. The input from population genetics: “the new ecological genetics”. Systematic Botany:233-245.

Avise, J. C. 2000. Phylogeography: the history and formation of species. Harvard Univ Press.

Bangert, R. K., R. J. Turek, G. D. Martinsen, G. M. Wimp, J. K. Bailey, and T. G. Whitham. 2005. Benefits of conservation of plant genetic diversity to arthropod diversity. Conservation Biology 19:379-390.

Barber, P. H., M. C. A. Ablan-Lagman, R. G. Berlinck, D. Cahyani, E. D. Crandall, R. Ravago-Gotanco, M. A. Juinio-Meñez, I. Mahardika, K. Shanker, and C. J. Starger. 2014. Advancing biodiversity research in developing countries: the need for changing paradigms. Bulletin of Marine Science 90:187-210.

Barber, P. H., M. V. Erdmann, and S. R. Palumbi. 2006. Comparative phylogeography of three codistributed stomatopods: Origins and timing of regional lineage diversification in the Coral Triangle. Evolution 60:1825-1839.

Barber, P. H., S. R. Palumbi, M. V. Erdmann, and M. K. Moosa. 2002. Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns,

causes, and consequences. Molecular Ecology 11:659-674.

Barnosky, A. D., N. Matzke, S. Tomiya, G. O. U. Wogan, B. Swartz, T. B. Quental, C. Marshall, J. L. McGuire, E. L. Lindsey, K. C. Maguire, B. Mersey, and E. A. Ferrer. 2011. Has the Earth's sixth mass extinction already arrived? Nature 471:51-57.

Begg, G., C. Keenan, and M. Sellin. 1998. Genetic variation and stock structure of school mackerel and spotted mackerel in northern Australian waters. Journal of Fish Biology 53:543-559.

Bellwood, D. R., W. Renema, and B. R. Rosen. 2012. Biodiversity hotspots, evolution and coral reef biogeography: a review.in D. Gower, Johnson, K., Richardson, R., Rosen, B., Rüber, L., Williams, S., editor. Biotic Evolution and Environmental Change in Southeast Asia Cambridge University Press., Cambridge, UK.

Benzie, J. 1998. Genetic structure of marine organisms and SE Asian biogeography. Biogeography and geological evolution of SE Asia 30:197-209.

Benzie, J. A. H. 1999. Genetic structure of coral reef organisms: Ghosts of dispersal past. American Zoologist 39:131-145.

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution 24:127-135.

Bowen, B. W. 1999. Preserving genes, species, or ecosystems? Healing the fractured foundations of conservation policy. Molecular Ecology 8:S5–S10.

Bowen, B. W., K. Shanker, N. Yasuda, M. Celia, M. C. M. D. Malay, S. von der Heyden, G. Paulay, L. A. Rocha, K. A. Selkoe, and P. H. Barber. 2014. Phylogeography unplugged: comparative surveys in the genomic era. Bulletin of Marine Science 90:13-46.

Briggs, J. C. 1974. Marine zoogeography. McGraw-Hill, New York.

Briggs, J. C. 2004. Older species: a rejuvenation on coral reefs? Journal of Biogeography 31:525-530.

Briggs, J. C. 2005. The marine East Indies: Diversity and speciation. Journal of Biogeography 32:1517-1522.

Brown, J. H., Maurer, B.A. 1989. Macroecology: the division of food and space among species on continents. Science 243:1145-1150.

Cadotte, M. W., B. J. Cardinale, and T. H. Oakley. 2008. Evolutionary history and the effect of biodiversity on plant productivity. Proceedings of the National Academy of Sciences of the United States of America 105:17012-17017.

Campbell, C. A., J. W. Valentine, and F. J. Ayala. 1975. High genetic variability in a population of *Tridacna maxima* from the Great Barrier Reef. Marine Biology 33:341-345.

Carpenter, K. E., P. H. Barber, E. D. Crandall, B. M. Manjaji-Matsumoto, M. A. Juinio-Menez, M. D. Santos, C. J. Starger, and A. H. A. Toha10. 2011. Comparative phylogeography of the Coral Triangle and implications for marine management. Journal of Marine Biology.

Carpenter, K. E. and V. G. Springer. 2005. The center of the center of marine shore fish biodiversity: the Philippine Islands. Environmental Biology of Fishes 72:467-480.

Charlesworth, D. and S. I. Wright. 2001. Breeding systems and genome evolution. Current Opinion in Genetics & Development 11:685-690.

Chenoweth, S. F., J. M. Hughes, C. P. Keenan, and S. Lavery. 1998. When oceans meet: a teleost shows secondary intergradation at an Indian-Pacific interface. Proceedings of the Royal Society of London Series B-Biological Sciences 265:415-420.

Coyne, J. A. and H. A. Orr. 1998. The evolutionary genetics of speciation. Philosophical Transactions of the Royal Society B: Biological Sciences 353:287-305.

Crandall, E. D., M. A. Frey, R. K. Grosberg, and P. H. Barber. 2008a. Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. Molecular Ecology 17:611-626.

Crandall, E. D., M. E. Jones, M. M. Muñoz, B. Akinronbi, M. V. Erdmann, and P. H. Barber. 2008b. Comparative phylogeography of two seastars and their ectosymbionts within the Coral Triangle. Molecular Ecology 17:5276-5290.

Crandall, E. D., E. J. Sbrocco, T. S. Deboer, P. H. Barber, and K. E. Carpenter. 2012. Expansion dating: calibrating molecular clocks in marine species from expansions onto the Sunda Shelf following the Last Glacial Maximum. Molecular Biology and Evolution 29:707-719.

Crandall, E. D., E. A. Treml, L. Liggins, L. Gleeson, N. Yasuda, P. H. Barber, G. Wörheide, and C. Riginos. 2014. Return of the ghosts of dispersal past: historical spread and contemporary gene flow in the blue sea star *Linckia laevigata*. Bulletin of Marine Science 90:399-425.

Crozier, R. 1997. Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. Annual Review of Ecology and Systematics 28:243-268.

Crutsinger, G. M., M. D. Collins, J. A. Fordyce, Z. Gompert, C. C. Nice, and N. J. Sanders. 2006. Plant genotypic diversity predicts community structure and governs an ecosystem process. Science 313:966-968.

Crutsinger, G. M., L. Souza, and N. J. Sanders. 2008. Intraspecific diversity and dominant genotypes resist plant invasions. Ecology Letters 11:16-23.

Darwin, C. 1859. The origin of species. Hayes Barton Press.

Davy, A., S. Noble, and R. Oliver. 1990. Genetic variation and adaptation to flooding in plants. Aquatic Botany 38:91-108.

Dawson, M. N. 2012. Parallel phylogeographic structure in ecologically similar sympatric sister taxa. Molecular Ecology 21:987-1004.

De Queiroz, K. 2007. Species concepts and species delimitation. Systematic Biology 56:879-886.

DeBoer, T. S., M. R. Naguit, M. V. Erdmann, M. C. A. Ablan-Lagman, Ambariyanto, K. E. Carpenter, A. H. A. Toha, and P. H. Barber. 2014. Concordant phylogenetic patterns inferred from mitochondrial and microsatellite DNA in the giant clam *Tridacna crocea*. Bulletin of Marine Science 90:301 - 331.

DeBoer, T. S., M. D. Subia, Ambariyanto, M. V. Erdmann, K. Kovitvongsa, and P. H. Barber. 2008. Phylogeography and limited genetic connectivity in the endangered boring giant clam across the Coral Triangle. Conservation Biology 22:1255-1266.

Derry, A. M., S. E. Arnott, J. A. Shead, P. D. N. Hebert, and P. T. Boag. 2009. Ecological linkages between community and genetic diversity in zooplankton among boreal shield lakes. Ecology 90:2275-2286.

Dethmers, K. E. M., D. Broderick, C. Moritz, N. N. Fitzsimmons, C. J. Limpus, S. Lavery, S. Whiting, M. Guinea, R. I. T. Prince, and R. Kennett. 2006. The genetic structure of Australasian green turtles (*Chelonia mydas*): exploring the geographical scale of genetic exchange. Molecular Ecology 15:3931-3946.

Di Falco, S. and C. Perrings. 2003. Crop genetic diversity, productivity and stability of agroecosystems. A theoretical and empirical investigation. Scottish Journal of Political Economy 50:207-216.

Donoghue, M. J. 1985. A critique of the biological species concept and recommendations for a phylogenetic alternative. Bryologist 88:172–181.

Drew, J. and P. H. Barber. 2009. Sequential cladogenesis of the reef fish *Pomacentrus moluccensis* (Pomacentridae) supports the peripheral origin of marine biodiversity in the Indo-Australian archipelago. Molecular Phylogenetics and Evolution 53:335-339.

Ehrlich, P. R. 1988. The loss of diversity. Biodiversity. National Academy Press, Washington DC:21-27.

Ekman, S. 1935. Tiergeographie des meeres. Akademische Verlagsgesellschaft MBH, Leipzig.

Elliott, N. G. 1996. Allozyme and mitochondrial DNA analysis of the tropical saddle-tail sea perch, *Lutjanus malabaricus* (Schneider), from Australian waters. Marine and Freshwater Research 47.

Féral, J.-P. 2002. How useful are the genetic markers in attempts to understand and manage marine biodiversity? Journal of Experimental Marine Biology and Ecology 268:121-145.

Fleminger, A. 1986. The Pleistocene equatorial barrier between the Indian and Pacific Oceans and a likely cause for Wallace's line.

Forbes, E. 1856. Map of the distribution of marine life. Page 31 in A. Johnston, editor. The physical atlas of natural phenomena W Johnston and AK Johnston, London.

Fridley, J. D., J. P. Grime, and M. Bilton. 2007. Genetic identity of interspecific neighbours mediates plant responses to competition and environmental variation in a species-rich grassland. Journal of Ecology 95:908-915.

Gaither, M. R., B. W. Bowen, T. R. Bordenave, L. A. Rocha, S. J. Newman, J. A. Gomez, L. van Herwerden, and M. T. Craig. 2011. Phylogeography of the reef fish *Cephalopholis argus* (Epinephelidae) indicates Pleistocene isolation across the Indo-Pacific barrier with contemporary overlap in the Coral Triangle. Bmc Evolutionary Biology 11.

Giles, J. L., J. R. Ovenden, D. AlMojil, E. Garvilles, K.-O. Khampetch, H. Manjebrayakath, and C. Riginos. 2014. Extensive genetic population structure in the Indo–West Pacific spot-tail shark, *Carcharhinus sorrah*. Bulletin of Marine Science 90:427-454.

Gordon, A. L., C. F. Giulivi, and A. G. Ilahude. 2003. Deep topographic barriers within the Indonesian seas. Deep Sea Research Part II: Topical Studies in Oceanography 50:2205-2228.

Grafen, A. 1989. The phylogenetic regression. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences:119-157.

Hajjar, R., D. I. Jarvis, and B. Gemmill-Herren. 2008. The utility of crop genetic diversity in maintaining ecosystem services. Agriculture, ecosystems & environment 123:261-270.

Hey, J. and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. Genetics 167:747-760.

Hickerson, M., B. Carstens, J. Cavender-Bares, K. Crandall, C. Graham, J. Johnson, L. Rissler, P. Victoriano, and A. Yoder. 2010. Phylogeography's past, present, and future: 10 years after Avise, 2000. Molecular Phylogenetics and Evolution 54:291-301.

Hickerson, M., M. Gilchrist, and N. Takebayashi. 2003. Calibrating a molecular clock from phylogeographic data: moments and likelihood estimators. Evolution 57:2216–2225.

Hickerson, M. J., G. Dolman, and C. Moritz. 2006. Comparative phylogeographic summary statistics for testing simultaneous vicariance. Molecular Ecology 15:209-223.

Hoeksema, B. 2007. Delineation of the Indo-Malayan centre of maximum marine biodiversity: the Coral Triangle. Biogeography, time, and place: distributions, barriers, and islands 29:117-178.

Horne, J. B., L. van Herwerden, J. H. Choat, and D. Robertson. 2008. High population connectivity across the Indo-Pacific: congruent lack of phylogeographic structure in three reef fish congeners. Molecular Phylogenetics and Evolution 49:629-638.

Hubert, N., C. P. Meyer, H. J. Bruggemann, F. Guerin, R. J. Komeno, B. Espiau, R. Causse, J. T. Williams, and S. Planes. 2012. Cryptic diversity in Indo-Pacific coral-reef fishes revealed by DNA-barcoding provides new support to the centre-of-overlap hypothesis. PLOS one 7:e28987.

Hughes, A. R., R. J. Best, and J. J. Stachowicz. 2010. Genotypic diversity and grazer identity interactively influence seagrass and grazer biomass. Marine Ecology-Progress Series 403:43-51.

Hughes, A. R., B. D. Inouye, M. T. Johnson, N. Underwood, and M. Vellend. 2008. Ecological consequences of genetic diversity. Ecology letters 11:609-623.

Hughes, A. R. and J. J. Stachowicz. 2004. Genetic diversity enhances the resistance of a seagrass ecosystem to disturbance. Proceedings of the National Academy of Sciences of the United States of America 101:8998.

Hughes, W. O. and J. J. Boomsma. 2004. Genetic diversity and disease resistance in leaf-cutting ant societies. Evolution 58:1251-1260.

Imron, B. Jeffrey, P. Hale, B. M. Degnan, and S. M. Degnan. 2007. Pleistocene isolation and recent gene flow in *Haliotis asinina*, an Indo-Pacific vetigastropod with limited dispersal capacity. Molecular Ecology 16:289-304.

Jackson, A. M., M. V. Erdmann, A. H. A. Toha, L. A. Stevens, and P. H. Barber. 2014. Phylogeography of commercial tuna and mackerel in the Indonesian Archipelago. Bulletin of Marine Science 90:471-492.

Keyse, J., E. D. Crandall, R. J. Toonen, C. P. Meyer, E. A. Treml, and C. Riginos. 2014. The scope of published population genetic data for Indo-Pacific marine fauna and future research opportunities in the region. Bulletin of Marine Science 90:47-78.

Knittweis, L., W. E. Kraemer, J. Timm, and M. Kochzius. 2009. Genetic structure of *Heliofungia actiniformis* (Scleractinia: Fungiidae) populations in the Indo-Malay Archipelago: implications for live coral trade management efforts. Conservation Genetics 10:241-249.

Laikre, L., F. W. Allendorf, L. C. Aroner, C. S. Baker, D. P. Gregovich, M. M. Hansen, J. A. Jackson, K. C. Kendall, K. McKelvey, and M. C. Neel. 2009. Neglect of genetic diversity in implementation of the Convention on Biological Diversity. Conservation Biology 24:86-88.

Laird, N. M. and J. H. Ware. 1982. Random-effects models for longitudinal data. Biometrics:963-974.

Laurance, W. F. 2010. Habitat destruction: death by a thousand cuts. Conservation biology for all 1:73-88.

Leffler, E. M., K. Bullaughey, D. R. Matute, W. K. Meyer, L. Segurel, A. Venkat, P. Andolfatto, and M. Przeworski. 2012. Revisiting an old riddle: what determines genetic diversity levels within species. Plos Biol 10:e1001388.

Lessios, H. A. and D. R. Robertson. 2006. Crossing the impassable: genetic connections in 20 reef fishes across the eastern Pacific barrier. Proceedings of the Royal Society B: Biological Sciences 273:2201.

Liu, S. Y. V., T. Kokita, and C. F. Dai. 2008. Population genetic structure of the neon damselfish (*Pomacentrus coelestis*) in the northwestern Pacific Ocean. Marine Biology 154:745-753.

Lukoschek, V., M. Waycott, and H. Marsh. 2007. Phylogeography of the olive sea snake, *Aipysurus*

laevis (Hydrophiinae) indicates Pleistocene range expansion around northern Australia but low contemporary gene flow. Molecular Ecology 16:3406-3422.

MacArthur, R. H. and E. O. Wilson. 1967. The theory of island biogeography. Princeton Univ Pr.

Magurran, A. E. 2004. Measuring biological diversity. Blackwell Science Ltd, London.

Mattila, H. R. and T. D. Seeley. 2007. Genetic diversity in honey bee colonies enhances productivity and fitness. Science 317:362-364.

Mayr, E. 1963. Animal species and evolution. Animal species and their evolution. 76:193-203.

McIntosh, R. P. 1967. An index of diversity and the relation of certain concepts to diversity. Ecology 48:392-404.

McRae, B. H. and P. Beier. 2007. Circuit theory predicts gene flow in plant and animal populations. Proceedings of the National Academy of Sciences 104:19885-19890.

Messmer, V., G. P. Jones, P. L. Munday, and S. Planes. 2012. Concordance between genetic and species diversity in coral reef fishes across the Pacific Ocean biodiversity gradient. Evolution 66:3902-3917.

Michener, C. D. 1970. Diverse approaches to systematics. Evolutionary Biology 4:1-38.

Mirams, A. G. K., Treml, E.A., Shields, J.L., Liggins, L. & Riginos, C. 2011. Vicariance and dispersal across an intermittent barrier: population genetic structure of marine animals across the Torres Strait land bridge. Coral Reefs.

Mittell, E. A., S. Nakagawa, and J. D. Hadfield. 2015. Are molecular markers useful predictors of adaptive potential? Ecology Letters.

Nishida, M. and J. S. Lucas. 1988. Genetic differences between geographic populations of the Crown-of-thorns starfish throughout the Pacific region. Marine Biology 98:359-368.

Norse, E. A., K. L. Rosenbaum, D. S. Wilcove, B. A. Wilcox, W. H. Romme, D. W. Johnston, and M. L. Stout. 1986. Conserving biological diversity in our national forests. Wilderness Society.

Oldroyd, B. P. and J. H. Fewell. 2007. Genetic diversity promotes homeostasis in insect colonies. Trends in Ecology & Evolution 22:408-413.

Ovenden, J. R., J. Salini, S. O'Connor, and R. Street. 2004. Pronounced genetic population structure

in a potentially vagile fish species (*Pristipomoides multidens*, Teleostei : Perciformes : Lutjanidae) from the East Indies triangle. Molecular Ecology 13:1991-1999.

Palumbi, S. 1996. Macrospatial genetic structure and speciation in marine taxa with high dispersal abilities. Pages 101-117 in J. Ferraris and S. Palumbi, editors. Molecular zoology: advances, strategies and protocols. Wiley, New York.

Palumbi, S. R. 1997. Molecular biogeography of the Pacific. Coral Reefs 16:S47-S52.

Pianka, E. R. 1966. Latitudinal gradients in species diversity: a review of concepts. The American Naturalist 100:33-46.

Pope L. C., L. Liggins, J. Keyse, S.B. Carvalho, C. Riginos (2015) Not the time or the place: the missing spatio-temporal link in publicly available genetic data. Molecular Ecology. DOI 10.1111/mec.13254

Puscas, M., P. Taberlet, and P. Choler. 2008. No positive correlation between species and genetic diversity in European alpine grasslands dominated by *Carex curvula*. Diversity and Distributions 14:852-861.

Rainey, P. B. and M. Travisano. 1998. Adaptive radiation in a heterogeneous environment. Nature 394:69-72.

Rapoport, E. 1994. Remarks on marine and continental biogeography: an areographical viewpoint. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 343:71-78.

Reece, J. S., B. W. Bowen, K. Joshi, V. Goz, and A. Larson. 2010. Phylogeography of two moray eels indicates high dispersal throughout the Indo-Pacific. Journal of Heredity 101:391-402.

Reed, D. H. and R. Frankham. 2001. How closely correlated are molecular and quantitative measures of genetic variation? a meta-analysis. Evolution 55:1095-1103.

Reusch, T. B. H., A. Ehlers, A. Hämmerli, and B. Worm. 2005. Ecosystem recovery after climatic extremes enhanced by genotypic diversity. Proceedings of the National Academy of Sciences of the United States of America 102:2826.

Reynolds, L. K., M. Waycott, K. J. McGlathery, R. J. Orth, and J. C. Zieman. 2012. Eelgrass restoration by seed maintains genetic diversity: case study from a coastal bay system. Marine Ecology Progress Series 448:223-233.

Roberts, C. M., C. J. McClean, J. E. Veron, J. P. Hawkins, G. R. Allen, D. E. McAllister, C. G. Mittermeier, F. W. Schueler, M. Spalding, and F. Wells. 2002a. Marine biodiversity hotspots and conservation priorities for tropical reefs. Science 295:1280-1284.

Robinson, J. D., E. Diaz-Ferguson, M. F. Poelchau, S. Pennings, T. D. Bishop, and J. Wares. 2010. Multiscale diversity in the marshes of the Georgia coastal ecosystems LTER. Estuaries and Coasts 33:865-877.

Rohde, K. 1992. Latitudinal gradients in species diversity: the search for the primary cause. Oikos 65:514-527.

Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, R. Dernat, L. Duret, N. Faivre, E. Loire, J. M. Lourenco, B. Nabholz, C. Roux, G. Tsagkogeorga, A. A. Weber, L. A. Weinert, K. Belkhir, N. Bierne, S. Glemin, and N. Galtier. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515:261-263.

Scheltema, R. S. 1988. Initial evidence for the transport of teleplanic larvae of benthic invertebrates across the East Pacific Barrier. The Biological Bulletin 174:145-152.

Schutz, W. M., Usanis, S.A. 1969. Inter-genotypic competition in plant populations II. Maintenance of allelic polymorphisms with frequency-dependent selection and mixed selfing and random mating. Genetics 61:875-891.

Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution:264-279.

Smith, H. M. 1965. More evolutionary terms. Systematic Biology 14:57-58.

Smithson, J. B. and J. M. Lenne. 1996. Varietal mixtures: a viable strategy for sustainable productivity in subsistence agriculture. Annals of Applied Biology 128:127-158.
Springer, V. G. and J. T. Williams. 1990. Widely distributed Pacific plate endemics and lowered sea level. Bulletin of Marine Science 47:631-640.

Stachowicz, J. J., J. F. Bruno, and J. E. Duffy. 2007. Understanding the effects of marine biodiversity on communities and ecosystems. Annual Review of Ecology Evolution and Systematics 38:739-766.

Struebig, M. J., T. Kingston, E. J. Petit, S. C. Le Comber, A. Zubaid, A. Mohd-Adnan, and S. J. Rossiter. 2011. Parallel declines in species and genetic diversity in tropical forest fragments. Ecology Letters 14:582-590.

Stuart-Smith, R. D., A. E. Bates, J. S. Lefcheck, J. E. Duffy, S. C. Baker, R. J. Thomson, J. F. Stuart-Smith, N. A. Hill, S. J. Kininmonth, L. Airoldi, M. A. Becerro, S. J. Campbell, T. P. Dawson, S. A. Navarrete, G. A. Soler, E. M. Strain, T. J. Willis, and G. J. Edgar. 2013. Integrating abundance and functional traits reveals new global hotspots of fish diversity. Nature 501:539-542.

Tarpy, D. R. 2003. Genetic diversity within honeybee colonies prevents severe infections and promotes colony growth. Proceedings of the Royal Society of London. Series B: Biological Sciences 270:99-103.

Templeton, A. R. 1989. The meaning of species and speciation: a genetic perspective. The units of evolution: Essays on the nature of species:159-183.

Tittensor, D. P., C. Mora, W. Jetz, H. K. Lotze, D. Ricard, E. V. Berghe, and B. Worm. 2010. Global patterns and predictors of marine biodiversity across taxa. Nature 466:1098-1101.

Toonen, R. J., K. R. Andrews, I. B. Baums, C. E. Bird, G. T. Concepcion, T. S. Daly-Engel, J. A. Eble, A. Faucci, M. R. Gaither, M. Iacchei, J. B. Puritz, J. K. Schultz, D. J. Skillings, M. A. Timmers, and B. W. Bowen. 2011. Defining boundaries for ecosystem-based management: A multispecies case study of marine connectivity across the Hawaiian archipelago. Journal of Marine Biology:1-13.

Van Herwerden, L., J. H. Choat, S. J. Newman, M. Leray, and G. Hillersoy. 2009. Complex patterns of population structure and recruitment of *Plectropomus leopardus* (Pisces: Epinephelidae) in the Indo-West Pacific: implications for fisheries management. Marine Biology 156:1595-1607.

Van Valen, L. 1976. Ecological species, multispecies, and oaks. Taxon 25:233-239.

Vellend, M. 2005. Species diversity and genetic diversity: parallel processes and correlated patterns. The American Naturalist 166:199-215.

Vellend, M., G. Lajoie, A. Bourret, C. Murria, S. W. Kembel, and D. Garant. 2014. Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. Molecular Ecology 23:2890-2901.

Veron, J. E. N. 1995. Corals in space and time. the biogeography and evolution of the scleractinia. Cornell University Press (Comstock), Ithaca, London.

Von Humboldt, A. 1849. Cosmos: a sketch of a physical description of the universe. HG Bohn.

Voris, H. K. 2000. Special paper 2: Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. Journal of Biogeography 27:1153-1167.

Wallace, A. R. 1876. The geographical distribution of animals; with a study of the relations of living and extinct faunas as elucidating the past changes of the Earth's surface. Macmillan & Co, London.

Whitham, T. G., J. K. Bailey, J. A. Schweitzer, S. M. Shuster, R. K. Bangert, C. J. LeRoy, E. V. Lonsdorf, G. J. Allan, S. P. DiFazio, B. M. Potts, D. G. Fischer, C. A. Gehring, R. L. Lindroth, J. C. Marks, S. C. Hart, G. M. Wimp, and S. C. Wooley. 2006. A framework for community and ecosystem genetics: from genes to ecosystems. Nature reviews. Genetics 7:510-523.

Whitlock, R. 2014. Relationships between adaptive and neutral genetic diversity and ecological structure and functioning: a meta-analysis. The Journal of Ecology 102:857-872.

Whittaker, R. H. 1972. Evolution and measurement of species diversity. Taxon 21:213-251.

Whittaker, R. J., K. J. Willis, and R. Field. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. Journal of Biogeography 28:453-470.

Williams, S. T. and J. A. H. Benzie. 1996. Genetic uniformity of widely separated populations of the coral reef starfish Linckia laevigata from the East Indian and West Pacific Oceans, revealed by allozyme electrophoresis. Marine Biology 126:99-107.

Willig, M. R., D. M. Kaufman, and R. D. Stevens. 2003. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. Annual Review of Ecology & Systematics 34:273-309.

Winans, G. A. 1980. Geographic variation in the milkfish *Chanos chanos*. I. Biochemical evidence. Evolution:558-574.

Wright, S. 1943. Isolation by distance. Genetics 28:114-138.

Wright, S. G. 1931. Evolution in Mendelian Populations. Genetics 16:0097-0159.

Wright, S. G. 1940. Breeding Structure of Populations in Relation to Speciation. The American Naturalist 74:232-248.

Wyrtki, K. 1961. Scientific results of marine investigations of the South China Sea and the Gulf of Thailand 1959-1961. NAGA report 2.

CHAPTER 2

Bull Mar Sci. 90(1):000–000. 2014
http://dx.doi.org/10.5343/bms.2012.1107

The scope of published population genetic data for Indo-Pacific marine fauna and future research opportunities in the region

Jude Keyse [1]*
Eric D Crandall [2]
Robert J Toonen [3]
Christopher P Meyer [4]
Eric A Treml [5]
Cynthia Riginos [1]

[1] School of Biological Sciences, University of Queensland, St Lucia, QLD 4072, Australia.

[2] NOAA Southwest Fisheries Science Center, Fisheries Ecology Division & Institute of Marine Sciences, UC Santa Cruz, 110 Shaffer Rd. Santa Cruz, California 95062.

[3] Hawai‘i Institute of Marine Biology, School of Ocean and Earth Science and Technology, University of Hawai‘i at Mānoa, P.O. Box 1346 Kāne‘ohe, Hawaii 96744.

[4] Smithsonian Institution, National Museum of Natural History, 10th & Constitution Ave., NW, Washington, DC 20560-0163.

[5] Department of Zoology, University of Melbourne, Parkville, VIC 3010, Australia.

* Corresponding author email: judith.keyse@uqconnect.edu.au

Date Submitted: 2 January, 2013.
Date Accepted: 1 July, 2013.
Available Online: 18 November, 2013.

Abstract — Marine biodiversity reaches its pinnacle in the tropical Indo-Pacific region, with high levels of both species richness and endemism, especially in coral reef habitats. While this pattern of biodiversity has been known to biogeographers for centuries, causal mechanisms remain enigmatic. Over the past 20 yrs, genetic markers have been employed by many researchers as a tool to elucidate patterns of biodiversity above and below the species level, as well as to make inferences about the underlying processes of diversification, demographic history, and dispersal. In a quantitative, comparative framework, these data can be synthesized to address questions about this bewildering diversity by treating species as “replicates.” However, the sheer size of the Indo-Pacific region means that the geographic and genetic scope of many species’ data sets are not complementary. Here, we describe data sets from 116 Indo-Pacific species (108 studies). With a mind to future synthetic investigations, we consider the strengths and omissions of currently published population genetic data for marine fauna of the Indo-Pacific region, as well as the geographic and taxonomic scope of the data, and suggest some ways forward for data collection and collation.

The waters of the Indian and Pacific oceans contain the greatest concentration of tropical marine biodiversity on Earth (Ekman 1953, Briggs 1974, Veron 1995). The question of why marine biodiversity is concentrated in this region, particularly at the juncture of the Indian and Pacific oceans, has been the topic of much study (Forbes 1856, Ekman 1935, Ladd 1960, Briggs 1974, 1999, Bellwood and Hughes 2001, Connolly et al. 2003, Carpenter and Springer 2005, Hoeksema 2007, Reaka et al. 2008, Renema et al. 2008, Bellwood et al. 2012). Tools from population genetics and phylogeography can enhance our understanding of how biodiversity is created and maintained in this region (Avise et al. 1987, Palumbi 1997, Barber and Bellwood 2005). Moreover, genetic approaches are essential for initial detection of the many cryptic species that apparently exist in this region (Knowlton 2000, Meyer et al. 2005, Barber and Boyce 2006, Vogler et al. 2008, Bowen et al. 2013) and can also be used to guide conservation (Moritz 1994, Moritz and Faith 2002, Rocha et al. 2007, Beger et al. 2014, von der Heyden et al. 2014).

Numerous studies have investigated population genetic and phylogeographic patterns in the Indo-Pacific region (for examples, see recent reviews by Crandall et al. 2008a, Carpenter et al. 2011, Toonen et al. 2011); however, the high levels of biodiversity, combined with the vast area of the Indian and Pacific oceans, poses substantial challenges for documenting spatial genetic patterns, much less inferring underlying processes. For instance, the coral reefs of eastern Indonesia, Malaysia, the Philippines, Papua New Guinea, East Timor, and the Solomon Islands (collectively referred to as the Coral Triangle), contain the world's greatest concentration of marine species, which is consistently estimated in the upper decile for most coastal marine taxa (Roberts et al. 2002, Carpenter and Springer 2005, Tittensor et al. 2010). Together, the Indian and Pacific oceans span two thirds of the globe, with most individual species ranges encompassing much of one or both ocean basins (Connolly et al. 2003). This area includes more than 65 nations of which 18 are classified by the UN as Least Developed Countries, and only four are classified as High Income Organisation for Economic Co-operation and Development (OECD) countries (http://www.data.worldbank.org, accessed December 2012).

Thus, population genetic surveys in the Indo-Pacific region are likely to involve fieldwork in locations that are distant from each other, potentially difficult to access, may be in developing countries, and will fall under diverse regulations and jurisdictions. These are significant logistical impediments for biological research. Moreover, with such high biodiversity, the degree to which one or a few species can represent entire communities is unknown, and recent studies argue against exemplar species representing patterns for the broader community (Bird et al. 2007, Toonen et al. 2011). In the face of such challenges, progress can be fostered if data are shared and properly catalogued in the interests of capturing emergent patterns in this complex system. Our purpose here is not to provide a review of previous work on genetic patterns in the region (see Palumbi 1994, Benzie 1998, Carpenter et al. 2011, Toonen et al. 2011 for examples focusing on particular regions within the Indo-Pacific). Rather, our goal is to provide a detailed overview of published data from population genetic studies of Indo-Pacific marine fauna, which could be used for synthetic studies. In addition, we aim to inform future empirical studies by determining the

scope, strengths, and omissions of collective work to date, considering both the geographic and taxonomic coverage. Finally, we discuss potential uses for these published data that could provide a basis for future synthetic work and suggest guidelines for the collation of such data and future empirical investigations. The 108 studies presented here are the product of many years' work by many researchers and, if consolidated, would provide a solid foundation for our understanding of processes generating biodiversity in the region. Here we attempt to aggregate these efforts, identify significant areas of overlap or gaps, and suggest a standard platform for synthesis and collaboration.

Methods

A literature search was conducted using Web of Science (Thomson Reuters) on 21 June, 2012. The search terms were chosen to maximize inclusion of articles containing georeferenced data on population genetic diversity from shallow water marine habitats in the Indo-Pacific region. The following sets of Boolean search terms were combined in the Advanced Search tool: (1) gene flow OR population structure OR genetic diversity OR phylogeograph* OR F statistic OR mtDNA OR microsatellite OR allozyme, AND (2) Marine OR coral OR reef OR intertidal OR subtidal OR estuar*, AND (3) Indo* OR Malay* OR Indo-Pacific OR Coral Triangle. The titles and abstracts of all papers were assessed and articles that did not fit the following criteria were discarded: (1) only marine animals; (2) only tropical Indo-Pacific studies bounded by the geographical limits of north to Tropic of Cancer (30°N), south to Tropic of Capricorn (30°S), west to Cape Town, South Africa (20°E), and east to the Eastern Pacific Barrier (125°W); (3) only data based on DNA sequences, microsatellites or allozymes; (4) at least three populations were sampled per included study; (5) at least five individuals from a population had to be sampled for that population to be included; (6) the study had to provide sample sizes and indices of genetic diversity at the population level; and (7) the study had to provide latitude and longitude or a map/description of the sampling sufficient to permit location of sampling sites to within 500 km.

Articles remaining in the data set after this first pass were more closely investigated (reading the text of the introduction, methods, results, or supplementary material) to assess their fit to these criteria. Details of the authors, year of publication, genetic marker, sample size, species name, and population geographic positions from articles meeting all criteria were recorded. These criteria targeted population genetic articles, so that purely phylogenetic studies were usually discarded due to low population number or sample sizes. The resulting list was checked by experts in the field attending a catalysis meeting at the National Evolutionary Synthesis Center (NESCent) on the "Molecular Ecology and Evolution of the Indo-Pacific" and some relevant papers not captured by the literature search were added.

All maps were produced in ArcMap (version 10, ESRI, Redlands, CA) using coastline data from the Global, Self-consistent, Hierarchical, High-resolution Shoreline Database (Wessel and Smith 1996). Geographic coverage of studies and species was estimated by measuring the area of a convex hull drawn around the point locations of each study or species. Probability-based species range maps

were downloaded from Aquamaps (http://www.aquamaps.org). Sampling locality polygons were generated by buffering each data point by 60 km and dissolving to merge points close to each other. Species range polygons were drawn in ArcMap 10 using a convex hull of the occurrence points listed in the Global Biodiversity Information Facility (GBIF; http:// www.gbif.org, accessed 13 November, 2012) supplemented by the genetic database points for each species and some points gleaned from the literature for those species not listed on GBIF. Two species [*Echinometra sp. C* (Uehara and Shingaki 1985) and *Euryhaliotrematoides grandis* (see Appendix 1 for species authorities)] were excluded from these analyses because occurrence data could not be found. The GBIF occurrence polygons were merged and joined with sampling locality polygons to generate a count of species for each locality. We then divided the number of species sampled by the total number from the data set present and converted to a percentage. We chose to standardize by the number of species from the dataset with ranges intersecting a location rather than attempting to derive species richness estimates because reliable estimates of species richness across the six phyla that our data set encompasses are difficult to make. The number of species sampled was also divided by total reef area within each polygon to generate sampling per unit area of habitat. Statistical analyses were conducted in R statistical software (R Core Team 2012). These included linear models to assess the relationship between study area and number of locations studied and the differences in study area or number of locations studied among studies using different genetic markers. Study areas were estimated by convex polygons in ArcMap v10 (ESRI, Redlands, CA) and were square-root transformed in R prior to analysis. We also performed a community-style cluster analysis using sampling localities as the groups of interest and sampled species as presence/absence information. Thus, inference from this cluster analysis is about sampling practices rather than actual community composition. We calculated Euclidean distances among sites with the vegdist command in the *vegan* package for R (Oksanen et al. 2012) and clustered them into groups using Ward's Minimum Variance criterion.

Results and Discussion

In total, 493 studies were returned from the initial Web of Science search. This number was reduced to 108 following application of the criteria given above (see Appendix 1). These 108 studies covered 116 species in six phyla. The data set contained 1451 genetic diversity data points, with each point representing a georeferenced collection of a given species (five or more individuals of that species) and genotyped by a category of marker (mtDNA sequencing, microsatellites, or allozymes). In the final data set, there were 725 different geographic locations in 50 different countries.

Taxonomic Patterns.—Among Indo-Pacific genetic studies, there was a clear bias toward ray-finned fishes; just over half (69 of 126) of all species studied were Actinopterygians. The remaining 57 studies surveyed were, in descending order of coverage: Mollusca (16 species), Arthropoda (11 species), Echinodermata (11 species), Cnidaria (4 species), other Chordata (2 species of reptile, 2 species of shark and 1 species of lancelet), and a single representative of the Platyhelminthes (Fig. 1). Thus, large and relatively firm-bodied taxa have been preferred, whereas speciose phyla such as Annelida, Cnidaria, and Porifera have been overlooked, perhaps due to difficulty in identification,

preservation, or subsequent DNA amplification. Even the relatively well-studied Mollusca were under-represented in comparison to fishes when considering their proportional species richness in marine habitats [more than 40,000 estimated species of molluscs (WoRMS Editorial Board 2012) vs 16,764 of Actinopterygians (Eschmeyer et al. 2010)].

This large discrepancy in studies across phyla does not simply result from investigator bias. Invertebrates are generally more difficult to identify to the species level for the non-expert and molecular work is often challenging due to a paucity of genomic information for primer design (Toonen 1997, Fernandez-Silva et al. 2013). In addition, mucus and other polysaccharides commonly found in marine invertebrates are known to inhibit PCR (Huelsken et al. 2011, Vargas et al. 2012). Finally, anthozoans and sponges have a notorious deficiency of variation in their mitochondria (Shearer 2002, Hellberg 2006), such that this useful genetic marker is usually not informative for these taxa (but see Forsman et al. 2009), and development of nuclear markers has generally lagged far behind mtDNA (Karl and Avise 1993, Hare 2001, Puritz et al. 2012). Conversely, fishes are good candidates for population genetic and phylogeographic studies due to their varied life histories and functional traits and their many readily identifiable species. Genetic work tends to be easier in fishes, whose vertebrate affiliation and economic importance mean that there is a plethora of genetic information available for primer design. Despite this overall skew towards fishes, however, the top five most studied species in this data set (based on the number of published studies of that species and by the number of total geographic locations sampled for each species) consisted of four invertebrates and a single fish (discussed in detail in the Text Box, pages 20-21).

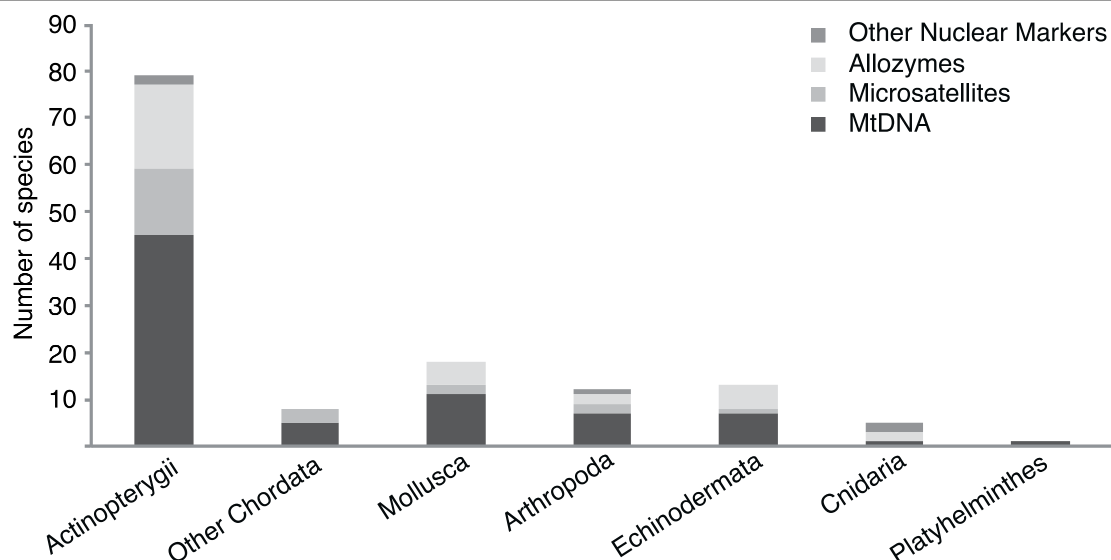


Figure 1. Histogram showing the number of taxa studied by the four categories of molecular marker type.

The bias toward Actinopterygii remains in multi-species studies. There are 19 studies in the final data set that include more than one species, of which 12 were of fishes (Doherty et al. 1995, Dudgeon et al. 2000, Fauvelot and Planes 2002, Drew et al. 2008, Magsino and Juinio-Meñez 2008, Ramon et al. 2008, Thacker et al. 2008, van Herwerden et al. 2009a, Gaither et al. 2010, Mirams et al.

2011, Lord et al. 2012, Ludt et al. 2012) and seven of invertebrates (Palumbi et al. 1997, Uthicke et al. 2001, Barber et al. 2002, Crandall et al. 2008a,b, Kochzius et al. 2009, Duda et al. 2012). Generally, studies include phylogenetically similar species (e.g., for fishes: Fauvelot and Planes 2002, Magsino and Juinio-Meñez 2008, Thacker et al. 2008, Lord et al. 2012, Ludt et al. 2012; and for invertebrates: Palumbi et al. 1997, Uthicke et al. 2001, Barber et al. 2002, Crandall et al. 2008a, Duda et al. 2012). However, an exception is that two multispecies studies have focused on the seastar *Linckia laevigata* and its gastropod parasite *Thyca crystallina* (Crandall et al. 2008b, Kochzius et al. 2009). The comparative context that is offered by multi-species studies is valuable to any attempt to establish general associations between genetic patterns and geography or biological traits (Bowen et al. 2014). It is hoped that future sampling efforts can be coordinated in such a way as to maximize the comparative value of data sets for individual species (see below).

Geographic Scope. — Given the vast area and logistical constraints to fieldwork in the Indo-Pacific, it is not surprising that few Indo-Pacific genetic studies encapsulate the entire geographic range of a species. We examined the geographic scope of studies using a variety of criteria: the geographic extent (area encompassed by sampling), the number of sampling sites, and the density of sampling locations within geographic extent (Fig. 2). The five species with the greatest geographic sampling extent are highlighted in the Text Box. A general perception of population genetic studies is that there is an inherent trade-off between the geographic extent of sampling and the number of sampling sites. That is, some sampling strategies might be expected to include geographically distant sites to maximize the geographic extent of the study but that the expense and logistics of widespread sampling would limit the total number of sites. Other studies might prioritize sampling density and limit themselves to a smaller geographic extent but include more total sites. In addition, it could be that studies of the latter type might preferentially use microsatellites so as to infer recent migration events.

These expectations, however, were not borne out. The area encompassed by individual studies varies widely from 14.8–9092 km^2 (with mean and median values of 2141 and 1892 km^2) and the maximum number of sites is 38 (with a mean and median of 10.4 and 9 per study). There was a slight but significant positive relationship between sampling area (i.e., geographic extent) and number of sites (that can be described by the equation: no. sites = $5.648 + 0.102 (\sqrt{\text{area}})$, $F_{1,150} = 24.96$, $R^2 = 0.143$, $P < 0.00001$, following removal of an outlier study (Johnson et al. 1994), which included eight sites over approximately 14 km^2. Sampling areas differ according to the genetic marker employed by each study (ANOVA: $F_{3,149} = 8.94$, $P < 0.0001$), with the greatest geographic extent for studies using “other” nuclear markers, followed by mtDNA sequences, microsatellites, and allozymes. The difference in area is only significant when comparing allozyme studies to either mtDNA sequence or “other” nuclear marker studies (Tukey’s post hoc tests: both $P < 0.004$). Thus, contrary to expectations, there was no significant difference in sampling area between microsatellite and mtDNA based studies. The number of sites surveyed had no effect on the choice of genetic marker (ANOVA: $F_{3,149} = 1.075$, $P = 0.361$). In summary, then, there was no evidence for trade-offs between sampling extent and number of sites among Indo-Pacific studies.

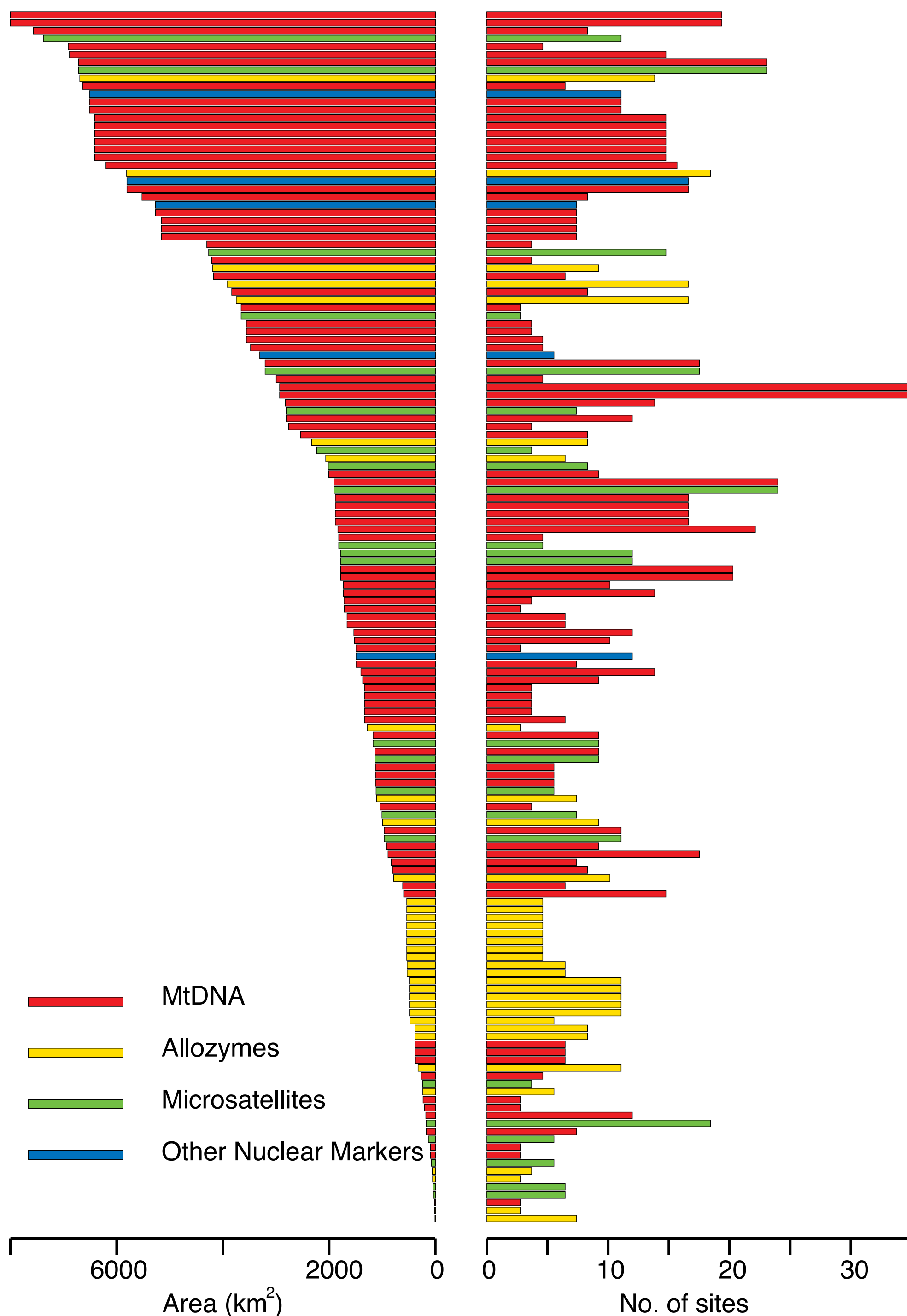


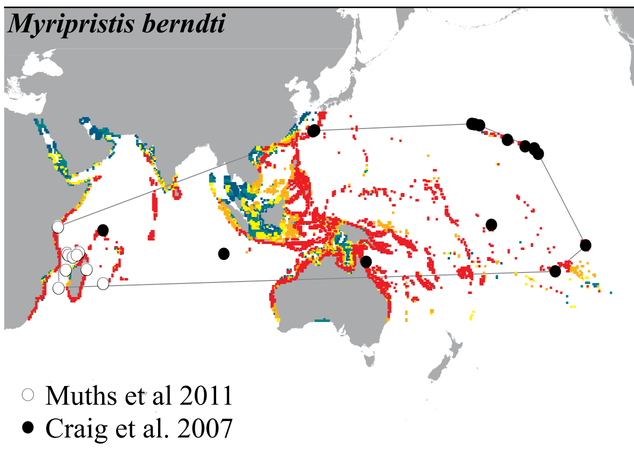
Figure 2. Summary of sampling for genetic surveys included in the present study. Total area surveyed (km^2) and the number sites survey are indicated per species

---

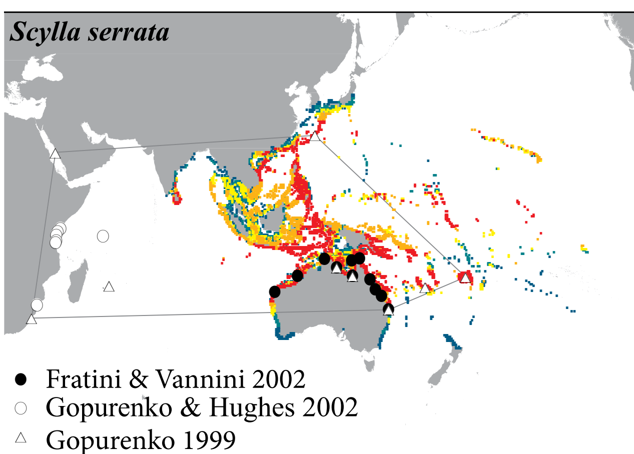
Two noteworthy studies illustrated the lack of inverse correlation between geographic extent of sampling and density of sampling locations. First, the study with the greatest geographical extent (9092 km^2) explored the phylogeographic patterns of *Nerita albicilla* and *Nerita plicata*, two intertidal gastropods (Crandall et al. 2008a) and included a number of evenly-spaced sites (21 sites included in this database) spanning most of the species' range (see Text Box). Second, the study with the maximum number of sampling locations (38 sites) encompassed 3336 km^2 of the Coral Triangle and

▼ The top five species with greatest sampling extent

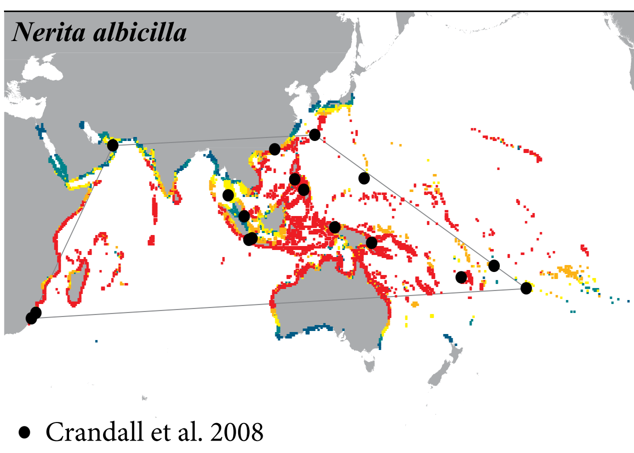
Species sampled across a wide area represent those for which broadscale patterns can be investigated. Heatmap colours show probabilistic occurrence from Aquamaps.org. Symbols show sampling events.



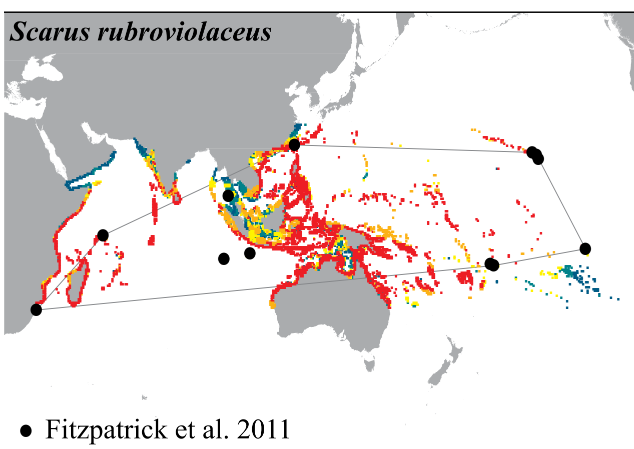
Myripristis berndti Jordan and Evermann, 1903, the bigeye soldierfish, is the species with the widest geographic coverage. The species has been the focus of two population genetic studies: one study restricted to sites around Madagascar (Muths et al. 2011) using mtDNA (*cytochrome oxidase b*) and microsatellites and one study with sites in both the Indian and Pacific Oceans (Craig et al. 2007) using mtDNA (*cyt b*) alone, yet neither study included locations in the Coral Triangle. Extending future coverage to include the Coral Triangle would be an obvious next step for this species.



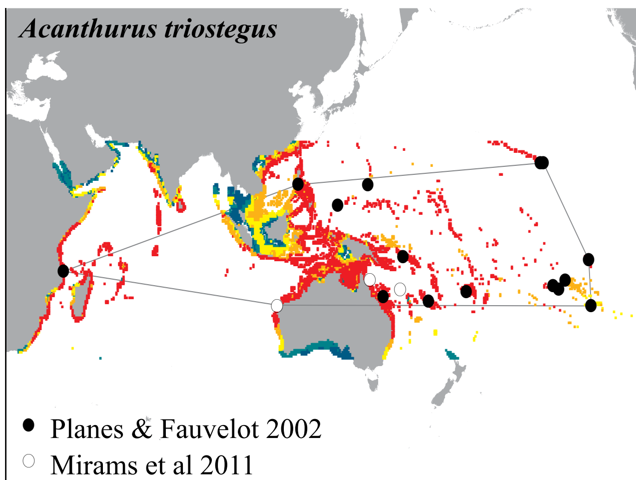
Scylla serrata Forskål, 1775, the mud crab, ranks second in the greatest geographic sampling extent. It has been the focus of three Indo-Pacific studies (Gopurenko 1999, Fratini and Vannini 2002, Gopurenko and Hughes 2002). Each employed mtDNA *COI*, so combination of the data is straightforward. For this reason, further studies on this species should include *COI* sequencing. With the exception of the Solomon Islands (Liu et al. 2007 (not captured by this search, part of a synthesis by Fratini et al., 2010)), there are no data for *S. serrata* from the Coral Triangle.



Nerita albicilla Linnaeus, 1758, an intertidal gastropod, is the species with the largest geographic coverage represented by a single study. Crandall et al. (2008a) included the species alongside its congener *Nerita plicata* in a comparative study that revealed markedly different patterns of genetic structure between these two closely related and ecologically similar species. It would seem that the majority of the species range has been covered by this study (and by Frey and Vermeij, 2008, although this study was excluded from the dataset as it did not report genetic diversity data).



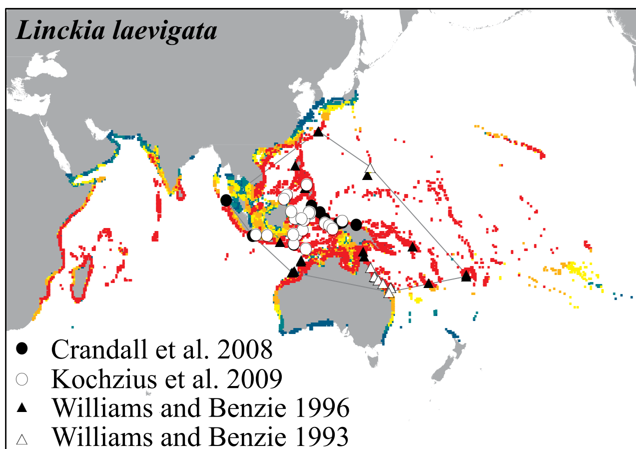
Scarus rubroviolaceus Bleeker, 1847, the Redlip parrotfish, has been surveyed from South Africa to the Marquesas within a single study. Fitzpatrick et al. (2011) used patterns of genetic structure in this species to distinguish between hypotheses explaining the diversity hotspot found in the Coral Triangle. The sampling in this study covered the edges of the species range fairly well, with the exception of the Coral Triangle itself. Sites in the Coral Triangle would enhance the understanding of processes behind patterns of high diversity at the juncture between the Indian and Pacific Oceans.



Acanthurus triostegus Linnaeus, 1758, the convict surgeonfish, has a range spanning the Indian and Pacific oceans. This is the fifth widest sampled species included, with 7879 km^2 covered by two studies. Planes and Fauvelot (2002) used allozymes to assess population structure in the Pacific Ocean, but sampled only a single location in the Indian Ocean and none in the Coral Triangle. Mirams et al. (2011) used mtDNA (*COI*) to investigate the effect of the Torres Strait landbridge, sampling two sites in the Pacific and one in Indian Ocean. The different molecular markers preclude combination of existing data, but there is scope for further sampling of the Indian Ocean and Coral Triangle.

▼ The top five species with the most locations sampled

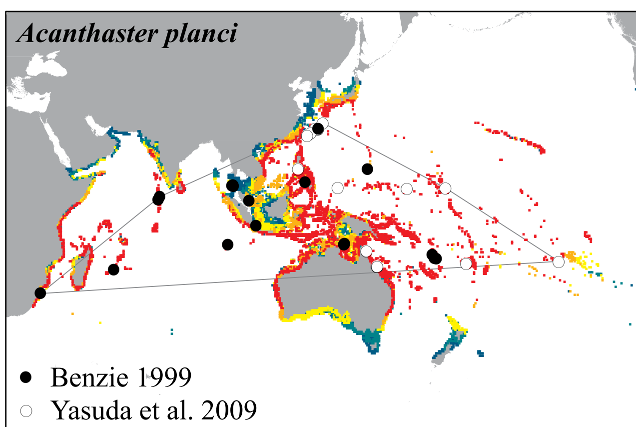
Species sampled from the highest number of locations represent those that may be a potentially fruitful focus for more geographically widespread sampling, even if individuals studies did not encompass a wide expanse of the ocean.



Linckia laevigata Linnaeus, 1758, the blue starfish, has been the focus of four studies covering 59 sites. *Linckia laevigata* and its parasite *Thyca crystallina* (Crandall et al. 2008b; Kochzius et al. 2009) were included in two mtDNA *COI* studies in the Coral Triangle. Two allozyme studies (Williams and Benzie 1993; 1996) sampled more widely, however there remains scope for work across the Indian and Pacific Oceans. Further work should include *COI* to allow data combination. A recent study has made a start on addressing this gap with *COI* sequences from Kenya and Madagascar (Otwoma 2012).

Scylla serrata, the mud crab, has been both densely and widely sampled (map shown above). The density is due to it having been the focus of two fine scale studies in different oceans (Fratini and Vannini 2002; Gopurenko and Hughes 2002), which drives up the sampling location numbers and area of the species sampling polygon. There remains scope for further work on this species in the region.

Pterapogon kauderni Koumans, 1933, the Banggai cardinalfish (map not shown), is endemic to Indonesia and Malaysia. It has been the focus of three studies (Bernardi and Vagelli 2004; Hoffman et al. 2005; Vagelli et al. 2008) over most of its range, each using different markers. Further studies on other endemics could reveal mechanisms maintaining small ranges and genetic health of such species.



Acanthaster planci Linnaeus, 1758**,** the Crown of Thorns seastar, has been studied twice in the Indo Pacific (Benzie 1999, Yasuda et al. 2009) at 36 locations. The sampling of the above studies overlaps in the west Pacific but coverage is lacking in the rest of the Pacific. This gap is partially filled by two recent mtDNA papers in the Central Pacific (Timmers et al. 2011 & 2012), however the addition of mtDNA (*control region*) work on this species from the Indian Ocean and Coral Triangle would allow combination of these data.

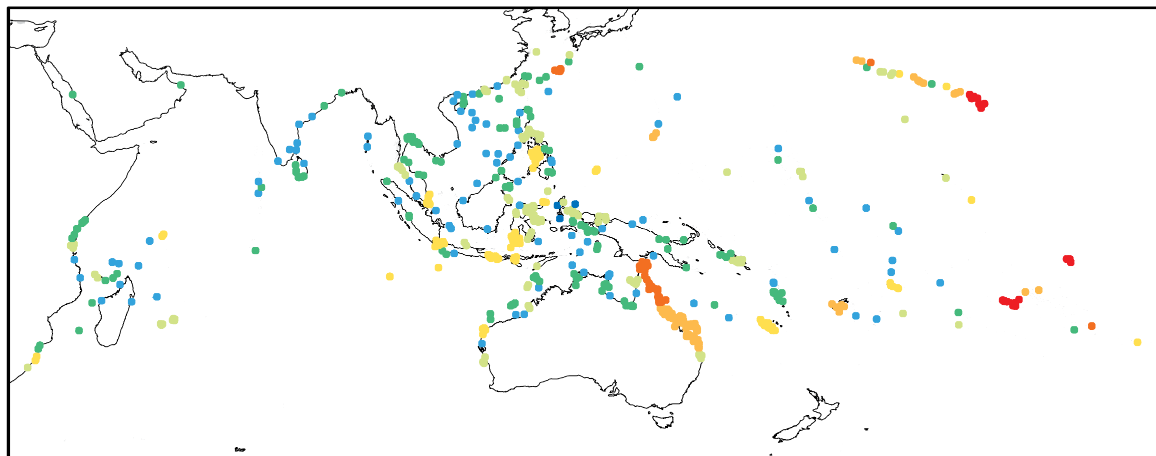
Tridacna crocea Lamarck, 1819, the boring giant clam, has been the focus of two studies at 35 sites in the Coral Triangle (DeBoer et al. 2008; Kochzius and Nuryanto 2008). Coverage in this region is substantial, but absent elsewhere in the species' range. There is wide opportunity for further work on *T. crocea*, particularly using mtDNA *COI* to fit with existing work and increase the geographic scope.

provided comparative data for two closely related species of mantis shrimp (Barber et al. 2002). These two studies have both managed to achieve substantial geographic coverage alongside maintaining a high number of sampling locations and focusing on more than one species.
Another aspect of geographic sampling scope concerns the total number of sampling locations, especially combined across multiple studies. Species that have been included in multiple studies represent opportunities for collaboration and data synthesis, whereby the total geographic scope could be maximized. The five species with the greatest total number of sampling locations are discussed in the Text Box and represent opportunities for synthetic analyses. An extension of the total number of sampling locations is the density of sampling per species (total area covered / number sampling locations). While species with wide geographic sampling covering substantial portions of their range are important for revealing broadscale phylogeographic patterns, studies with dense sampling provide detailed knowledge of connectivity in a small area that may be particularly relevant to marine conservation management actions (Harrison et al. 2012). The top five species in terms of the density of sampling points were: *Craterocephalus capreoli, Pterapogon kauderni, Siganus guttatus, Chromis atripectoralis,* and *Stegastes nigricans*. Such data sets can complement wide-ranging data sets by illuminating population genetic patterns at a small scale but the direct applicability of their findings is necessarily limited to the region in question.

Identifying Anchor Locations. — Given the many difficulties associated with field work in the Indo-Pacific region, it might be expected that researchers would choose to sample in places they or colleagues have sampled before, that are easy to access, or might have colleagues collect for them to reduce the costs. These locations might be established marine stations run by universities or non-governmental organizations, or they may simply be places where a “pioneer” researcher has established a connection and opened up the way for other researchers to follow. This would lead to a pattern of a few sites being the focus of multiple studies on multiple species. From the perspective of future work in the region, such sites can provide “anchor” locations with which direct comparisons can be made among taxa and studies, and therefore their inclusion might be prioritized in future empirical research projects.

Most of the point locations sampled to date were represented for only one species (484 locations out of a total of 682 locations) or by one study (490 locations); however, several localities stand out for the number of species sampled there. Localities are defined here as the polygons created by buffering each data point by 60 km and dissolving to merge points close to each other into a single locality. Figure 3 illustrates these patterns of uneven sampling across the Indo-Pacific region. Figure 3A shows localities colored according to the number of species that have been sampled as a proportion of those species in the data set with ranges intersecting that locality. The Society Islands, the Marquesas, and Main Hawaiian Islands in the central Pacific are localities where sampling has been high relative to the number of species occurring there (>40% of species from the dataset have been sampled). Other potential “anchor” localities identifiable from Figure 3A are the Northern Great Barrier Reef (GBR), Okinawa, Marutea Atoll in the central Pacific, and Pearl and Hermes Atoll in the Northwestern Hawaiian Islands. However, when the percentage of species sampled in a locality is

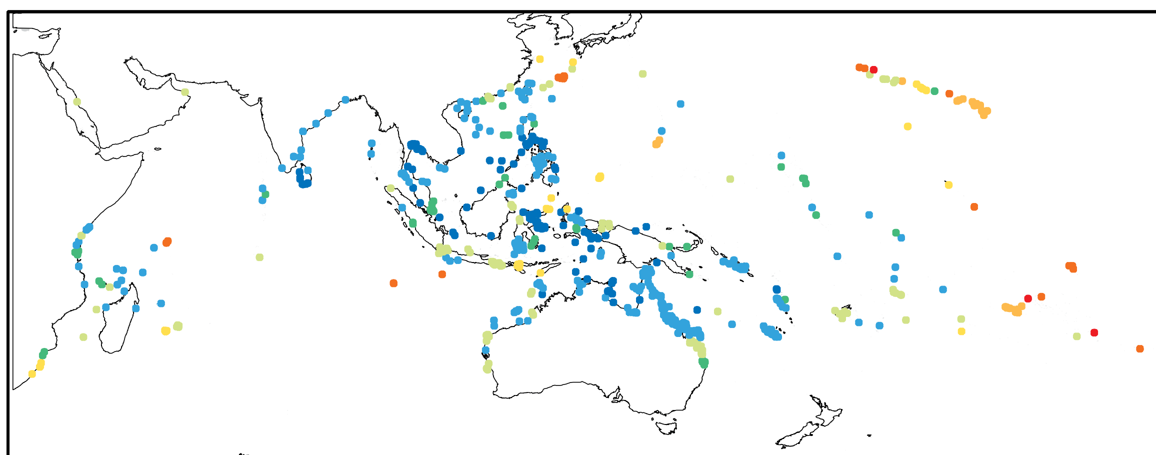
Figure 3. (A) Sampling intensity for the 116 species surveyed. A heatmap coloured by the proportion of studied species is shown per site with a correction for species range. For example, from the main Hawaiian Islands, 22 species have been surveyed and 39 species from the 116 in the data set have species ranges that encompass this location, which gives a percentage of 56.4. (B) Sampling intensity for the 116 species corrected for the area of the study locality. As the locality polygons are of different area depending on the proximity of sampling locations, this correction allows us to see intensity of sampling per unit area. For example, the Main Hawaiian Islands locality has an area of 69,063 km^2, so the corrected sampling intensity is 56.4 / 69,063 or 0.0008.



A

% of species sampled

- 0 - 1
- 1.1 - 2
- 2.1 - 5
- 5.1 - 10
- 10.1 - 20
- 20.1 - 30
- 30.1 - 40
- 40.1 - 56.5

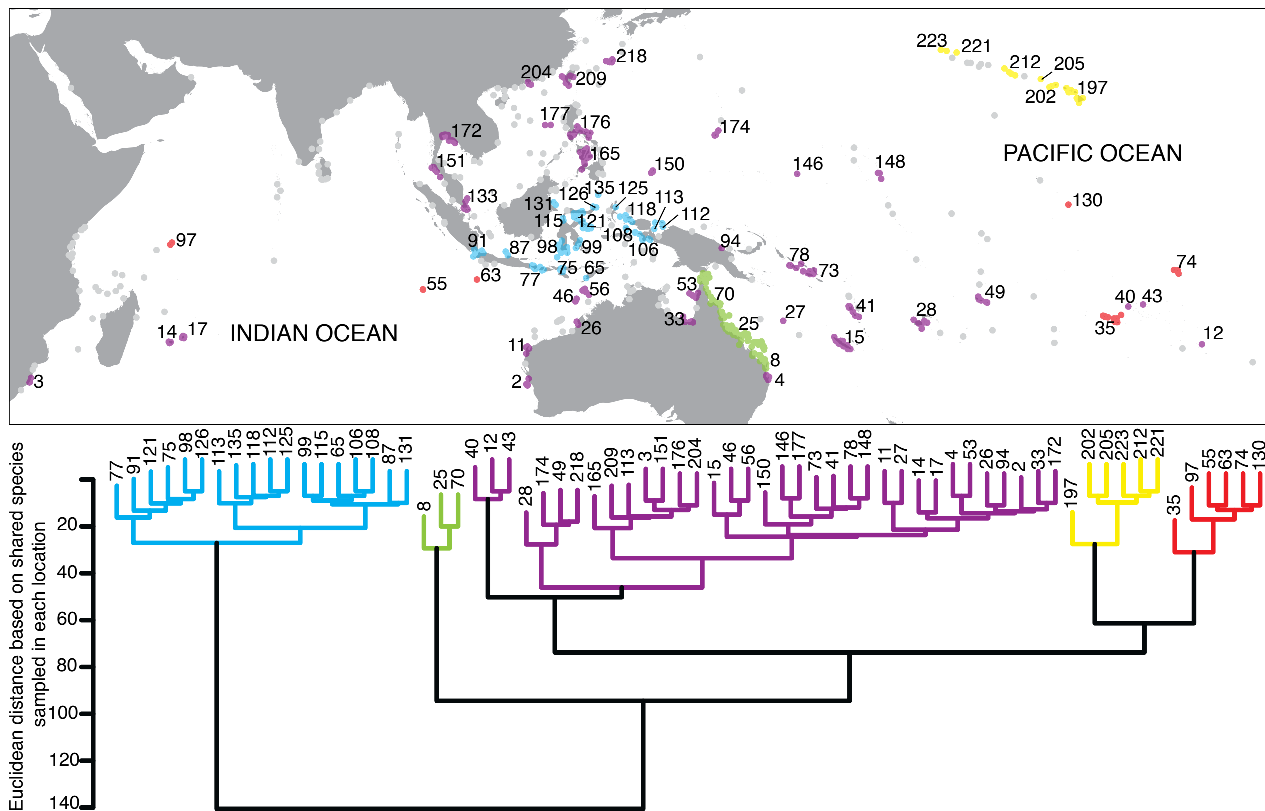


B

(% of species sampled / area of region) x 1000

- 0 - 0.1
- 0.101 - 0.2
- 0.201 - 0.3
- 0.301 - 0.5
- 0.501 - 0.7
- 0.701 - 1.0
- 1.001 - 2.0
- 2.001 - 3.3

Figure 4. Analysis of species co-sampling. Cluster dendrogram is based on squared Euclidean distances among sampling localities, derived from the composition of species that have been co-sampled in each locality. Localities with a higher number of co-sampled species have a lower Euclidean distance between them. Colors show the geographic spread of clusters of co-sampled localities across the Indo-Pacific region. Only localities where more than five species have been surveyed are shown on the map.



---

divided by the area of the locality polygon, we can see patterns of coverage per unit area (Fig. 3B). After this correction has been made, Rangiroa, Takapoto, and Kiritimati in the central Pacific, along with the Seychelles, Christmas Island, and Cocos-Keeling Island in the Indian Ocean emerge as important nodes. Not surprisingly, the GBR localities fade in their influence due to the large area they encompass. Nevertheless, the dense sampling within the GBR, as evidenced by the sizes of the polygons resulting from buffering of point locations, and the relative ease of accessibility of remote reefs here, argues for its inclusion in any list of target locations.

Co-sampled Localities. — The above section illustrates that sampling effort has been uneven across the Indo-Pacific region, with some locations attracting more sampling events than others. Here we investigate whether certain sites are commonly co-sampled, such as might be expected from a situation of reusing the same anchor locations, combining the sampling of many species in a single sampling expedition, or planned multispecies investigations. Figure 4 shows this tendency for subsets of locations to be co-sampled across species. The Hawaiian islands locations form a single cluster (yellow in online version) reflecting the many studies that have sampled multiple locations within

Hawaii. Similarly, sites from the Coral Triangle form a distinct cluster (blue in Fig. 4). The Great Barrier Reef (green in Fig. 4) is strikingly unconnected to other localities, indicative of many studies that have sampled within the GBR only and not included additional locations. Conversely, a suite of isolated oceanic islands from both the Pacific and Indian oceans (red in Fig. 4) have been intensely co-sampled despite their geographic breadth (the Seychelles to the Marquesas, >165° of longitude).

These clusters of sampling effort highlight opportunities for multispecies syntheses (within sampling blocks), but also show how the currently available data limit our ability to make inferences on an oceanic scale. For example, whereas there has been strong sampling effort in the highly biodiverse Coral Triangle (blue in Fig. 4), these results cannot be directly compared to other localities due to a tendency not to co-sample species. Designing future empirical work to link clusters would greatly enhance broadscale geographic inferences, for instance future species sampling from oceanic islands (red in Fig. 4) that targeted well-sampled species from the Coral Triangle (blue in online version) or vice versa, would permit direct comparisons between core and peripheral locations. Some sampling clusters may be driven by certain species only occurring in restricted areas (e.g., endemics), especially Hawaii. The lack of locational co-sampling involving the GBR is noteworthy given the geographic proximity of the GBR to other high profile regions (namely, the Coral Triangle and west Pacific) and its importance as a World Heritage Site.

Mitochondrial Sequencing and Prospects for Combining Data.— Mitochondrial DNA sequences have been the markers of choice for genetic studies in the Indo-Pacific region (Fig. 1). For studies of invertebrates, mitochondrial COI is clearly the locus of preference (with 28 of 30 studies using COI). For chordates, including bony fishes, there is a greater diversity of target loci, with mitochondrial control region being the most common (32 studies), followed by mitochondrial cytochrome b (17) and cytochrome oxidase subunit I (COI) (12). Thus, a researcher embarking on a genetic survey of a chordate would be well advised to determine with which marker previous work has been conducted on their focal species and pick a target locus accordingly.

Sequence based markers are especially amenable for combining data sets, provided that different studies target the same locus. Markers based on fragment size or charge differences (microsatellites and allozymes, respectively) are not directly comparable across research groups without sharing standards (such as tissues genotyped by each group to form a common frame of reference). Therefore, microsatellite and allozyme studies are often limited to stand-alone examples of genetic patterns for a particular species. Regardless of the marker used, qualitative patterns of divergence can be recognized, however, quantitative analyses rely on data produced from the same marker to control for different mutation rates. Sequence data can be exchanged with fewer concerns about reliability, and most studies currently upload their edited sequences to public repositories [e.g., NCBI Genbank (http://www.ncbi.nlm.nih.gov/genbank), EMBL-Bank (http://www.ebi.ac.uk/embl), and the DNA Data Bank of Japan (http://www.ddbj.nig.ac.jp)]. Thus, the growing collection of mtDNA sequence data is a valuable public resource for the Indo-Pacific research community. However, these data are most useful to other researchers when properly georeferenced (see best practice recommendations below).

As a measure of the current standard of georeferencing, 20 studies were excluded from the data set owing to vague reporting of geographical locations. If a population sampling location could not be identified to within approximately 500 km the location was excluded from the dataset; this resulted in 19 data points covering nine species in 10 localities being excluded from the set of accepted studies.

The emergence of several DNA barcoding initiatives in recent years has led to the gathering of large volumes of mtDNA sequence data for the purposes of identification and cataloguing of biodiversity. DNA barcoding involves the sequencing of a common gene that is informative of species-level differences; the accepted barcode for most animals is a fragment of the mitochondrial COI gene. Global marine barcoding projects include Barcode of Life Datasystems (BOLD), Tree of Life, FishBOL, MarBOL, CnidToL, and the Sponge Barcode Project. When the data from these initiatives are made public, they will greatly increase the geographic and taxonomic scope of available mtDNA COI data. This further argues for the inclusion of this locus in population genetic studies in the region.

Recommendations for Future Work

Data Synthesis as an Approach for Understanding Indo-Pacific Biodiversity. — Greater knowledge regarding the spatial genetics of Indo-Pacific taxa will inform long-standing questions regarding the origin and dynamics of marine biodiversity in the Indian and Pacific oceans. Simply put, these oceans are far too large and their communities far too diverse for any single research group to empirically summarize spatial genetic diversity. Only by combining data across locations and taxa can broadscale emergent patterns be identified. For instance, where are the geographic locations of genetic disjunctions and how do they differ among species? Are there biological traits that influence the permeability of a barrier to gene flow? In the Coral Triangle, at the juncture of the Indian and Pacific oceans, there appear to be many instances of genetic breaks (Carpenter et al. 2011), but how such barriers differ among taxa is poorly resolved. Conversely, within the Hawaiian Archipelago, concordant genetic breaks are observed across broad taxonomic lines that are not obvious from any of the single-species studies to date (Toonen et al. 2011). Competing hypotheses regarding broadscale patterns of species diversity (Bellwood et al. 2012) invoke asymmetric migration or colonization. Population genetics provides tools to estimate these asymmetries and this could be done for many taxa. While the main objective of the present study has been toward compiling studies listing genetic diversity data, the practice of using phylogenies combined with information about range size and location alongside species traits has been gaining ground in recent years (Meyer 2003, Paulay and Meyer 2006, Selkoe et al. 2010, Choat et al. 2012). These recent papers provide examples of the kind of synthetic work that can be done with existing data to make sense of the bewildering array of biodiversity in the Indo-Pacific region.

Traditional population genetic reviews have been based on qualitative assessment of published works, which are being complemented by a growing literature using quantitative tests of specific hypotheses (examples from the Indo-Pacific region include: Meyer 2003, Lessios and Robertson 2006, Paulay and Meyer 2006, Hickerson and Meyer 2008, Crandall et al. 2012), and some rely upon reusing previously

published data (Bradbury and Bentzen 2007, Weersing and Toonen 2009, Mirams et al. 2011, Riginos et al. 2011, Selkoe and Toonen 2011). Multiple-species studies are essential for addressing questions about how geography and biological traits affect genetic diversity and partitioning, as species are the unit of replication. Concordant patterns among species support scenarios whereby shared geographic features contribute to similar population genetic structure (Avise 2000), and comparisons among closely related taxa can reduce evolutionary variance when searching for commonalities or points of contrast between species (Dawson 2012). Many research programs are purposefully co-sampling numerous taxa, although the theory for simultaneous statistical evaluation of multiple species is not well developed (see Hickerson and Meyer 2008 for an important exception and example).

Thus, the potential value of any single study exceeds one or two standalone publications. The value of such data synthesis is becoming apparent across the fields of ecology and evolutionary biology and, concurrently, a cultural shift is underway whereby many funding bodies (including the National Science Foundation in the USA, the National Environment Research Council in the UK, and the Deutsche Forschungsgemeinschaft in Germany) and journals (see http://www.datadryad.org/pages/jdap for a list of journals in evolution and ecology) are requiring that raw data be accessible. Ensuring raw data are made public provides direct benefits to the scientific community, including long-term preservation, verifiability, and availability for data reuse (Tenopir et al. 2011, Whitlock 2011), and also to the publishing author, as citation rates are higher for papers that make their data available (Whitlock 2011). Despite this apparent shift, there are no official guidelines or consensus as to what constitutes essential data elements for population genetics so that what is reported across studies varies widely.

Best Practices for Reporting Population Genetic Data.— Here, we outline the minimal scope of a population genetic survey and the aspects of data that should be reported for Indo-Pacific studies to maximize the continued utility of published work to the scientific community. We recommend that a population genetic survey include an absolute minimum of three populations to allow partitioning of diversity among locations. Whereas in this survey we include studies sampling as few as five individuals per population in the interests of representing a full range of studies, we believe that targets of at least 15 individuals per population for mitochondrial sequence data and 20 for microsatellites and SNPs may provide reasonable estimates of diversity by location but that sample size per location should be targeted higher (n=50) if possible (Nei 1978; Kalinowski 2005; Pruett and Winker 2008; Hale et al 2012). Power analyses should be run to estimate the sample size required for the number of loci under consideration. Data reporting should include the latitude and longitude of each sampled population with a detailed map as a beneficial complement. Dates of collections are essential to allow the consideration of temporal change.

In addition to the sample size and exact sampling locations, our opinion is that a population genetic study should report frequently used summary statistics that provide commonality across studies. This includes reporting diversity per location [haplotype diversity (h) and average pairwise differences (π) for sequence data, allelic diversity/number of alleles (A) for microsatellites, minor allele frequency

for SNPs, and both observed (H_O) and expected (H_E) heterozygosity for genotype data]. Authors should also report measures of population differentiation (with F-statistics being the most commonly reported differentiation statistics). Although reporting of F-statistics has differed based on data type (G_{ST} and similar statistics being used for genotype data, and N_{ST} statistics for sequence data), recent work has identified the need to consider allelic diversity (Hedrick 2005) so that metrics correcting for this diversity might be in order (see Bird et al. 2011 for an extended discussion). Neutrality test scores for sequence data are also useful, with Tajima's D and FS (Tajima 1989, Fu 1997) being commonly used, though there is disagreement about the suitability of current null models for mitochondrial sequence data (Wares 2009). If nothing more, the diversity of possible estimators underscores the necessity of making all raw data available so that new approaches can be applied to old data.

Edited, georeferenced sequence data of unique haplotypes should be accessioned at NCBI, EMBL, or DDBJ and accession numbers reported. Unedited, georeferenced sequences (e.g., FASTA files) for all individuals, along with input files for all reported statistics (e.g., NEXUS, XML, or .parm files, etc.), should be deposited in a flexible online data repository such as Dryad (http://www.datadryad.org) for studies to be fully transparent and repeatable. Many studies (examples include work from some of the authors on this paper) only take partial steps toward such accessioning, such as depositing sequences of unique haplotypes only or labeling accessions in a manner whereby the geographic origins are unclear.

For multilocus genotype data, ideally full genotypes of all individuals should be made available and their geographic origins explicit. However, there is no easily searchable public repository designed for such data at present. The creation of a shared database including all Indo-Pacific population genetic data would allow such data to be housed and, if integrated with a collaborative online research forum, would facilitate further progress in the field. Until such infrastructure exists, placing full georeferenced genotype information in Dryad or appending files as supplements to the published paper would represent best practice. Because sequence data, especially from mtDNA, can be readily consolidated among research groups (see previous section), there is a distinct advantage to including mtDNA sequences as part of all future genetic surveys.

While the above practices will help maintain consistency across population genetic studies and facilitate collation of data, sampling for these studies also yields data useful to users outside the field of molecular ecology. For example, the locations of sampled populations can add data to occurrence databases for the species, allowing refinement of species range maps and the mapping of species richness patterns. At present, there is a notable mismatch between the locations of occurrence data points held in GBIF and those for the same species from population genetic studies; this can be easily solved by integrating these useful online repositories.

The recommendations laid out in this final section are likely to be familiar to most readers; we are not suggesting a major shift, merely a strengthening of the system already in place and the potential addition of more streamlined workflows. In an age where genetic data are increasingly numerous and

funding agencies are increasingly frugal, we have a responsibility to make the most out of the existing data, compile new data in easily accessible ways and foster collaborative synthesis across regions with a view to tackling some of the “big” questions regarding marine biodiversity in the Indo-Pacific region.

Acknowledgments

The concept for this study arose from a Catalysis meeting funded by the National Evolutionary Synthesis Center (NESCent) on the Molecular Ecology and Evolution of the Indo-Pacific to EC and CR. This material is based upon work supported by the National Science Foundation through the National Evolutionary Synthesis Center (NESCent) under grant number NSF #EF-0905606. JK is supported by a University of Queensland International and University of Queensland Research Scholarships. Thanks to L Rocha and H Lessios for helpful suggestions during the data collection phase.

Literature Cited

Antoro S, Na-Nakorn U, Koedprang W. 2006. Study of genetic diversity of orange-spotted grouper, *Epinephelus coioides*, from Thailand and Indonesia using microsatellite markers. Mar Biotechnol. 8(1):17–26. PMid:16331343. http://dx.doi.org/10.1007/s10126-005-5026-0

Avise JC. 2000. Phylogeography: the history and formation of species. Cambridge, MA: Harvard University Press.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Syst. 18:489–522.

Barber P, Boyce SL. 2006. Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. Proc Roy Soc London Ser B-Biol Sci. 273(1597):2053– 2061. PMid:16846913. PMCid:PMC1635474. http://dx.doi.org/10.1098/rspb.2006.3540

Barber PH, Bellwood DR. 2005. Biodiversity hotspots: evolutionary origins of biodiversity in wrasses (Halichoeres: Labridae) in the Indo-Pacific and new world tropics. Mol Phylogenet Evol. 35(1):235–253. PMid:15737594. http://dx.doi.org/10.1016/j.ympev.2004.10.004

Barber PH, Moosa MK, Palumbi SR. 2002. Rapid recovery of genetic populations on Krakatau: diversity of stomatopod temporal and spatial scales of marine larval dispersal. Proc Roy Soc London Ser B-Biol Sci. 269(1500):1591–1597. PMid:12184829. PMCid:PMC1691063. http://dx.doi.org/10.1098/rspb.2002.2026

Bastidas C, Benzie JAH, Uthicke S, Fabricius KE. 2001. Genetic differentiation among populations of a broadcast spawning soft coral, Sinularia flexibilis, on the Great Barrier Reef. Mar Biol. 138(3):517–525. http://dx.doi.org/10.1007/s002270000480

Bastidas C, Benzie JAH, Fabricius KE. 2002. Genetic differentiation among populations of the brooding soft coral *Clavularia koellikeri* on the Great Barrier Reef. Coral Reefs. 21:233–241.

Bay LK, Choat JH, van Herwerden L, Robertson DR. 2004. High genetic diversities and complex genetic structure in an Indo-Pacific tropical reef fish (*Chlorurus sordidus*): evidence of an unstable evolutionary past? Mar Biol. 144(4):757–767. http://dx.doi.org/10.1007/s00227-003-1224-3

Beger M, Selkoe K, Treml E, Barber PH, von der Heyden S, Crandall E, Toonon RJ, Riginos C. 2014. Evolving coral reef conservation with genetic information. Bull Mar Sci. 90(1):159-185.

Bellwood DR, Hughes TP. 2001. Regional-scale assembly rules and biodiversity of coral reefs. Science. 292(5521):1532–1535. PMid:11375488. http://dx.doi.org/10.1126/science.1058635

Bellwood DR, Renema W, Rosen BR. 2012. Biodiversity hotspots, evolution and coral reef biogeography: a review. In: Gower D, Johnson K, Richardson R, Rosen B, Rüber L, Williams S, editors. Biotic evolution and environmental change in southeast Asia. Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511735882.011

Benzie J. 1998. Genetic structure of marine organisms and SE Asian biogeography. In: Hall R, Hollway JD, editors. Biogeography and geological evolution of SE Asia. Kerkwerve, the Netherlands: Backhuys Publishers. p. 197–209.

Benzie JAH. 1999. Major genetic differences between crown-of-thorns starfish (*Acanthaster planci*) populations in the Indian and Pacific oceans. Evolution. 53(6):1782–1795. http://dx.doi.org/10.2307/2640440

Benzie JAH, Williams ST. 1995. Gene flow among giant clam (*Tridacna gigas*) populations in Pacific does not parallel ocean circulation. Mar Biol. 123(4):781–787. http://dx.doi.org/10.1007/BF00349121

Benzie JAH, Williams ST. 1997. Genetic structure of giant clam (*Tridacna maxima*) populations in the west Pacific is not consistent with dispersal by present-day ocean currents. Evolution. 51(3):768–783. http://dx.doi.org/10.2307/2411153

Bernardi G, Vagelli A. 2004. Population structure in Banggai cardinalfish, *Pterapogon kauderni,* a coral reef species lacking a pelagic larval phase. Mar Biol. 145:803–810.

Bird CE, Holland BS, Bowen BW, Toonen RJ. 2007. Contrasting phylogeography in three endemic

Hawaiian limpets (*Cellana* spp.) with similar life histories. Mol Ecol. 16:3173–3186. PMid:17651195. http://dx.doi.org/10.1111/j.1365-294X.2007.03385.x

Bird CE, Karl SA, Smouse PE, Toonen RJ. 2011. Detecting and measuring genetic differentiation. In: Held C, Koenemann S, Schubart CD, editors. Phylogeography and population genetics in crustacea. CRC Press. p. 400. http://dx.doi.org/10.1201/b11113-4

Bourjea J, Lapegue S, Gagnevin L, Broderick D, Mortimer JA, Ciccione S, Roos D, Taquet C, Grizel H. 2007. Phylogeography of the green turtle, *Chelonia mydas*, in the southwest Indian Ocean. Mol Ecol. 16(1):175-186. PMid:17181729. http://dx.doi.org/10.1111/j.1365-294X.2006.03122.x

Bowen BW, Bass AL, Rocha LA, Grant WS, Robertson DR. 2001. Phylogeography of the trumpetfishes (*Aulostomus*): ring species complex on a global scale. Evolution. 55(5):1029–1039. http://dx.doi.org/10.1554/0014-3820(2001)055[1029:POTTAR]2.0.CO;2

Bowen BW, Rocha LA, Toonen RJ, Karl SA, Craig MT, DiBattista JD, Eble JA, Gaither MR, Skillings D, Bird CJ. 2013. The origins of tropical marine biodiversity. Trends Ecol Evol. 28(6):359–366. PMid:23453048. http://dx.doi.org/10.1016/j.tree.2013.01.018

Bowen BW, Shanker K, Yasuda N, Malay MCD, von der Heyden S, Paulay G, Rocha LA, Selkoe KA, Barber PH, Williams ST, et al. 2014. Phylogeography unplugged: comparative geographic surveys in the genomic era. Bull Mar Sci. 90(1):13-46.

Bradbury IR, Bentzen P. 2007. Non-linear genetic isolation by distance: implications for dispersal estimation in anadromous and marine fish populations. Mar Ecol Prog Ser. 340:245–257. http://dx.doi.org/10.3354/meps340245

Briggs JC. 1974. Marine zoogeography. New York: McGraw-Hill.

Briggs JC. 1999. Coincident biogeographic patterns: Indo-West Pacific Ocean. Evolution. 53(2):326–335. http://dx.doi.org/10.2307/2640770

Carpenter KE, Barber PH, Crandall ED, Manjaji-Matsumoto BM, Juinio-Me-ez MA, Santos MD, Starger CJ, Toha AHA. 2011. Comparative phylogeography of the Coral Triangle and implications for marine management. J Mar Biol. 2011:Article ID 396982, 14 p.

Carpenter KE, Springer VG. 2005. The center of the center of marine shore fish biodiversity: the Philippine Islands. Environ Biol Fish. 72(4):467–480. http://dx.doi.org/10.1007/s10641-004-3154-4

Chenoweth SF, Hughes JM. 2003. Oceanic interchange and nonequilibrium population structure in the estuarine dependent Indo-Pacific tasselfish, *Polynemus sheridani*. Mol Ecol. 12(9):2387-2397.

PMid:12919476. http://dx.doi.org/10.1046/j.1365-294X.2003.01921.x

Chenoweth SF, Hughes JM, Keenan CP, Lavery S. 1998. Concordance between dispersal and mitochondrial gene flow: isolation by distance in a tropical teleost, *Lates calcarifer* (Australian barramundi). Heredity. 80:187–197. http://dx.doi.org/10.1046/j.1365-2540.1998.00292.x

Choat JH, Klanten OS, Van Herwerden L, Robertson DR, Clements KD. 2012. Patterns and processes in the evolutionary history of parrotfishes (Family Labridae). Biol J Linnean Soc. 107(3):529–557. http://dx.doi.org/10.1111/j.1095-8312.2012.01959.x

Concepcion GT, Kahng SE, Crepeau MW, Franklin EC, Coles SL, Toonen RJ. 2010. Resolving natural ranges and marine invasions in a globally distributed octocoral (genus *Carijoa*). Mar Ecol Prog Ser. 401:113–127. http://dx.doi.org/10.3354/meps08364

Connolly SR, Bellwood DR, Hughes TP. 2003. Indo-Pacific biodiversity of coral reefs: deviations from a mid-domain model. Ecology. 84(8):2178–2190. http://dx.doi.org/10.1890/02-0254

Craig MT, Eble JA, Bowen BW, Robertson DR. 2007. High genetic connectivity across the Indian and Pacific oceans in the reef fish *Myripristis berndti* (Holocentridae). Mar Ecol Prog Ser. 334:245–254. http://dx.doi.org/10.3354/meps334245

Crandall ED, Frey MA, Grosberg RK, Barber PH. 2008a. Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. Mol Ecol. 17(2):611–626. PMid:18179436. http://dx.doi.org/10.1111/j.1365-294X.2007.03600.x

Crandall ED, Jones ME, Mu-oz MM, Akinronbi B, Erdmann MV, Barber PH. 2008b. Comparative phylogeography of two seastars and their ectosymbionts within the Coral Triangle. Mol Ecol. 17(24):5276–5290. PMid:19067797. http://dx.doi.org/10.1111/j.1365-294X.2008.03995.x

Crandall ED, Treml EA, Barber PH. 2012. Coalescent and biophysical models of long-distance dispersal dynamics in Neritid snails. Mol Ecol. 21(22):5579–5598. PMid:23050562. http://dx.doi.org/10.1111/mec.12031

Daly-Engel TS, Randall JE, Bowen BW. 2012. Is the great barracuda (*Sphyraena barracuda*) a reef fish or a pelagic fish? The phylogeographic perspective. Mar Biol. 159(5):975–985. http://dx.doi.org/10.1007/s00227-012-1878-9

Dawson MN. 2012. Parallel phylogeographic structure in ecologically similar sympatric sister taxa. Mol Ecol. 21(4):987–1004. PMid:22229665. http://dx.doi.org/10.1111/j.1365-294X.2011.05417.x

DeBoer TS, Subia MD, Ambariyanto, Erdmann MV, Kovitvongsa K, Barber PH. 2008. Phylogeography

and limited genetic connectivity in the endangered boring giant clam across the Coral Triangle. Conserv Biol. 22(5):1255–1266. PMid:18637905. http://dx.doi.org/10.1111/j.1523-1739.2008.00983.x

de Croos M, Palsson S. 2010. Mitochondrial DNA variation and population genetic structure of white shrimp *Fenneropenaeus indicus* along the coastal belt of Sri Lanka. Aquat Living Resour. 23(3):315–323. http://dx.doi.org/10.1051/alr/2010027

Dethmers KEM, Broderick D, Moritz C, Fitzsimmons NN, Limpus CJ, Lavery S, Whiting S, Guinea M, Prince RIT, Kennett R. 2006. The genetic structure of Australasian green turtles (*Chelonia mydas*): exploring the geographical scale of genetic exchange. Mol Ecol. 15(13):3931–3946. PMid:17054494. http://dx.doi.org/10.1111/j.1365-294X.2006.03070.x

DiBattista JD, Waldrop E, Bowen BW, Schultz JK, Gaither MR, Pyle RL, Rocha LA. 2012. Twisted sister species of pygmy angelfishes: discordance between taxonomy, coloration, and phylogenetics. Coral Reefs. 31:839–851. http://dx.doi.org/10.1007/s00338-012-0907-y

DiBattista JD, Wilcox C, Craig MT, Rocha LA, Bowen BW. 2011. Phylogeography of the Pacific blueline surgeonfish, *Acanthurus nigroris*, reveals high genetic connectivity and a cryptic endemic species in the Hawaiian Archipelago. J Mar Biol. 2011:1–17. http://dx.doi.org/10.1155/2011/839134

Doherty PJ, Planes S, Mather P. 1995. Gene flow and larval duration in seven species of fish from the Great Barrier Reef. Ecology. 76(8):2373–2391. http://dx.doi.org/10.2307/2265814

Drew J, Allen GR, Kaufman L, Barber PH. 2008. Endemism and regional color and genetic differences in five putatively cosmopolitan reef fishes. Conserv Biol. 22(4):965–975. PMid:18786099. http://dx.doi.org/10.1111/j.1523-1739.2008.01011.x

Duda TF, Lee T. 2009. Isolation and population divergence of a widespread Indo-west Pacific marine gastropod at Easter Island. Mar Biol. 156(6):1193-1202. http://dx.doi.org/10.1007/s00227-009-1161-x

Duda TF, Lessios HA. 2009. Connectivity of populations within and between major biogeographic regions of the tropical Pacific in *Conus ebraeus*, a widespread marine gastropod. Coral Reefs. 28(3):651–659. http://dx.doi.org/10.1007/s00338-009-0485-9

Duda TF, Palumbi SR. 1999. Population structure of the black tiger prawn, *Penaeus monodon*, among western Indian Ocean and western Pacific populations. Mar Biol. 134(4):705–710. http://dx.doi.org/10.1007/s002270050586

Duda TF, Terbio M, Chen G, Phillips S, Olenzek AM, Chang D, Morris DW. 2012. Patterns of population structure and historical demography of *Conus* species in the tropical Pacific. Am Malacological Bull.

30(1):175–187. http://dx.doi.org/10.4003/006.030.0116

Dudgeon CL, Broderick D, Ovenden JR. 2009. IUCN classification zones concord with, but underestimate, the population genetic structure of the zebra shark *Stegostoma fasciatum* in the Indo-West Pacific. Mol Ecol. 18(2):248–261. PMid:19192179. http://dx.doi.org/10.1111/j.1365-294X.2008.04025.x

Dudgeon CL, Gust N, Blair D. 2000. No apparent genetic basis to demographic differences in scarid fishes across continental shelf of the Great Barrier Reef. Mar Biol. 137(5–6):1059–1066. http://dx.doi.org/10.1007/s002270000403

Eble JA, Rocha LA, Craig MT, Bowen BW. 2011a. Not all larvae stay close to home: insights into marine population connectivity with a focus on the brown surgeonfish (*Acanthurus nigrofuscus*). J Mar Biol. 2011:1–12. http://dx.doi.org/10.1155/2011/518516

Eble JA, Toonen RJ, Sorenson L, Basch LV, Papastamatiou YP, Bowen BW. 2011b. Escaping paradise: larval export from Hawaii in an Indo-Pacific reef fish, the yellow tang *Zebrasoma flavescens*. Mar Ecol Prog Ser. 428:245–258. http://dx.doi.org/10.3354/meps09083

Ekman S. 1935. Tiergeographie des meeres. Leipzig: Akademische Verlagsgesellschaft MBH.

Ekman S. 1953. Zoogeography of the sea. London: Sidgwick and Jackson.

Eschmeyer WN, Fricke R, Fong JD, Polack DA. 2010. Marine fish diversity: history of knowledge and discovery (Pisces). Zootaxa. 2525:19–50.

Fauvelot C, Planes S. 2002. Understanding origins of present-day genetic structure in marine fish: biologically or historically driven patterns? Mar Biol. 141(4):773–788. http://dx.doi.org/10.1007/s00227-002-0869-7

Fernandez-Silva I, Whitney J, Wainwright B, Andrews KR, Toonen RJ, Ylitalo-Ward H, Bowen BW, Goetze E, Karl SA. 2013. Microsatellites for next-generation ecologists: a post-sequencing bioinformatics pipeline. PLoS One. 8(2):e55990. PMid:23424642. PMCid:PMC3570555. http://dx.doi.org/10.1371/journal.pone.0055990

Fitzpatrick JM, Carlon DB, Lippe C, Robertson DR. 2011. The west Pacific diversity hotspot as a source or sink for new species? Population genetic insights from the Indo-Pacific parrotfish *Scarus rubroviolaceus*. Mol Ecol. 20(2):219–234. PMid:21143329. http://dx.doi.org/10.1111/j.1365-294X.2010.04942.x

Forbes E. 1856. Map of the distribution of marine life. In: Johnston AK, editors. The physical atlas of

natural phenomena (new edition). London: W Johnston and AK Johnston. p. 31.

Forsman Z, Barshis D, Hunter C, Toonen R. 2009. Shape-shifting corals: molecular markers show morphology is evolutionarily plastic in *Porites*. BMC Evol Biol. 9(1):45. PMid:19239678. PMCid:PMC2656472. http://dx.doi.org/10.1186/1471-2148-9-45

Fratini S, Ragionieri L, Cannicci S. 2010. Stock structure and demographic history of the Indo-West Pacific mud crab *Scylla serrata*. Estuar Coast Shelf Sci. 86(1):51–61. http://dx.doi.org/10.1016/j.ecss.2009.10.009

Fratini S, Vannini M. 2002. Genetic differentiation in the mud crab *Scylla serrata* (Decapoda: Portunidae) within the Indian Ocean. J Exp Mar Biol Ecol. 272(1):103–116. http://dx.doi.org/10.1016/S0022-0981(02)00052-7

Frey MA, Vermeij GJ. 2008. Molecular phylogenies and historical biogeography of a circumtropical group of gastropods (Genus: *Nerita*): implications for regional diversity patterns in the marine tropics. Mol Phylogenet Evol. 48(3):1067–1086. PMid:18586528. http://dx.doi.org/10.1016/j.ympev.2008.05.009

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 147(2):915–925. PMid:9335623. PMCid:PMC1208208.

Gaither MR, Bowen BW, Bordenave TR, Rocha LA, Newman SJ, Gomez JA, van Herwerden L, Craig MT. 2011a. Phylogeography of the reef fish *Cephalopholis argus* (Epinephelidae) indicates Pleistocene isolation across the Indo-Pacific barrier with contemporary overlap in the Coral Triangle. BMC Evol Biol. 11:189. PMid:21722383. PMCid:PMC3145601. http://dx.doi.org/10.1186/1471-2148-11-189

Gaither MR, Jones SA, Kelley C, Newman SJ, Sorenson L, Bowen BW. 2011b. High connectivity in the deepwater snapper *Pristipomoides filamentosus* (Lutjanidae) across the Indo-Pacific with Isolation of the Hawaiian Archipelago. PLoS One. 6(12):e28913. PMid:22216141. PMCid:PMC3245230. http://dx.doi.org/10.1371/journal.pone.0028913

Gaither MR, Toonen RJ, Robertson DR, Planes S, Bowen BW. 2010. Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira* and *Lutjanus fulvus*). J Biogeogr. 37(1):133–147. http://dx.doi.org/10.1111/j.1365-2699.2009.02188.x

Gopurenko D, Hughes JM. 2002. Regional patterns of genetic structure among Australian populations of the mud crab, *Scylla serrata* (Crustacea: Decapoda): evidence from mitochondrial DNA. Mar Freshwat Res. 53(5):849–857. http://dx.doi.org/10.1071/MF01225

Gopurenko D, Hughes JM, Keenan CP. 1999. Mitochondrial DNA evidence for rapid colonisation of the Indo-West Pacific by the mudcrab *Scylla serrata*. Mar Biol. 134:227–233. http://dx.doi.org/10.1007/s002270050541

Haig JA, Connolly RM, Hughes JM. 2010. Little shrimp left on the shelf: the roles that sea-level change, ocean currents and continental shelf width play in the genetic connectivity of a seagrass-associated species. J Biogeogr. 37(8):1570–1583.

Hale ML, Burg TM, Steeves TE (2012) Sampling for Microsatellite-Based Population Genetic Studies: 25 to 30 Individuals per Population Is Enough to Accurately Estimate Allele Frequencies. PLoS ONE 7(9): e45170. doi: 10.1371/journal.pone.0045170

Han ZQ, Li YZ, Chen GB, Gao TX. 2008. Population genetic structure of coral reef species *Plectorhinchus flavomaculatus* in South China Sea. Afr J Biotechnol. 7(11):1774–1781.

Hare M. 2001. Prospects for nuclear gene phylogeography. Trends Ecol Evol. 16:700–706. http://dx.doi.org/10.1016/S0169-5347(01)02326-6

Harrison HB, Williamson DH, Evans RD, Almany GR, Thorrold SR, Russ GR, Feldheim KA, Van Herwerden L, Planes S, Srinivasan M. 2012. Larval export from marine reserves and the recruitment benefit for fish and fisheries. Curr Biol. 22(11):1023–1028. PMid:22633811. http://dx.doi.org/10.1016/j.cub.2012.04.008

Hedrick PW. 2005. A standardized genetic differentiation measure. Evolution. 59(8):1633–1638. PMid:16329237.

Hellberg ME. 2006. No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. BMC Evol Biol. 6(1):24. PMid:16542456. PMCid:PMC1431588. http://dx.doi.org/10.1186/1471-2148-6-24

Hickerson MJ, Meyer CP. 2008. Testing comparative phylogeographic models of marine vicariance and dispersal using a hierarchical Bayesian approach. BMC Evol Biol. 8(1):322. PMid:19038027. PMCid:PMC2614435. http://dx.doi.org/10.1186/1471-2148-8-322

Hoareau TB, Boissin E, Berrebi P. 2012. Evolutionary history of a widespread Indo-Pacific goby: the role of Pleistocene sea-level changes on demographic contraction/expansion dynamics. Mol Phylogenet Evol. 62(1):566–572. PMid:22037473. http://dx.doi.org/10.1016/j.ympev.2011.10.004

Hoeksema B. 2007. Delineation of the Indo-Malayan centre of maximum marine biodiversity: the Coral Triangle. Biogeogr Time Place: Distributions Barriers Islands. 29:117–178. http://dx.doi.org/10.1007/978-1-4020-6374-9_5

Hoffman EA, Kolm N, Berglund A, Arguello JR, Jones AG. 2005. Genetic structure in the coral- reef-associated Banggai cardinalfish, *Pterapogon kauderni*. Mol Ecol. 14(5):1367–1375. PMid:15813777. http://dx.doi.org/10.1111/j.1365-294X.2005.02538.x

Horne JB, Momigliano P, Welch DJ, Newman SJ, van Herwerden L. 2011. Limited ecological population connectivity suggests low demands on self-recruitment in a tropical inshore marine fish (*Eleutheronema tetradactylum*: Polynemidae). Mol Ecol. 20(11):2291–2306. PMid:21518062. http://dx.doi.org/10.1111/j.1365-294X.2011.05097.x

Horne JB, van Herwerden L, Choat JH, Robertson DR. 2008. High population connectivity across the Indo-Pacific: congruent lack of phylogeographic structure in three reef fish congeners. Mol Phylogenet Evol. 49(2):629–638. PMid:18804542. http://dx.doi.org/10.1016/j.ympev.2008.08.023

Huelsken T, Schreiber S, Hollmann M. 2011. COI amplification success from mucus-rich marine gastropods (Gastropoda: Naticidae) depends on DNA extraction method and preserving agent. Mitteilungen der Deutschen Malakozoologischen Gesellschaft. 85:17–26.

Imron JB, Hale P, Degnan BM, Degnan SM. 2007. Pleistocene isolation and recent gene flow in *Haliotis asinina*, an Indo-Pacific vetigastropod with limited dispersal capacity. Mol. Ecol. 16(2):289–304. PMid:17217345.

Iwamoto K, Takemura A, Yoshino T, Imai H. 2009. Molecular ecological study of *Siganus spinus* and *S. guttatus* from Okinawan waters based on mitochondrial DNA control region sequences. J. Oceanogr. 65(1):103–112. http://dx.doi.org/10.1007/s10872-009-0010-3

Johnson MS, Watts RJ, Black R. 1994. High levels of genetic subdivision on peripherally isolated populations of the atherinid fish *Craterocephalus capreoli* in the Houtman-Abrolhos Islands, Western Australia. Mar Biol. 119(2):179–184. http://dx.doi.org/10.1007/BF00349554

Kalinowski ST. 2005. Do polymorphic loci require large sample sizes to estimate genetic distances? Heredity 94(1):33-6.

Karl SA, Avise JC. 1993. PCR-based assays of Mendelian polymorphisms from anonymous single-copy nuclear DNA: techniques and applications for population genetics. Mol Biol Evol. 10:342–361. PMid:8098128.

Klanten OS, Choat JH, van Herwerden L. 2007. Extreme genetic diversity and temporal rather than spatial partitioning in a widely distributed coral reef fish. Mar Biol. 150(4):659–670. http://dx.doi.org/10.1007/s00227-006-0372-7

Knittweis L, Kraemer WE, Timm J, Kochzius M. 2009. Genetic structure of *Heliofungia actiniformis* (Scleractinia: Fungiidae) populations in the Indo-Malay Archipelago: implications for live coral trade management efforts. Conserv Genet. 10(1):241–249. http://dx.doi.org/10.1007/s10592-008-9566-5

Knowlton N. 2000. Molecular genetic analyses of species boundaries in the sea. Hydrobiologia. 420(1):73–90. http://dx.doi.org/10.1023/A:1003933603879

Kochzius M, Nuryanto A. 2008. Strong genetic population structure in the boring giant clam, *Tridacna crocea*, across the Indo-Malay Archipelago: implications related to evolutionary processes and connectivity. Mol Ecol. 17:3775–3787. PMid:18662232. http://dx.doi.org/10.1111/j.1365-294X.2008.03803.x

Kochzius M, Seidel C, Hauschild J, Kirchhoff S, Mester P, Meyer-Wachsmuth I, Nuryanto A, Timm J. 2009. Genetic population structures of the blue starfish *Linckia laevigata* and its gastropod ectoparasite *Thyca crystallina*. Mar Ecol Prog Ser. 396:211–219. http://dx.doi.org/10.3354/meps08281

Kon T, Nohara M, Nishida M, Sterrer W, Nishikawa T. 2006. Hidden ancient diversification in the circumtropical lancelet *Asymmetron lucayanum* complex. Mar Biol. 149(4):875–883. http://dx.doi.org/10.1007/s00227-006-0271-y

Lacson JM, Clark S. 1995. Genetic divergance of Maldivian and Micronesian demes of the damselfishes *Stegastes nigricans, Chriysiptera biocellata, C. glauca* and *C. leucopoma* (Pomacentridae). Mar Biol. 121(4):585–590. http://dx.doi.org/10.1007/BF00349293

Ladd HS. 1960. Origin of the Pacific island molluscan fauna. Am J Sci. 258A:137–150.

Lavery S, Moritz C, Fielder DR. 1995. Changing patterns of population structure and gene flow at different spatial scales in *Birgus latro* (the coconut crab) Heredity. 74:531–541. http://dx.doi.org/10.1038/hdy.1995.75

Lessios HA, Robertson DR. 2006. Crossing the impassable: genetic connections in 20 reef fishes across the eastern Pacific barrier. Proc Roy Soc B: Biol Sci. 273(1598):2201. PMid:16901840. PMCid:PMC1635520. http://dx.doi.org/10.1098/rspb.2006.3543

Lin TY, Liu LL. 2008. Low levels of genetic differentiation among populations of the coral inhabiting snail *Coralliophila violacea* (Gastropoda: Coralliophilidae) in regions of the Kuroshio and South China Sea. Zool Stud. 47(1):17–24.

Lind CE, Evans BS, Taylor JJU, Jerry DR. 2007. Population genetics of a marine bivalve, *Pinctada maxima*, throughout the Indo-Australian Archipelago shows differentiation and decreased diversity at range limits. Mol Ecol. 16(24):5193–5203. PMid:18028303. http://dx.doi.org/10.1111/j.1365-

294X.2007.03598.x

Liu S-YV, Kokita T, Dai C-F. 2008. Population genetic structure of the neon damselfish (*Pomacentrus coelestis*) in the northwestern Pacific Ocean. Mar Biol. 154(4):745–753. http://dx.doi.org/10.1007/s00227-008-0967-2

Liu Q, Li SJ, Li Z, Wang G. 2007. Species composition in genus *Scylla* from the coast of southeast China. J Fish China. 31:211–219.

Lord C, Lorion J, Dettai A, Watanabe S, Tsukamoto K, Cruaud C, Keith P. 2012. From endemism to widespread distribution: phylogeography of three amphidromous *Sicyopterus* species (Teleostei: Gobioidei: Sicydiinae). Mar Ecol Prog Ser. 455:269–285. http://dx.doi.org/10.3354/meps09617

Ludt WB, Bernal MA, Bowen BW, Rocha LA. 2012. Living in the past: phylogeography and population histories of Indo-Pacific wrasses (genus *Halichoeres*) in shallow lagoons versus outer reef slopes. PLoS ONE. 7(6):e38042. PMid:22701597. PMCid:PMC3368945. http://dx.doi.org/10.1371/journal.pone.0038042

Lukoschek V, Waycott M, Keogh JS. 2008. Relative information content of polymorphic microsatellites and mitochondrial DNA for inferring dispersal and population genetic structure in the olive sea snake, *Aipysurus laevis*. Mol Ecol. 17(13):3062–3077. PMid:18494766. http://dx.doi.org/10.1111/j.1365-294X.2008.03815.x

Lukoschek V, Waycott M, Marsh H. 2007. Phylogeography of the olive sea snake, *Aipysurus laevis* (Hydrophiinae) indicates Pleistocene range expansion around northern Australia but low contemporary gene flow. Mol Ecol. 16(16):3406–3422. PMid:17688542. http://dx.doi.org/10.1111/j.1365-294X.2007.03392.x

Macaranas JM, Ablan CA, Pante MJR, Benzie JAH, Williams ST. 1992. Genetic structure of giant clam (*Tridacna derasa*) populations from reefs in the Indo-Pacific. Mar Biol. 113(2):231–238.
Magsino RM, Juinio-Meñez MA. 2008. The influence of contrasting life history traits and oceanic processes on genetic structuring of rabbitfish populations *Siganus argenteus* and *Siganus fuscescens* along the eastern Philippine coasts. Mar Biol. 154(3):519–532. http://dx.doi.org/10.1007/s00227-008-0946-7

Mahidol C, Na-Nakorn U, Sukmanomon S, Taniguchi N, Nguyen TTT. 2007. Mitochondrial DNA diversity of the Asian moon scallop, *Amusium pleuronectes* (Pectinidae), in Thailand. Mar Biotechnol. 9(3):352–359. PMid:17549563. http://dx.doi.org/10.1007/s10126-006-6137-y

Mandal A, Rao D, Karuppaiah D, Gopalakrishnan A, Pozhoth J, Samraj YCT, Doyle RW. 2012. Population genetic structure of *Penaeus monodon*, in relation to monsoon current patterns in

southwest, east and Andaman coastal waters of India. Gene. 491(2):149–157. PMid:22020227. http://dx.doi.org/10.1016/j.gene.2011.10.002

Meyer CP. 2003. Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. Biol J Linnean Soc. 79(3):401–459. http://dx.doi.org/10.1046/j.1095-8312.2003.00197.x

Meyer CP, Geller JB, Paulay G. 2005. Fine scale endemism on coral reefs: archipelagic differentiation in turbinid gastropods. Evolution. 59(1):113–125. PMid:15792232.

Miller-Sims VC, Gerlach G, Kingsford MJ, Atema J. 2008. Dispersal in the spiny damselfish, *Acanthochromis polyacanthus*, a coral reef fish species without a larval pelagic stage. Mol Ecol. 17(23):5036–5048. PMid:19120989. http://dx.doi.org/10.1111/j.1365-294X.2008.03986.x

Mirams AGK, Treml EA, Shields JL, Liggins L, Riginos C. 2011. Vicariance and dispersal across an intermittent barrier: population genetic structure of marine animals across the Torres Strait land bridge. Coral Reefs. 30(4):937–949. http://dx.doi.org/10.1007/s00338-011-0767-x

Moritz C. 1994. Applications of mitochondrial DNA analysis in conservation: a critical review. Mol Ecol. 3(4):401–411. http://dx.doi.org/10.1111/j.1365-294X.1994.tb00080.x

Moritz C, Faith DP. 2002. Comparative phylogeography and the identification of genetically divergent areas for conservation. Mol Ecol. 7(4):419–429. http://dx.doi.org/10.1046/j.1365-294x.1998.00317.x

Muths D, Tessier E, Gouws G, Craig M, Mwale M, Mwaluma J, Mwandya A, Bourjea J. 2011. Restricted dispersal of the reef fish *Myripristis berndti* at the scale of the SW Indian Ocean. Mar Ecol Prog Ser. 443:167–180. http://dx.doi.org/10.3354/meps09394

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89(3):583-590.

Nuryanto A, Kochzius M. 2009. Highly restricted gene flow and deep evolutionary lineages in the giant clam *Tridacna maxima*. Coral Reefs. 28(3):607-619. http://dx.doi.org/10.1007/s00338-009-0483-y

Oksanen J, Guillaume Blanchet F, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H et al. 2012. Vegan: community ecology package. R package version 2.0-4.

Otwoma LM. 2012. Genetic population structure of the coral reef starfish *Linckia laevigata* in the Indo-West Pacific. MS thesis. Free University of Brussels. p. 73.

Ovenden JR, Lloyd J, Newman SJ, Keenan CP, Slater LS. 2002. Spatial genetic subdivision between northern Australian and southeast Asian populations of *Pristipomoides multidens*: a tropical marine reef fish species. Fish Res. 59(1–2):57–69. http://dx.doi.org/10.1016/S0165-7836(01)00415-5

Ovenden JR, Salini J, O'Connor S, Street R. 2004. Pronounced genetic population structure in a potentially vagile fish species (*Pristipomoides multidens*, Teleostei: Perciformes: Lutjanidae) from the East Indies triangle. Mol Ecol. 13(7):1991–1999. PMid:15189219. http://dx.doi.org/10.1111/j.1365-294X.2004.02210.x

Palumbi SR. 1994. Genetic divergence, reproductive isolation, and marine speciation. Ann Rev Ecol Syst. 25:547–572. http://dx.doi.org/10.1146/annurev.es.25.110194.002555

Palumbi SR. 1997. Molecular biogeography of the Pacific. Coral Reefs. 16:S47–S52. http://dx.doi.org/10.1007/s003380050241

Palumbi SR, Grabowsky G, Duda T, Geyer L, Tachino N. 1997. Speciation and population genetic structure in tropical Pacific sea urchins. Evolution. 51(5):1506–1517. http://dx.doi.org/10.2307/2411203

Paulay G, Meyer C. 2006. Dispersal and divergence across the greatest ocean region: Do larvae matter? Integrative Comp Biol. 46(3):269–281. PMid:21672741. http://dx.doi.org/10.1093/icb/icj027

Plaisance L, Rousset V, Morand S, Littlewood DTJ. 2008. Colonization of Pacific islands by parasites of low dispersal ability: phylogeography of two monogenean species parasitizing butterflyfishes in the South Pacific Ocean. J Biogeogr. 35(1):76–87.

Planes S, Fauvelot C. 2002. Isolation by distance and vicariance drive genetic structure of a coral reef fish in the Pacific Ocean. Evolution. 56(2):378–399. PMid:11926506.

Prakoon W, Tunkijjanukij S, Nguyen TTT, Na-Nakorn U. 2010. Spatial and temporal genetic variation of green mussel, *Perna viridis* in the Gulf of Thailand and implication for aquaculture. Mar Biotechnol. 12(5):506–515. PMid:19941027. http://dx.doi.org/10.1007/s10126-009-9234-x

Pruett CL, Winker K. 2008. The effects of sample size on population genetic diversity estimates in song sparrows *Melospiza melodia*. Journal of Avian Biology 39(2):252-256.

Puritz JB, Addison JA, Toonen RJ. 2012. Next-generation phylogeography: a targeted approach for multilocus sequencing of non-model organisms. PloS One. 7(3):e34241. PMid:22470543. PMCid:PMC3314618. http://dx.doi.org/10.1371/journal.pone.0034241

R Core Team. 2012. R: a language and environment for statistical computing. 2.15.1 ed. Vienna, Austria: R Foundation for Statistical Computing.

Ramon ML, Nelson PA, Martini E, Walsh WJ, Bernardi G. 2008. Phylogeography, historical demography, and the role of post-settlement ecology in two Hawaiian damselfish species. Mar Biol. 153(6):1207–1217. http://dx.doi.org/10.1007/s00227-007-0894-7

Ravago-Gotanco RG, Juinio-Menez MA. 2010. Phylogeography of the mottled spinefoot *Siganus fuscescens*: Pleistocene divergence and limited genetic connectivity across the Philippine archipelago. Mol Ecol. 19(20):4520–4534. PMid:20875063. http://dx.doi.org/10.1111/j.1365-294X.2010.04803.x

Reaka ML, Rodgers PJ, Kudla AU. 2008. Patterns of biodiversity and endemism on Indo- West Pacific coral reefs. Proc Natl Acad Sci USA. 105:11474–11481. PMid:18695226. PMCid:PMC2556415. http://dx.doi.org/10.1073/pnas.0802594105

Renema W, Bellwood D, Braga J, Bromfield K, Hall R, Johnson K, Lunt P, Meyer C, McMonagle L, Morley R. 2008. Hopping hotspots: global shifts in marine biodiversity. Science. 321(5889):654–657. PMid:18669854. http://dx.doi.org/10.1126/science.1155674

Rhodes KL, Lewis RI, Chapman RW, Sadovy Y. 2003. Genetic structure of camouflage grouper, *Epinephelus polyphekadion* (Pisces: Serranidae), in the western central Pacific. Mar Biol. 142(4):771–776.

Riginos C, Douglas KE, Jin Y, Shanahan DF, Treml EA. 2011. Effects of geography and life history traits on genetic differentiation in benthic marine fishes. Ecography. 34(4):566–575. http://dx.doi.org/10.1111/j.1600-0587.2010.06511.x

Roberts CM, McClean CJ, Veron JEN, Hawkins JP, Allen GR, McAllister DE, Mittermeier CG, Schueler FW, Spalding M, Wells F. 2002. Marine biodiversity hotspots and conservation priorities for tropical reefs. Science. 295(5558):1280. PMid:11847338. http://dx.doi.org/10.1126/science.1067728

Rocha L, Craig M, Bowen B. 2007. Phylogeography and the conservation of coral reef fishes. Coral Reefs. 26(3):501–512. http://dx.doi.org/10.1007/s00338-007-0261-7

Schultz JK, Feldheim KA, Gruber SH, Ashley MV, McGovern TM, Bowen BW. 2008. Global phylogeography and seascape genetics of the lemon sharks (genus *Negaprion*). Mol Ecol. 17(24):5336–5348. PMid:19121001. http://dx.doi.org/10.1111/j.1365-294X.2008.04000.x

Selkoe K, Toonen RJ. 2011. Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. Mar Ecol Prog Ser. 436:291–305. http://dx.doi.org/10.3354/meps09238

Selkoe KA, Watson J, White C, Ben-Horin T, Iacchei M, Miterai S, Siegel D, Gaines SD, Toonen RJ. 2010. Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. Mol Ecol. 19(17):3708–3726. PMid:20723063. http://dx.doi.org/10.1111/j.1365-294X.2010.04658.x

Shearer TL, Van Oppen MJ, Romano SL, Worheide G. 2002. Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). Mol Ecol. 11:2475–2487. PMid:12453233. http://dx.doi.org/10.1046/j.1365-294X.2002.01652.x

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595. PMid:2513255. PMCid:PMC1203831.

Tenopir C, Allard S, Douglass KL, Aydinoglu AU, Wu L, Read E, Manoff M, Frame M. 2011. Data sharing by scientists: practices and perceptions. PloS One. 6(6):e21101. PMid:21738610. PMCid:PMC3126798. http://dx.doi.org/10.1371/journal.pone.0021101

Teske PR, Hamilton H, Palsboll PJ, Choo CK, Gabr H, Lourie SA, Santos M, Sreepada A, Cherry MI, Matthee CA. 2005. Molecular evidence for long-distance colonization in an Indo-Pacific seahorse lineage. Mar Ecol Prog Ser. 286:249–260. http://dx.doi.org/10.3354/meps286249

Thacker CE, Thompson AR, Roje DM, Shaw EY. 2008. New expansions in old clades: population genetics and phylogeny of *Gnatholepis* species (Teleostei: Gobioidei) in the Pacific. Mar Biol. 153(3):375–385. http://dx.doi.org/10.1007/s00227-007-0814-x

Timm J, Kochzius M. 2008. Geological history and oceanography of the Indo-Malay Archipelago shape the genetic population structure in the false clown anemonefish (*Amphiprion ocellaris*). Mol Ecol. 17(18):3999–4014. PMid:19238702. http://dx.doi.org/10.1111/j.1365-294X.2008.03881.x

Timm J, Planes S, Kochzius M. 2012. High similarity of genetic population structure in the false clown anemonefish (*Amphiprion ocellaris*) found in microsatellite and mitochondrial control region analysis. Conserv Genet. 13(3):693–706. http://dx.doi.org/10.1007/s10592-012-0318-1

Timmers MA, Andrews KR, Bird CE, deMaintenton MJ, Brainard RE, Toonen RJ. 2011. Widespread dispersal of the crown-of-thorns sea star, *Acanthaster planci*, across the Hawaiian Archipelago and Johnston Atoll. J Mar Biol. 2011:Article ID 934269, 10 p.

Timmers MA, Bird CE, Skillings DJ, Smouse PE, Toonen RJ. 2012. There's no place like home: crown-of-thorns outbreaks in the central Pacific are regionally derived and independent events. PLoS ONE. 7(2):e31159. PMid:22363570. PMCid:PMC3281911. http://dx.doi.org/10.1371/journal pone.0031159

Tittensor DP, Mora C, Jetz W, Lotze HK, Ricard D, Berghe EV, Worm B. 2010. Global patterns and predictors of marine biodiversity across taxa. Nature. 466(7310):1098–1101. PMid:20668450. http://dx.doi.org/10.1038/nature09329

Toonen RJ. 1997. Microsatellites for ecologists: non-radioactive isolation and amplification protocols for microsatellite markers. Unpublished manuscript available from the author online: http://www.academia.edu/950049/Microsatellites_for_ecologists_non-radioactive_isolation_and_amplification_protocols_for_microsatellite_markers

Toonen RJ, Andrews KR, Baums IB, Bird CE, Concepcion GT, Daly-Engel TS, Eble JA, Faucci A, Gaither MR, Iacchei M. 2011. Defining boundaries for ecosystem-based management: a multispecies case study of marine connectivity across the Hawaiian Archipelago. J Mar Biol. 2011:Article ID 460173, 13 p.

Tsang LM, Chan BKK, Wu TH, Ng WC, Chatterjee T, Williams GA, Chu KH. 2008. Population differentiation in the barnacle *Chthamalus malayensis*: postglacial colonization and recent connectivity across the Pacific and Indian oceans. Mar Ecol Prog Ser. 364:107–118. http://dx.doi.org/10.3354/meps07476

Tsoi KH, Chan TY, Chu KH. 2007. Molecular population structure of the kuruma shrimp *Penaeus japonicus* species complex in western Pacific. Mar Biol. 150(6):1345–1364. http://dx.doi.org/10.1007/s00227-006-0426-x

Uehara T, Shingaki M. 1985. Taxonomic studies on the four types of sea urchin, *Echinometra mathaei* from Okinawa, Japan. Zoo Sci. 2:1009–1015.

Uthicke S, Benzie JAH. 2001. Restricted gene flow between *Holothuria scabra* (Echinodermata: Holothuroidea) populations along the north-east coast of Australia and the Solomon Islands. Mar Ecol Prog Ser. 216:109–117. http://dx.doi.org/10.3354/meps216109

Uthicke S, Benzie JAH. 2003. Gene flow and population history in high dispersal marine invertebrates: mitochondrial DNA analysis of *Holothuria nobilis* (Echinodermata: Holothuroidea) populations from the Indo-Pacific. Mol Ecol. 12(10):2635–2648. PMid:12969467. http://dx.doi.org/10.1046/j.1365-294X.2003.01954.x

Uthicke S, Benzie JAH, Ballment E. 1999. Population genetics of the fissiparous holothurian *Stichopus chloronotus* (Aspidochirotida) on the Great Barrier Reef, Australia. Coral Reefs. 18(2):123–132. http://dx.doi.org/10.1007/s003380050166

Uthicke S, Conand C, Benzie JAH. 2001. Population genetics of the fissiparous holothurians *Stichopus chloronotus* and *Holothuria atra* (Aspidochirotida): a comparison between the Torres Strait and La

Reunion. Mar Biol. 139(2):257–265. http://dx.doi.org/10.1007/s002270100579

Uthicke S, Purcell S. 2004. Preservation of genetic diversity in restocking of the sea cucumber *Holothuria scabra* investigated by allozyme electrophoresis. Can J Fish Aquat Sci. 61(4):519–528. http://dx.doi.org/10.1139/f04-013

Vagelli A, Burford M, Bernardi G. 2008. Fine scale dispersal in Banggai Cardinalfish, *Pterapogon kaudemi*, a coral reef species lacking a pelagic larval phase. Mar Genom. 1(3–4):129–134. PMid:21798164. http://dx.doi.org/10.1016/j.margen.2009.01.001

van Herwerden L, Aspden WJ, Newman SJ, Pegg GG, Briskey L, Sinclair W. 2009a. A comparison of the population genetics of *Lethrinus miniatus* and *Lutjanus sebae* from the east and west coasts of Australia: evidence for panmixia and isolation. Fish Res. 100(2):148–155. http://dx.doi.org/10.1016/j.fishres.2009.07.003

van Herwerden L, Benzie J, Davies C. 2003. Microsatellite variation and population genetic structure of the red throat emperor on the Great Barrier Reef. J Fish Biol. 62(5):987–999. http://dx.doi.org/10.1046/j.1095-8649.2003.00075.x

van Herwerden L, Choat JH, Newman SJ, Leray M, Hillersoy G. 2009b. Complex patterns of population structure and recruitment of *Plectropomus leopardus* (Pisces: Epinephelidae) in the Indo-West Pacific: implications for fisheries management. Mar Biol. 156(8):1595–1607. http://dx.doi.org/10.1007/s00227-009-1195-0

Vargas S, Schuster A, Sacher K, Büttner G, Schätzle S, Läuchli B, Hall K, Hooper JNA, Erpenbeck D, Wörheide G. 2012. Barcoding sponges: an overview based on comprehensive sampling. PloS One. 7(7):e39345. PMid:22802937. PMCid:PMC3389008. http://dx.doi.org/10.1371/journal.pone.0039345

Veron JEN. 1995. Corals in space and time. The biogeography and evolution of the scleractinia. Ithaca, NY, USA: Cornell University Press (Comstock).

Visram S, Yang M-C, Pillay RM, Said S, Henriksson O, Grahn M, Chen CA. 2010. Genetic connectivity and historical demography of the blue barred parrotfish (*Scarus ghobban*) in the western Indian Ocean. Mar Biol. 157(7):1475–1487. http://dx.doi.org/10.1007/s00227-010-1422-8

Vogler C, Benzie J, Lessios H, Barber P, Wörheide G. 2008. A threat to coral reefs multiplied? Four species of crown-of-thorns starfish. Biol Lett. 4(6):696–699. PMid:18832058. PMCid:PMC2614177. http://dx.doi.org/10.1098/rsbl.2008.0454

von der Heyden S, Beger M, Toonen RJ, van Herwerden L, Juinio-Meñez MA, Ravago-Gotanco R,

Fauvelot C, Bernardi G. 2014. The application of genetics to marine management and conservation: examples from the Indo-Pacific. Bull Mar Sci 90(1):123-158.

Wares JP. 2009. Natural distributions of mitochondrial sequence diversity support new null hypotheses. Evolution. 64(4):1136–1142. PMid:19863588. http://dx.doi.org/10.1111/j.1558-5646.2009.00870.x

Weersing K, Toonen RJ. 2009. Population genetics, larval dispersal, and connectivity in marine systems. Mar Ecol Prog Ser. 393(1):12.

Wessel P, Smith WHF. 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. J Geophys Res. 101(B4):8741–8743. http://dx.doi.org/10.1029/96JB00104

Whitlock MC. 2011. Data archiving in ecology and evolution: best practices. Trends Ecol Evol. 26(2):61–65. PMid:21159406. http://dx.doi.org/10.1016/j.tree.2010.11.006

Williams ST, Benzie JAH. 1993. Genetic consequences of long larval life in the starfish *Linckia laevigata* (Echinodermata, Asteroidea) on the Great Barrier Reef. Mar Biol. 117(1):71–77. http://dx.doi.org/10.1007/BF00346427

Williams ST, Benzie JAH. 1996. Genetic uniformity of widely separated populations of the coral reef starfish *Linckia laevigata* from the east Indian and west Pacific oceans, revealed by allozyme electrophoresis. Mar Biol. 126(1):99–107. http://dx.doi.org/10.1007/BF00571381

Winters KL, van Herwerden L, Choat JH, Robertson DR. 2010. Phylogeography of the Indo-Pacific parrotfish *Scarus psittacus*: isolation generates distinctive peripheral populations in two oceans. Mar Biol. 157(8):1679–1691. http://dx.doi.org/10.1007/s00227-010-1442-4

WoRMS Editorial Board. 2012. World Register of Marine Species. Available from: http://www.marinespecies.org. Accessed 15 December, 2012.

Xia JH, Huang JH, Gong JB, Jiang SG. 2008. Significant population genetic structure of yellowfin seabream *Acanthopagrus latus* in China. J Fish Biol. 73(8):1979–1992. http://dx.doi.org/10.1111/j.1095-8649.2008.02010.x

Yap CK, Chong CM, Tan SG. 2011. Allozyme polymorphisms in horseshoe crabs, *Carcinoscorpius rotundicauda*, collected from polluted and unpolluted intertidal areas in Peninsular Malaysia. Environ Monit Assess. 174(1–4):389–400. PMid:20437264. http://dx.doi.org/10.1007/s10661-010-1464-x

Yap CK, Tan SG, Ismail A, Omar H. 2004. Allozyme polymorphisms and heavy metal levels in the green-lipped mussel *Perna viridis* (Linnaeus) collected from contaminated and uncontaminated sites in Malaysia. Environ Int. 30(1):39–46. http://dx.doi.org/10.1016/S0160-4120(03)00144-2

Yasuda N, Nagai S, Hamaguchi M, Okaji K, Gerard K, Nadaoka K. 2009. Gene flow of *Acanthaster planci* (L.) in relation to ocean currents revealed by microsatellite analysis. Mol Ecol. 18(8):1574–1590. PMid:19302361. http://dx.doi.org/10.1111/j.1365-294X.2009.04133.x

Yue GH, Zhu ZY, Lo LC, Wang CM, Lin G, Fenf F, Pang HY, Li J, Gong P, Liu HM, et al. 2009. Genetic variation and population structure of Asian seabass (*Lates calcarifer*) in the Asia-Pacific region. Aquaculture. 293(1–2):22–28. http://dx.doi.org/10.1016/j.aquaculture.2009.03.053

Zhang JB, Cai ZP, Huang LM. 2006. Population genetic structure of crimson snapper *Lutjanus erythropterus* in East Asia, revealed by analysis of the mitochondrial control region. Ices J Mar Sci. 63(4):693–704. http://dx.doi.org/10.1016/j.icesjms.2006.01.00

Appendix 1. All species included in the dataset, with their citations and the genetic marker used in the study. Names in black are those species for which species authority was derived from the Integrated Taxonomic Information System, [#] represents those derived from the World Register of Marine Species and [‡] represents those derived from Fishbase.

Species name and authority	Citations for studies included in this data set	Marker
Acanthaster planci (Linnaeus, 1758)	Benzie 1999, Yasuda et al. 2009	Allozymes, Microsatellites
Acanthochromis polyacanthus (Bleeker, 1855)	Doherty et al. 1995, Miller-Sims et al. 2008	Allozymes, Microsatellites
Acanthopagrus latus (Houttuyn, 1782)	Xia et al. 2008	MtDNA
Acanthurus nigrofuscus (Forsskål, 1775)	Eble et al. 2011a	MtDNA
Acanthurus nigroris Valenciennes in Cuvier and Valenciennes, 1835	DiBattista et al. 2011	MtDNA
Acanthurus triostegus (Linnaeus, 1758)	Mirams et al. 2011, Planes and Fauvelot 2002	Allozymes, MtDNA
Aipysurus laevis Lacépède, 1804	Lukoschek et al. 2008, Lukoschek et al. 2007	Microsatellites, MtDNA
Amphiprion melanopus Bleeker, 1852	Doherty et al. 1995, Drew et al. 2008	Allozymes, MtDNA
Amphiprion ocellaris Cuvier in Cuvier and Valenciennes, 1830	Timm and Kochzius 2008, Timm et al. 2012	MtDNA, Microsatellites
Amusium pleuronectes (Linnaeus, 1758) [#]	Mahidol et al. 2007	MtDNA
Apogon doederleini Jordan and Snyder, 1901	Mirams et al. 2011	MtDNA
Asymmetron lucayanum Andrews, 1893	Kon et al. 2006	MtDNA
Aulostomus chinensis (Linnaeus, 1766)	Bowen et al. 2001	MtDNA
Birgus latro (Linnaeus, 1767)[#]	Lavery et al. 1995	Allozymes
Carcinoscorpius rotundicauda (Latreille, 1802)[#]	Yap et al. 2011	Allozymes
Carijoa riisei (Duchassaing and Michelotti, 1860)	Concepcion et al. 2010	MtDNA, other nuclear markers
Centropyge flavissima (Cuvier in Cuvier and Valenciennes, 1831)	DiBattista et al. 2012	MtDNA
Cephalopholis argus Bloch and Schneider, 1801	Gaither et al. 2011a	MtDNA, other nuclear markers
Chaetodon citrinellus Cuvier in Cuvier and Valenciennes, 1831	Fauvelot and Planes 2002	Allozymes
Chaetodon quadrimaculatus Gray, 1831	Fauvelot and Planes 2002	Allozymes
Chelonia mydas (Linnaeus, 1758)	Bourjea et al. 2007, Dethmers et al. 2006	MtDNA
Chlorurus sordidus (Forsskål, 1775)	Bay et al. 2004, Dudgeon et al. 2000	MtDNA
Chromis atripectoralis Welander and Schultz, 1951	Doherty et al. 1995	Allozymes
Chromis xanthura (Bleeker, 1854)	Fauvelot and Planes 2002	Allozymes
Chrysiptera glauca (Cuvier in Cuvier and Valenciennes, 1830)	Fauvelot and Planes 2002	Allozymes
Chrysiptera leucopoma (Cuvier in Cuvier and Valenciennes, 1830)	Lacson and Clark 1995	Allozymes

Appendix 1. Continued.

Species name and authority	Citations for studies included in this data set	Marker
Chrysiptera talboti (Allen, 1975)	Drew et al. 2008	MtDNA
Chthamalus malayensis Pilsbry, 1916[#]	Tsang et al. 2008	MtDNA
Clavularia koellikeri (Dean, 1927)[#]	Bastidas 2002	Allozymes
Conus chaldaeus (Röding, 1798)[#]	Duda et al. 2012	MtDNA
Conus ebraeus Linnaeus, 1758[#]	Duda and Lessios 2009	MtDNA
Conus miliaris Hwass in Bruguière, 1792[#]	Duda and Lee 2009	MtDNA
Conus sanguinolentus Quoy and Gaimard, 1834 [#]	Duda et al. 2012	MtDNA
Coralliophila violacea (Kiener, 1836)[#]	Lin and Liu 2008	Allozymes
Craterocephalus capreoli (Rendahl, 1922)	Johnson et al. 1994	Allozymes
Dascyllus albisella Gill, 1862	Ramon et al. 2008	MtDNA
Dascyllus aruanus (Linnaeus, 1758)	Fauvelot and Planes 2002	Allozymes
Dascyllus trimaculatus (Rüppell, 1829)	Mirams et al. 2011	MtDNA
Echinometra mathaei (Blainville, 1825)[#]	Palumbi et al. 1997	MtDNA
Echinometra oblonga (Blainville, 1825)[#]	Palumbi et al. 1997	MtDNA
Echinometra sp. C. [first mentioned in Uehara and Shingaki 1985]	Palumbi et al. 1997	MtDNA
Eleutheronema tetradactylum (Shaw, 1804)	Horne et al. 2011	MtDNA, Microsatellites
Epinephelus coioides (Hamilton, 1822)	Antoro et al. 2006	Microsatellites
Epinephelus polyphekadion (Bleeker, 1849)	Rhodes et al. 2003	Microsatellites
Euryhaliotrematoides grandis (Mizelle and Kritsky, 1969) [# revised to *Euryhaliotrema grandis* Kritsky 2012]	Plaisance et al. 2008	MtDNA
Fenneropenaeus indicus (H.Milne-Edwards, 1837)	de Croos and Palsson 2010	MtDNA
Forcipiger flavissimus Jordan and McGregor in Jordan and Evermann, 1898	Fauvelot and Planes 2002	Allozymes
Gnatholepis anjerensis (Bleeker, 1851)	Thacker et al. 2008	Microsatellites
Gnatholepis scapulostigma Herre, 1953[# synonym of *G. cauerensis* (Bleeker, 1853)]	Thacker et al. 2008	Microsatellites
Halichoeres claudia Randall & Rocha, 2009[#]	Ludt et al. 2012	MtDNA
Halichoeres ornatissimus (Garrett, 1863)	Ludt et al. 2012	MtDNA
Halichoeres trimaculatus (Quoy and Gaimard, 1834)	Ludt et al. 2012	MtDNA

Appendix 1. Continued.

Species name and authority	Citations for studies included in this data set	Marker
Haliotis asinina Linnaeus, 1758	Imron et al. 2007	MtDNA
Haptosquilla glyptocercus (Wood-Mason, 1875)[#]	Barber et al. 2002	MtDNA
Haptosquilla pulchella (Miers, 1880)[#]	Barber et al. 2002	MtDNA
Heliofungia actiniformis (Quoy and Gaimard, 1833)	Knittweis et al. 2009	Other Nuclear Markers
Hippocampus kuda Bleeker, 1852	Teske et al. 2005	MtDNA
Holothuria (Halodeima) atra Jaeger, 1833	Uthicke et al. 2001	Allozymes
Holothuria (Microthele) nobilis (Selenka, 1867)	Uthicke and Benzie 2003	MtDNA
Holothuria (Metriatyla) scabra Jaeger, 1833	Uthicke and Benzie 2001, Uthicke and Purcell 2004	Allozymes
Lates calcarifer (Bloch, 1790)	Chenoweth et al. 1998, Yue et al. 2009	MtDNA, Microsatellites
Lethrinus miniatus (Forster in Bloch and Schneider, 1801)	van Herwerden et al. 2009a, van Herwerden et al. 2003	Microsatellites, MtDNA
Linckia laevigata (Linnaeus, 1758)[#]	Crandall et al. 2008b, Kochzius et al. 2009, Williams and Benzie 1993, 1996	MtDNA, Allozymes
Lutjanus erythropterus Bloch, 1790	Zhang et al. 2006	MtDNA
Lutjanus fulvus (Forster in Bloch and Schneider, 1801)	Gaither et al. 2010	MtDNA
Lutjanus kasmira (Forsskål, 1775)	Gaither et al. 2010	MtDNA
Lutjanus sebae (Cuvier, 1816)	van Herwerden et al. 2009a	MtDNA
Penaeus japonicus (Bate, 1888) synonym of *Marsupenaeus japonicus*	Tsoi et al. 2007	Microsatellites
Mugil cephalus Linnaeus, 1758	Livi et al. 2011	MtDNA
Myripristis berndti Jordan and Evermann, 1903	Craig et al. 2007, Muths et al. 2011	MtDNA, Microsatellites
Naso brevirostris (Cuvier, 1829)	Horne et al. 2008	MtDNA
Naso vlamingii (Valenciennes in Cuvier and Valenciennes, 1835)	Klanten et al. 2007	MtDNA
Negaprion acutidens (Rüppell, 1837)	Schultz et al. 2008	MtDNA, Microsatellites
Nerita albicilla Linnaeus, 1758[#]	Crandall et al. 2008a	MtDNA
Nerita plicata Linnaeus, 1758[#]	Crandall et al. 2008a	MtDNA
Penaeus monodon Fabricius, 1798	Duda and Palumbi 1999, Mandal et al. 2012	Other nuclear markers, Microsatellites
Periclimenes soror Nobili, 1904	Crandall et al. 2008b	MtDNA
Perna viridis (Linnaeus, 1758)[#]	Prakoon et al. 2010, Yap et al. 2004	Microsatellites, Allozymes

Appendix 1. Continued.

Species name and authority	Citations for studies included in this data set	Marker
Phycomenes zostericola Bruce, 2008[#]	Haig et al. 2010	MtDNA
Pinctada maxima (Jameson, 1901)[#]	Lind et al. 2007	Microsatellites
Plectorhinchus flavomaculatus (Cuvier in Cuvier and Valenciennes, 1830)	Han et al. 2008	MtDNA
Plectroglyphidodon dickii (Liénard, 1839)	Fauvelot and Planes 2002	Allozymes
Plectropomus leopardus (Lacépède, 1802)	Van Herwerden et al. 2009b	MtDNA
Polynemus sheridani MacLeay, 1884[#] synonym of *Polydactylus macrochir* (Günther, 1867)	Chenoweth and Hughes 2003	MtDNA
Pomacentrus coelestis Jordan and Starks, 1901	Liu et al. 2008, Mirams et al. 2011	MtDNA
Pomacentrus moluccensis Bleeker, 1853	Drew et al. 2008	MtDNA
Pomacentrus pavo (Bloch, 1787)	10.1007/s00227-002-0869-7	Allozymes
Pristipomoides filamentosus (Valenciennes in Cuvier and Valenciennes, 1830)	Gaither et al. 2011b	MtDNA, Microsatellites
Pristipomoides multidens (Day, 1871)	Ovenden et al. 2002, Ovenden et al. 2004	MtDNA
Protoreaster nodosus (Linnaeus, 1758)[#]	Crandall et al. 2008b	MtDNA
Pterapogon kauderni Koumans, 1933	Bernardi and Vagelli 2004, Hoffman et al. 2005, Vagelli et al. 2009	MtDNA, Microsatellites
Pterocaesio chrysozona (Cuvier in Cuvier and Valenciennes, 1830)	Doherty et al. 1995	Allozymes
Scarus frenatus Lacepède, 1802	Dudgeon et al. 2000	MtDNA
Scarus ghobban Forsskål, 1775	Visram et al. 2010	MtDNA
Scarus psittacus Forsskål, 1775	Winters et al. 2010	MtDNA
Scarus rubroviolaceus Bleeker, 1847	Fitzpatrick et al. 2011	Microsatellites
Scylla serrata (Forsskål, 1775)	Fratini and Vannini 2002, Gopurenko and Hughes 2002, Gopurenko 1999	MtDNA
Sicyopterus aiensis Keith, Watson & Marquet, 2004[‡]	Lord et al. 2012	MtDNA
Sicyopterus lagocephalus (Pallas, 1770)	Hoareau et al. 2012, Lord et al. 2012	MtDNA
Sicyopterus sarasini Weber and de Beaufort, 1915	Lord et al. 2012	MtDNA
Siganus argenteus (Quoy and Gaimard, 1825)	Magsino and Juinio-Meñez 2008	Allozymes
Siganus fuscescens (Houttuyn, 1782)	Magsino and Juinio-Meñez 2008, Ravago-Gotanco and Juinio-Menez 2010	Allozymes, MtDNA

Appendix 1. Continued.

Species name and authority	Citations for studies included in this data set	Marker
Siganus guttatus (Bloch, 1787)	Iwamoto et al. 2009	MtDNA
Sinularia flexibilis (Quoy and Gaimard, 1833)[#]	Bastidas et al. 2001	Allozymes
Sphyraena barracuda (Edwards in Catesby, 1771)	Daly-Engel et al. 2012	MtDNA
Stegastes fasciolatus (Ogilby, 1889)	Ramon et al. 2008	MtDNA
Stegastes nigricans (Lacépède, 1802)	Doherty et al. 1995	Allozymes
Stegostoma fasciatum (Hermann, 1783)	Dudgeon et al. 2009	MtDNA, Microsatellites
Stichopus chloronotus Brandt, 1835[#]	Uthicke et al. 1999, Uthicke et al. 2001	Allozymes
Thyca crystallina (Gould, 1846)[#]	Crandall et al. 2008b, Kochzius et al. 2009	MtDNA
Tridacna crocea Lamarck, 1819[#]	DeBoer et al. 2008, Kochzius and Nuryanto 2008	MtDNA
Tridacna derasa (Röding, 1798)[#]	Macaranas et al. 1992	Allozymes
Tridacna gigas (Linnaeus, 1758)[#]	Benzie and Williams 1995	Allozymes
Tridacna maxima (Röding, 1798)[#]	Benzie and Williams 1997, Nuryanto and Kochzius 2009	Allozymes, MtDNA
Zebrasoma flavescens (Bennett, 1828)	Eble et al. 2011b	MtDNA, Microsatellites

CHAPTER 3
Multi-species approach to testing the species-genetic diversity correlation in Indo-Pacific reef animals.

ABSTRACT

Both community ecology and population genetics seek to uncover general principles describing the effects of dispersal, extinction, and diversification on spatial patterns of biodiversity. Community ecology focuses on species diversity whereas population genetics focuses on intraspecific genetic diversity; but, because landscape context affects dispersal and extinction likelihoods (of species or alleles) in a similar manner, species and genetic diversity should show similar spatial patterns. Here we test this expectation (the Species-Genetic Diversity Correlation: SGDC) using published genetic diversity data for 75 marine species from the Indo-Pacific region. We present a novel Bayesian approach to these analyses using summary statistics of within-species genetic diversity (H_E from microsatellites, h and π from mtDNA) as the response variables. Our results suggest an overall positive correlation between species richness and genetic diversity across Indo-Pacific reefs (SGDC slopes greater than zero for each of the three genetic diversity indices). There are, however, high levels of variation among species in the direction and strength of the SGDC. These results are consistent with the spatial configuration of Indo-Pacific reefs affecting both inter- and intraspecific diversity in a concordant way, but highlight the need to account for inter-specific variation.

INTRODUCTION

The key processes driving the spatial distribution of both species diversity and neutral genetic diversity are strongly analogous: dispersal, extinction and diversification (either speciation or mutation) mediate patterns of both species and neutral genetic diversity. Given the similarity between the factors influencing species and genetic diversity, it is reasonable to expect spatial patterns in species and genetic diversity to be correlated (i.e., *alpha* diversity of both species and alleles). This pattern was first noted by Antonovics (1976) and was formalised by Vellend (2003) as the Species-Genetic Diversity Correlation (SGDC). The SGDC draws together ideas from the fields of community ecology and population genetics, but has applicability for fields such as conservation biology (Kahilainen et al. 2014) and evolutionary ecology (Laroche et al. 2015). Knowing where diversity accumulates across multiple levels of biological organisation can identify areas useful for studying the processes of biodiversity creation. Furthermore, this knowledge can aid conservation decision-making processes, particularly where decisions are based on species diversity rather than genetic diversity (Beger et al. 2014).

The SGDC can be conceptualized by considering the parallels between the Theory of Island Biogeography (MacArthur and Wilson 1967) and the Island Model of Population Genetics (Wright

1940). The Theory of Island Biogeography is a central theory for understanding species diversity and predicts the effects of area and isolation on the number of species in a habitat patch through the processes of immigration and extinction. The Island Model of Population Genetics, similarly, considers the effects of population size and isolation on the diversity of alleles in a population through the processes of gene flow and genetic drift. In ecological terms, the similarity between these two theories can be explained as follows: large habitat patches that are well connected to other patches are likely to support more species and more genetic diversity than smaller, more isolated patches. These two neutral theories downplay the role of selection in determining patterns of diversity, focusing instead on the processes of extinction and dispersal. Clearly species vary in their responses to the environment, so selection at the species level should increase among-species variance (Endler 1986; Via 2001; Rundle and Nosil 2005; Bird et al. 2012) but may be unlikely to show general concordance across taxa. Thus, if dispersal and extinction are the predominant processes underlying the spatial patterns of species and genetic diversity, then geography should influence these processes similarly and lead to a positive correlation between species and genetic diversity. Maximising the number of species considered, as we do here, greatly increases the chances of seeing past any species-specific effects of selection to the general concordance driven by the neutral processes of dispersal and extinction.

In addition to the role of geography (i.e., environmental characteristics and habitat isolation) in creating similar spatial patterns of both species and genetic diversity, there may be direct feedbacks between these two levels of diversity in which one causes change in the other (Whittaker 1972; Vellend and Geber 2005). This is easiest to understand when diversity is measured within organisms forming a habitat (such as forest trees) and the communities they support. Speciose or genetically diverse habitats may create more niches for different species or genotypes to occupy than less diverse habitats. For example, a recent meta-analysis showed positive effects of non-neutral genetic diversity within focal plant species on the diversity of their associated communities (Whitlock 2014). Although theory suggests that causation can work the other way (Vellend and Geber 2005), an effect of the species diversity of habitat-forming organisms on the genetic diversity of one or a few resident species, this is not commonly studied (for an exception, see Robinson et al. 2010).

Modelling of terrestrial plant communities (Vellend 2005) reinforced the expectation that species and genetic diversity positively correlate when driven predominantly by the effects of habitat area and isolation. Vellend (2005) found that simulations dominated by area or immigration resulted in more predictably positive correlations between species and genetic diversity than did those in which environmental heterogeneity was the dominant force. This emphasizes the dependence of the SGDC on neutral dynamics and, along with the same study’s finding that species abundance influenced the strength of the correlation in simulations, identifies the strong effect that the choice of focal species (within which to measure genetic diversity) can have on the strength of the SGDC. Variation among species in abundance and responses to heterogeneity argues for maximising the number of species in such analyses.

Empirical tests of the SGDC have been strongly terrestrially focused and have typically measured neutral genetic diversity within a single or few species (the highest number of species = 12, Papadopoulou et al. 2011). For example, positive SGDCs have been reported from communities of forest herbs (Vellend 2004), butterflies (Cleary et al. 2006), bats (Struebig et al. 2011), trees (Wehenkel et al. 2006; Wei and Jiang 2012), stream fishes (Blum et al. 2012), freshwater invertebrates (Evanno et al. 2009; Finn and Poff 2011; Lamy et al. 2013), sand beetles (Papadopoulou et al. 2011), salt marsh communities (Robinson et al. 2010) and coral reef fishes (Messmer et al. 2012). However, several of these studies also report negative or non-significant correlations, depending on the focal species targeted for genetic diversity (Robinson et al. 2010; Papadopoulou et al. 2011; Struebig et al. 2011), the choice of diversity metric calculated (Wehenkel et al. 2006; Evanno et al. 2009; Blum et al. 2012) or the demographic history of the community (Wei and Jiang 2012).

Although many of these studies individually show support the SGDC, they contain a number of shortcomings such that determining the generality of the SGDC in ecology is not yet possible. First, few studies of the SGDC have measured genetic diversity across many species, resulting in poor explanatory power to assess trends across species (Quinn and Keough 2002). Second, several studies report results for analyses of genetic diversity summed or averaged across species (Robinson et al. 2010; Papadopoulou et al. 2011; Messmer et al. 2012), thereby ignoring variation among species in mutation rate, evolutionary history, life history or demography. Ideally, species should be treated as replicates which, if they show general concordance, can allow us to infer shared process (Avise et al. 1987; Dawson 2012). One robust approach for evaluating the generality of the SGDC, meta-analysis of the effect sizes from previous studies, supports the existence of a positive SGDC (Vellend et al. 2014; Whitlock 2014). Meta-analyses, although increasing analytical power by increasing the number of species considered, do not solve the flaws in experimental design of their constituent studies, as identified above. Third, whether the SGDC holds true for marine ecosystems is largely unknown. There are, to our knowledge, only four examples of explicit considerations of the SGDC in the ocean (Reusch et al. 2005; Robinson et al. 2010; Noyer and Becerro 2011; Messmer et al. 2012) providing mixed support for the SGDC. Three of these studies found no positive correlation between genetic and species diversity in systems of habitat-forming organisms and their associated communities (Reusch et al. 2005; Robinson et al. 2010; Noyer and Becerro 2012). The single study correlating species and genetic diversity within the same trophic group (coral reef fishes) found positive, zero and negative slopes depending on the species used for genetic analysis (Messmer et al. 2012).

Here we take a multi-species approach to testing the SGDC hypothesis, using genetic data from 75 shallow-water marine species and focusing on the species-rich coral reefs of the Indian and Pacific Oceans. These ecosystems offer an excellent system in which to study the SGDC in that they form discrete habitat patches in an inhospitable matrix of deep-water soft sediments, the landscape configuration most likely to yield concordant patterns between species and genetic diversity (Vellend 2014). Much work has been done to examine the underlying cause of the gradient

of species diversity in this region (Connolly et al. 2003; Renema et al. 2008; Bellwood et al. 2012), where species richness is highest at the confluence of the Indian and Pacific Oceans. Studies considering intra-specific genetic structure in this region are numerous (for reviews, see Carpenter et al. 2011; Keyse et al. 2014), but rarely encompass more than a handful of species, so seeking common patterns of genetic diversity is difficult. With such a strong gradient of species diversity, a network of habitat islands of varying size and isolation, and a substantial repository of genetic data, the Indo-Pacific is an excellent system within which to test the generality of the SGDC both spatially and taxonomically.

Our broad-scale data synthesis approach to testing the SGDC hypothesis uses genetic diversity data for 75 species of reef animals across 277 reef sites in the Indo-Pacific and species richness data for six families of tropical marine vertebrates (see Supp. Table 1). We predict that, as genetic and species diversity should respond similarly to the landscape configurations of coral reef habitat, so neutral genetic diversity within reef species will be positively correlated with the species richness of their community. We take a Bayesian modelling approach to *beta* regression of genetic diversity as a response variable while taking into account the effect of using genetic diversity data from multiple species and genetic markers on the slope of the SGDC. This is the first time that the SGDC has been assessed across such a broad spatial and taxonomic extent and the first time that variation among species has been explicitly factored into such analyses.

METHODS

Genetic diversity data

A literature search was conducted in Web of Science on Sept 21, 2013, to capture studies reporting genetic diversity summary statistics for marine animals of the Indo-Pacific. The search terms below were chosen to maximise inclusion of articles containing georeferenced data on population genetic diversity from shallow water marine habitats in the Indo-Pacific region. The following sets of Boolean search terms were combined in the Advanced Search tool: (gene flow OR population structure OR genetic diversity OR phylogeograph* OR F statistic OR mtDNA OR microsatellite OR allozyme) AND (marine OR coral OR reef OR intertidal OR subtidal OR estuar*) AND (Indo* OR Malay* OR Indo-Pacific OR Coral Triangle).

The resulting collection of articles were assessed based on information in the title and abstract for their fit to the following criteria: a) presents data for shallow water marine animal(s), b) only tropical Indo-Pacific studies bounded by the geographical limits of: north to Tropic of Cancer (30˚ N), south to Tropic of Capricorn (30˚ S), west to Cape Town, South Africa (20˚ E), and east to the Eastern Pacific Barrier (140˚ W), c) contained genetic diversity data based on DNA sequences or microsatellites d) at least 3 populations were sampled per species, e) at least 15 individuals from a population were sampled for a population to be included, f) the study provided sample sizes and indices of genetic diversity at a population level, and g) the study provided latitude and longitude or

a map/description sufficient to identify the location of each sampling site to within 500 km. The final data set included 1544 genetic diversity data points across 75 species from 84 published studies captured by the literature search (Fig. 1). We recorded measures of expected heterozygosity (H_E), the probability of two copies of a locus sampled at random from the population have different allelic states (Wright 1922; Wright 1931; Nei 1978); haplotype diversity (*h*), the chance of two haplotypes sampled at random from a population being different from each other (Nei and Li 1979; Nei and Tajima 1981); and nucleotide diversity (π), the average number of nucleotide differences per site between two randomly chosen DNA sequences (Nei and Li 1979; Nei 1987). The haplotype diversity data set comprised 61 species, the nucleotide diversity data set comprised 57 species and the expected heterozygosity data set comprised 22 species. Data were compiled as reported from the original studies, or calculated from tables if single locus data were given for microsatellites. The data for the three diversity indices (*h*, π, H_E) were analysed separately, given that *h*, π and H_E measure different aspects of genetic diversity (DeWoody and Avise 2000). The haplotype and nucleotide diversity data sets contained almost the same set of studies given that these two indices are often reported together for studies using sequence data.

Species richness data

Data on species richness were derived from range polygons for reef animals in the families Acanthuridae, Labridae (wrasse and parrotfishes), Chaetodontidae, Epinephelidae, Elapidae and Pomacanthidae (IUCN 2014) (Fig. 2). We assembled range maps from 64 species of Elapidae, 86 species of Pomacanthidae, 128 species of Chaetodontidae, 163 species of Epinephelidae, 611 species of Labridae, 80 species of Acanthuridae within ArcMap 10 (ESRI, Redlands, CA). To determine a proxy of species richness at each location for which genetic diversity was available, locations were buffered by 100 km to account for the coarse scale of the range maps. The sample locations were used to sum the number of species ranges each intersects. All spatial analyses were conducted in ArcMap 10 (ESRI, Redlands, CA) using a Cylindrical Equal Area projection centred on 130˚E. Those species represented in both the genetic diversity and species richness data can be seen in Supp. Tables 1-3.

Statistical models

Standard indices of genetic diversity, such as haplotype diversity (*h*), nucleotide diversity (π) and heterozygosity (H_E), are based on proportions (bounded by 0 and 1) and are often skewed (nucleotide diversity is generally right-skewed and haplotype diversity, the opposite). These data violate the assumptions of standard linear models. Therefore we used *beta* regression, while not traditionally used in ecology, because it allows for the modelling of variables that are continuous between zero and one (Ferrari and Cribari-Neto 2004).

Multi-species approaches to inferring general patterns, such as the SGDC, demand consideration of the variation among species. We explicitly accounted for this variation by including a random effect of species on the slope and intercept of our *beta* regression models. This allowed us to assess the correlation between species and genetic diversity while controlling for the variation existing

Figure 1. The study region boundary is shown with sampling intensity of genetic diversity data points.

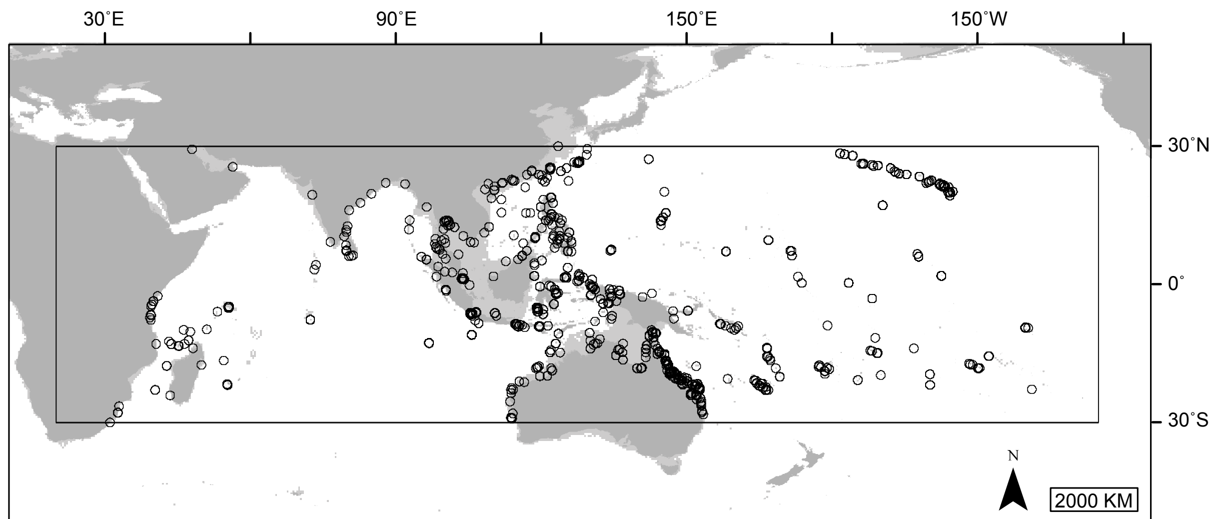
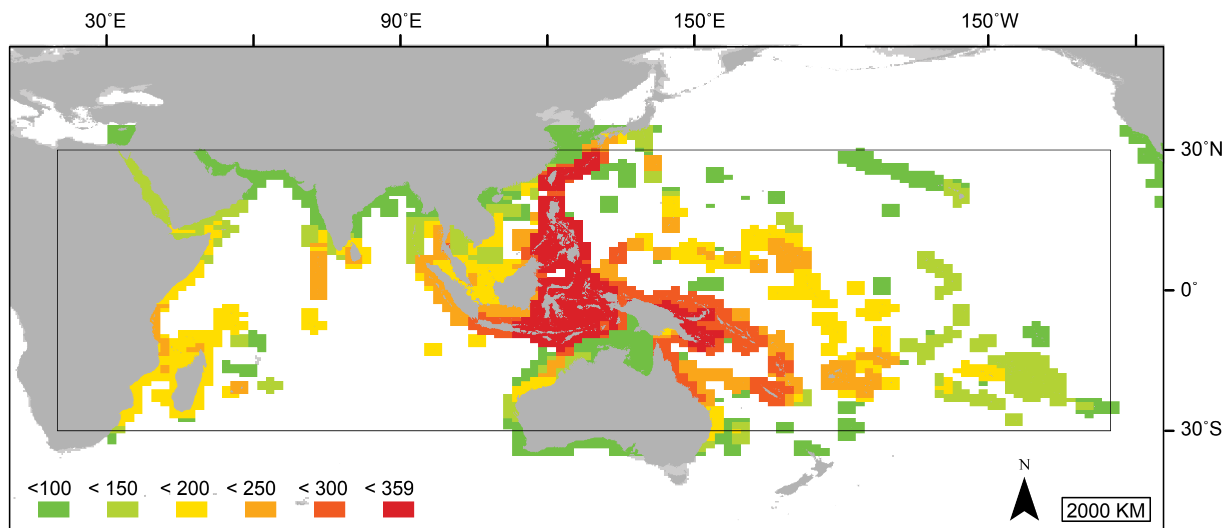


Figure 2: Estimated species richness within six families of marine taxa (Acanthuridae, Labridae, Chaetodontidae, Epinephelidae, Elapidae and Pomacanthidae) represented as a heatmap. Red indicates the highest and green is the lowest species richness.



between species. Below, we detail the data, the *beta* regression model and model parameters we used to test the SGDC across multiple tropical marine species.

Beta regression is recommended when regressing continuous variables bounded by zero and one, as forcing such variables into a linear model violates assumptions of normality resulting in unreliable estimates of explained variance (Ferrari and Cribari-Neto 2004; Branscum et al. 2007). *Beta* regression allows the modelling of both the central tendency and dispersion of heteroscedastic data without making restrictive distributional assumptions (Smithson and Verkuilen 2006).

Analysis of multi-species data brings with it the need to account for the differences among species and different genetic markers. In order to account for these differences statistically, random

effects for species and marker can be incorporated into regression models (Bolker et al. 2009). Grouping factors allow the modelled relationship to vary depending on the value of the grouping factor (in this case a unique index representing the species/marker combination) (Moulton 1986). Computational Bayesian approaches offer the capacity to handle *beta* regression with arbitrary numbers of random effects within a flexible framework (Buckley 2002; Smithson and Verkuilen 2006; Figueroa-Zúñiga et al. 2013).

We were interested in determining whether the levels of within-species genetic diversity correlate with species richness of reef fauna allowing for the fact that correlations may vary among species-by-marker combinations. To account for this we modelled variation among species-by-marker combinations using random-effects. However, it was not possible to estimate parameters with random-effects for both species identity and marker simultaneously because, in general, a species was represented by data for only a single genetic marker, so that species identity and marker effects were highly confounded. We therefore removed all occurrences of more than one marker for a species from the data set, deleting the marker which allowed us to retain the greatest number of data points across the greatest geographic scope. Where single individuals had been genotyped with more than one marker we chose the marker showing the greatest genetic variation. The correlation between haplotype and nucleotide diversity for those species with both sets of data was assessed in the *beta* regression framework described below, but with nucleotide diversity substituted for species richness to ascertain whether these two marker types behaved similarly.

Prior to analysis, diversity statistics were transformed using the following formula to prevent the inclusion of zeroes or ones in the analysis to meet the assumption that the *beta* distribution is bounded (0,1). The transformation we applied is below, following Ferrari and Cribari-Neto (2004).

$$g' = \begin{cases} g \text{ if } 0 < g > 1 \\ g\,(n\text{-}1) + 0.5/n \text{ if } g = 0 \text{ or } g = 1 \end{cases}$$

where g' is the transformed diversity statistic, g is the original diversity statistic, and n is the total number of data points. Species richness estimates were standardised prior to analyses to have a mean of zero and standard deviation of one.

Our model for the relationship between species richness and genetic diversity was:

$$\begin{aligned} y_{i,s} &\sim \text{Beta}\,(\mu_{i,s}\phi, (1\text{-}\mu_{i,s})\,\phi) \\ \text{logit}(\mu_{i,s}) &= \beta_{0,s} + \beta_{1,s} x_i \\ \beta_{0,s} &\sim \text{Normal}\,(\beta_0, \sigma^2_0) \\ \beta_{1,s} &\sim \text{Normal}\,(\beta_1, \sigma^2_1) \end{aligned}$$

where $y_{i,s}$ is the genetic diversity for species s at location i; x_i is the species richness at location i; μ is the expected genetic diversity for species s at location i; ϕ is the precision parameter; $\beta_{0,s}$ is a normally distributed random-effect for the intercept; and $\beta_{1,s}$ is a normally distributed random effect for the slope.

Uninformative priors for the mean intercept and slope parameters were assumed to have a normal distribution with mean of 0 and precision of 0.0001. Priors on the precision parameter (ϕ) and the standard deviation of the random effects were assumed to follow a uniform distribution with a minimum of zero and maximum of 1000. Initial parameter values were chosen so that they were overdispersed according to Gilks (2005). We randomly generated initial values from a uniform (-10,10) distribution for mean slopes and intercepts and a uniform (0,10) distribution for random-effect variance parameters. Initial values for individual species slopes and intercepts were set at zero.

For each of the three diversity indices three Markov chains were run for 500,000 iterations with the first 50,000 discarded and a thinning rate of 100. Convergence to a stationary distribution was assumed when the potential scale reduction factor ($\hat{R}$) for each parameter reached one (Gelman 1996; Sturtz et al. 2005). Parameter estimates for the mean slope and mean intercept, as well as the individual species slopes and intercepts, were recorded.

The support for a positive relationship between genetic diversity (h, π, H_E) and species richness of reef fauna was tested by comparing the model described above to the same model but with the mean slope fixed at zero. We calculated the Deviance Information Criteria (DIC) of the normal and zero slope models and compared the DICs to identify the most parsimonious model. The lower the DIC the more parsimonious the model (Spiegelhalter et al. 2002). Models were also assessed based on the credible intervals around the estimate of the mean slope. If 95% credible intervals did not contain zero this was interpreted as support for a positive SGDC. The fit of each model to the data was assessed using posterior predictive checks (Gelman 1996) comparing the sums of squared residuals of the actual data and data predicted by the model based on the following statistic:

$$T = \sum_{i=1}^{n} \left(\frac{\left(\mu_i - H_{predicted_i}\right)^2}{\left(\mu_i \times (1 - \mu_i)\right)} \right) - \sum_{i=1}^{n} \left(\frac{\left(\mu_i - H_{actual_i}\right)^2}{\left(\mu_i \times (1 - \mu_i)\right)} \right)$$

Values of T greater than zero indicate that the model represents variability in the data at least as well as data generated from the model. Poor model fit was indicated if the percentage of times T greater than zero was less than 5% (i.e., a 5% significance level) (Gelman 1996). This is presented in Table 1 as PPC p-value.

Spatial autocorrelation

We assessed spatial autocorrelation among genetic diversity data within each species using Moran's I statistic in ArcMap 10 (ESRI, Redlands, CA). We also assessed the spatial autocorrelation in model residuals by constructing correlograms in the R package *ncf* (Bjornstad 2009) with increments set at 100 km. to reflect the coarsest resolution of the input data (species richness).

RESULTS

All three genetic diversity indices (h, π, H_E) showed positive mean slopes in our *beta* regression models for reef animals (Table 1). DIC support for model predicting H_E with species richness was strong, indicating high confidence that the slope of the SGDC was different from zero (Table 1). DIC support for models predicting h and π, however, was marginal. Credible intervals did not include zero for h, π, and H_E thus providing further evidence that the mean slope of the SGDC is positive for these diversity indices (Fig. 3).

Variation in slopes for individual species:marker combinations

For models predicting haplotype diversity (h) with species richness, 53 out of 61 species showed a positive slope (Table 2). For models predicting nucleotide diversity (π) with species richness, 48 out of 57 species showed a positive slope (Table 2). The few negative SGDC slopes represented species from three phyla and three different genetic markers so there was no apparent taxonomic or molecular signal in the species showing negative slopes. All 22 species considered in heterozygosity (H_E) models showed positive mean slopes.

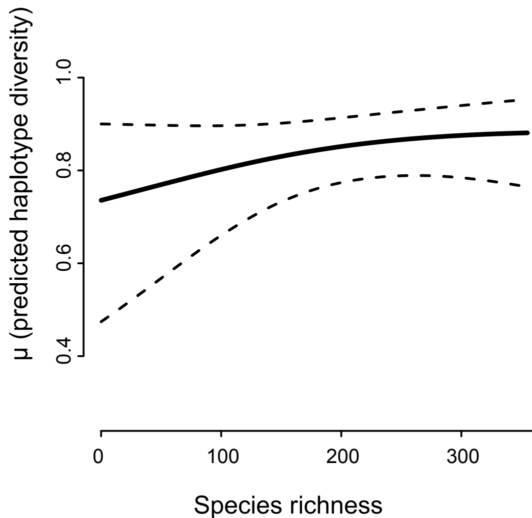
Posterior predictive checks

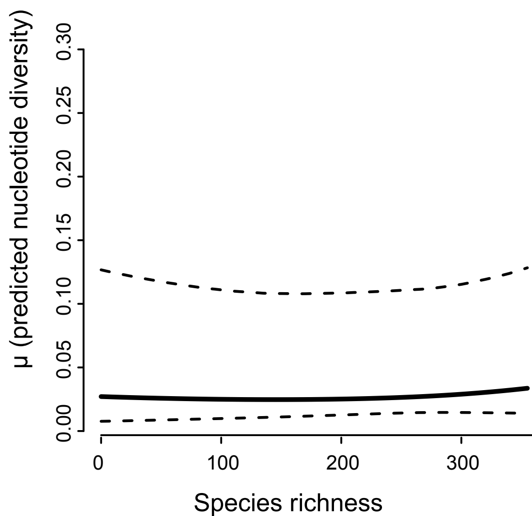
Models predicting h and H_E with species richness fit the data well: comparisons of the residuals of the actual data compared to the residuals of predicted data favoured the actual data 54% of the time for haplotype diversity (h) models and 51% of the time for heterozygosity (H_E) calculated for microsatellites. Therefore there was no evidence that these models were as significantly poor fit to the data at the 0.05 significance level. Nucleotide diversity (π) models fitted poorly, only 2.5% of the comparisons favoured the actual data.

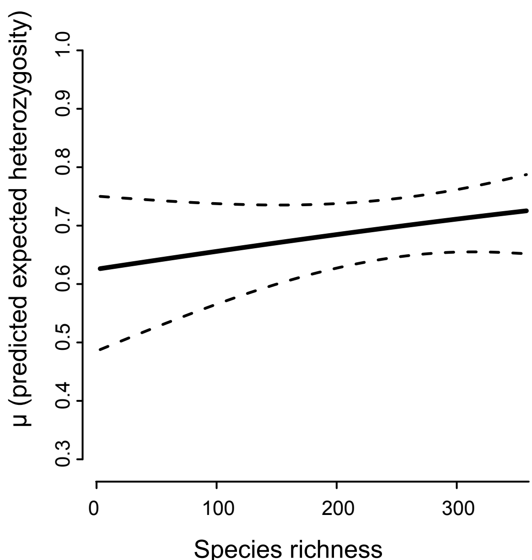
Table 1. Posterior estimates from each SGDC beta regression model. Mean slope values are averaged across all species. PPC values . Support for the full model is derived from DIC scores lower than the null model where the slope is fixed at zero. MC Error is a measure of the accuracy of the Markov Chain used to infer posterior estimates.

Genetic diversity predictor	Parameter	Mean	SD	2.5%	97.5%	MC Error	PPC p-value	DIC Full Model	DIC Slope as 0	DIC support for full model
Haplotype diversity (*h*)	SGDC slope	0.297	0.091	0.117	0.478	0.00042	0.541	-1821.6	-1820.9	Marginal
	SGDC intercept	2.217	0.162	1.902	2.539	0.00072				
	Inverse precision	0.060	0.004	0.052	0.070	0.00002				
Nucleotide diversity (π)	SGDC slope	0.127	0.048	0.037	0.226	0.00066	0.024	-4622	-4620.3	Marginal
	SGDC intercept	-4.631	0.142	-4.911	-4.357	0.00125				
	Inverse precision	0.002	0.000	0.001	0.002	0.00011				
Heterozygosity (HE)	SGDC slope	0.122	0.048	0.027	0.215	0.00108	0.508	-583.2	-567.5	Strong
	SGDC intercept	0.909	0.152	0.605	1.213	0.00143				
	Inverse precision	0.016	0.002	0.013	0.019	0.00002				

Figure 3. Graphical plots of the predicted genetic diversity across all species (μ) as a function of species richness. The solid line shows the mean posterior estimate of the relationship between haplotype diversity (A), nucleotide diversity (B) and heterozygosity (C); and standardised species richness. 95% credible limits are shown as dashed lines.







Spatial dependence

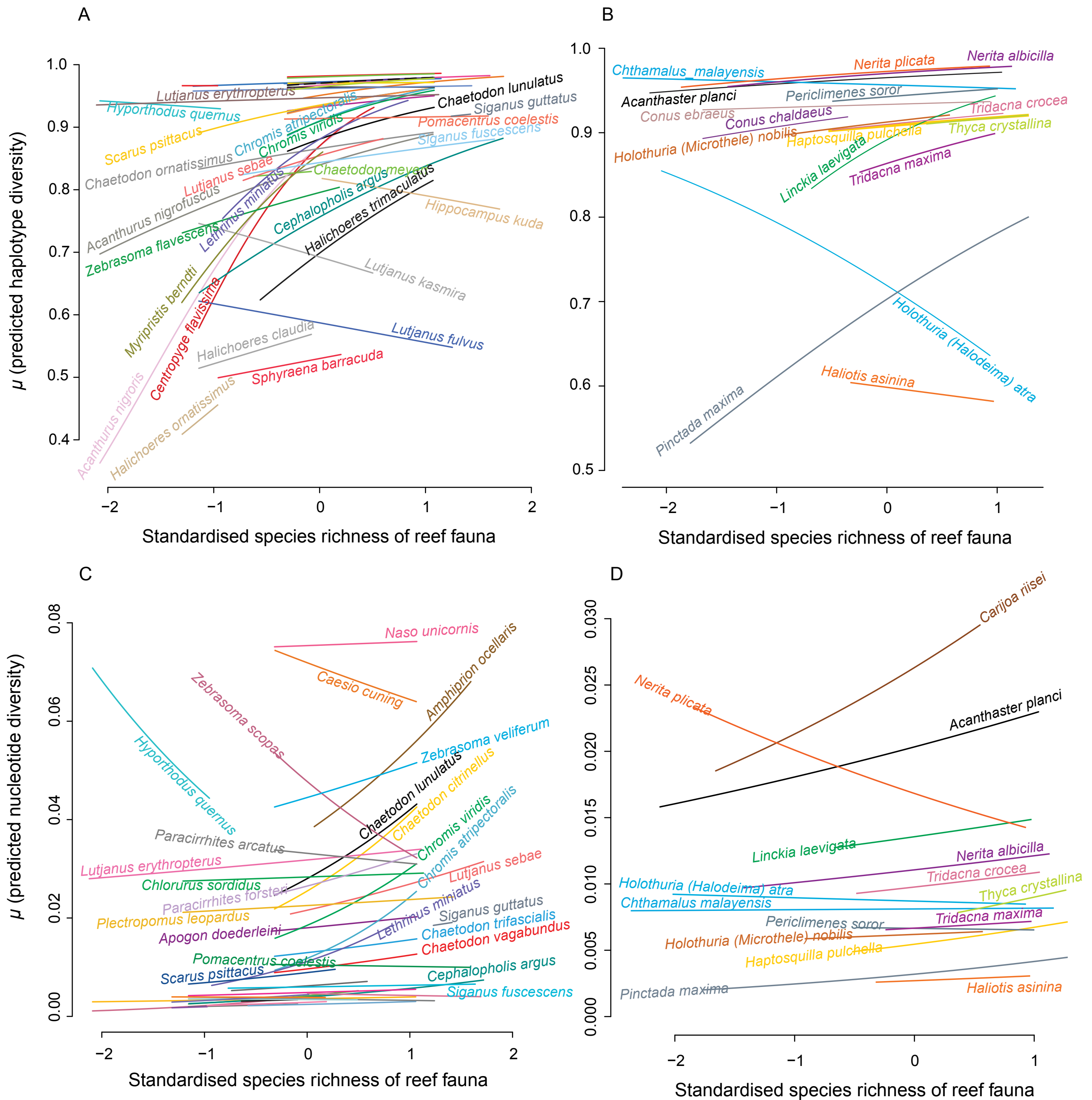
The genetic diversity data for most species were not significantly spatially autocorrelated (Supp. Tables 1-3). We found no evidence of spatial dependence of the *beta* regression residuals. Correlograms for each model (Fig. 6) did not display clear patterns associated with spatial autocorrelation. The correlograms all centre around zero, with the few significant correlations falling in equal parts above and below the line with no directional change across the graph, indicating no correlation between the model residuals and spatial distance.

DISCUSSION

In the first test of the species-genetic diversity correlation to explicitly account for variation among species we found consistent support for a positive correlation between species richness and genetic diversity. By synthesizing published data for 75 species of marine reef animals from the Indo-Pacific we were able to test the SGDC across an unprecedented spatial and taxonomic extent. Our results matched expectations of a positive SGDC. Mean slopes for the three genetic diversity categories were all positive (Table 1), with the strongest DIC support for HE and moderate support for mtDNA h and π. Similarly, while the majority of individual species slopes were positive (Fig. 4), and a few were negative (Supp. Tables 4-6).

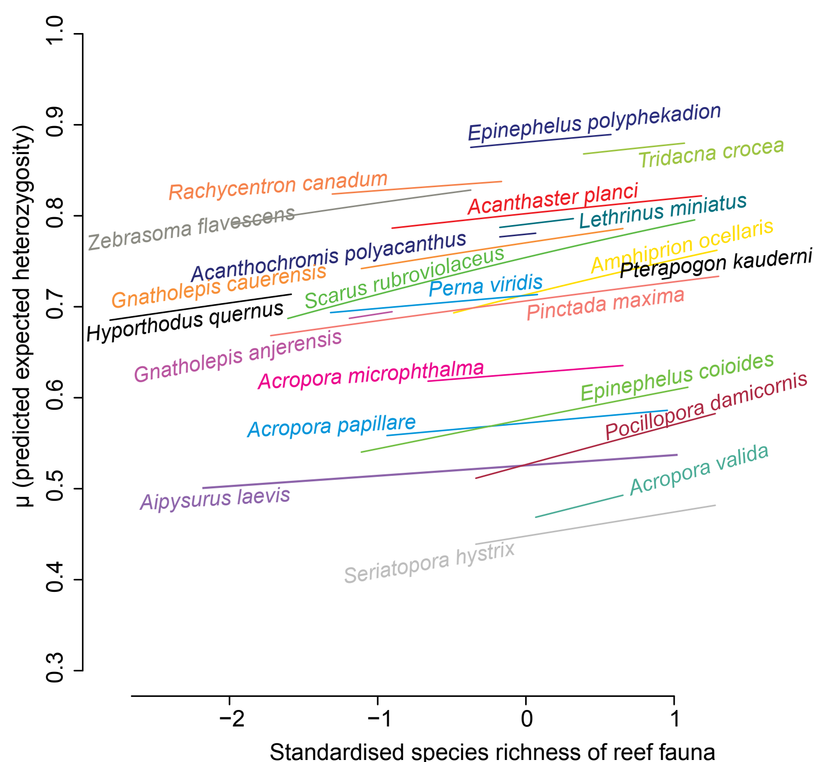
By focusing on shallow water reef species, we predicted a positive SGDC because communities existing in discrete habitat patches have been been found to demonstrate this relationship most strongly

Figure 4. Individual species SGDC slopes for mtDNA diversity indices. Species are divided into reef fishes (A, C) and invertebrates (B, D) for ease of viewing. Colours are the same for species across the upper (haplotype diversity) and lower (nucleotide diversity) graphs.



(Vellend et al. 2014). In the present study, this expectation received support on average across species and individually for most species. Haplotype diversity and expected heterozygosity are equivalent measures in that they estimate the probability of picking different hapotypes or alleles. Nucleotide diversity, however, measures the magnitude of the differences between individual DNA sequences and would be more akin to a phylogenetic metric of species diversity. Nucleotide diversity is greatest when divergent clades are sympatric and thus could be biased in cases of secondary contact, whereas haplotype diversity should be less affected by such circumstances. Despite these differences, in the instances where we were able to compare haplotype and nucleotide

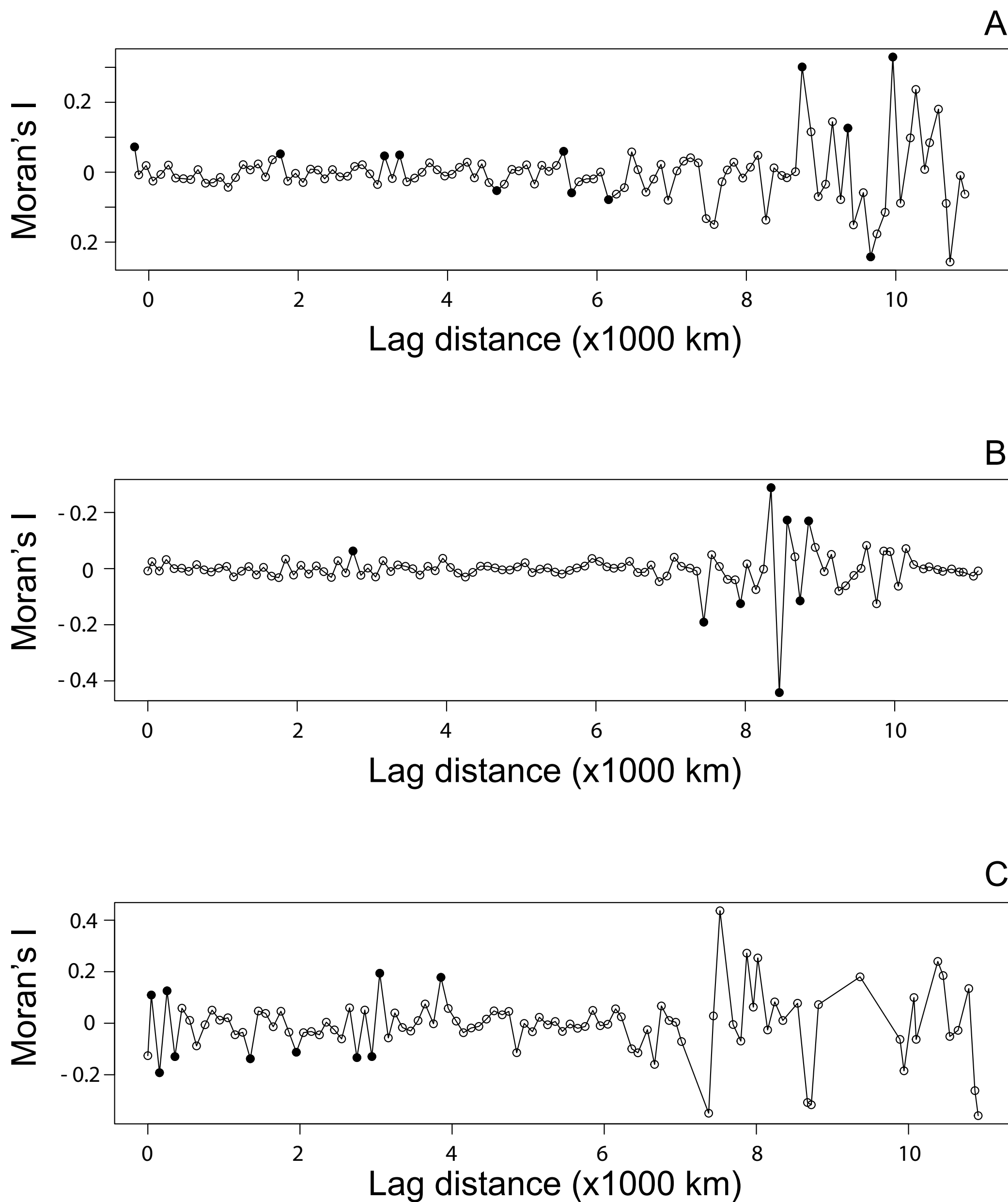
Figure 5. Individual species mean SGDC slopes for heterozygosity calculated from microsatellites. All species are shown on a single axis. Colours for species also occurring in Fig. 4 are the same.



diversity for the same sets of studies the two indices showed a positive correlation, suggesting that they are not behaving differently in these species. Alternatively, the poor model fit could be due to the upper bound of nucleotide diversity being poorly defined, meaning that, while it theoretically scales from zero to one, intraspecific nucleotide diversity in marine fishes is generally low (less than 0.01 in Australian marine fishes, Ward et al. 2005). In this sense, nucleotide diversity data may not be appropriate for modelling with *beta* regression where the assumption is that data are continuous between known upper and lower bounds.

Despite the overall tendency towards a positive relationship between species and genetic diversity, individual species slopes were highly variable and such high levels of variation among species genetic responses seem to characterise SGDC investigations (Robinson et al. 2010; Messmer et al. 2012). Thus, although a SGDC might arise from parallel responses to geography or reciprocal influences of species diversity on genetic diversity and *vice versa,* the exact dynamics of dispersal and genetic drift are likely to vary among species due to their life history attributes. That we see high levels of among species variation in the genetic response variables points to the strong effects of species differences in ecology, demography and evolutionary history (along with scaling issues and other sources of sampling noise – detailed in the following subsection). Including additional factors in the predictive model, such as life history traits of species might reveal consistent genetic patterns among sets of species that are predictable by natural history attributes. This is an important area for future work.

Figure 6. Correlograms of the correlation among model residuals at 100 km distance classes (horizontal axis). Correlation among residuals for models predicting A. haplotype diversity, B. nucleotide diversity and C. expected heterozygosity is displayed as correlation coefficients on the vertical axis. Filled circles indicate significant correlations at $p < 0.05$ (Moran's I statistic using 1000 permutations).



In addition, emergent patterns of genetic and species diversity may differ in spatial and temporal scales, reducing the signal of the relationship between them. Species richness gradients may be fairly insensitive to stochastic changes in population sizes, provided species do not go extinct. Conversely, genetic diversity of populations can be strongly affected by such changes, meaning that genetic diversity is perhaps more likely to fluctuate over time than is species diversity. Indeed, there is growing evidence that these single time-point snapshots of genetic diversity (such as those used in this study) are unreliable for some species, especially those with short generation times (Hedgecock 1994; Planes and Lenfant 2002; Klanten et al. 2007; Christie et al. 2010).

The only other study to examine the SGDC across multiple marine species (Messmer et al. 2012) reported a positive association between haplotype diversity and species richness in coral reef fishes across three locations in the Pacific Ocean. Where Messmer et al. (2012) summed genetic diversity indices across species prior to correlation, we kept these data separate and included species grouping factors to account for variation among species. This is important to account for species-specific variation such as demographic history, abundance or dispersal capacity that will affect the genetic diversity within populations. For the eleven species included in both analyses, we find individual species mean slopes that qualitatively match the individual species plots for both h and π in Fig. 3 of Messmer et al (2012), but the credible intervals of our estimates overlap zero in each case (see Supp Tables 4 & 5). Nonetheless the overall positive relationship found in both studies is consistent with the SGDC.

Our novel approach maximises the power of SGDC analyses by using genetic diversity data from a large number of species while allowing for the variation introduced by species-specific differences. Because our approach did not compare empirical estimates of species and genetic diversity at comparable resolutions, however, we may have a low capacity to detect a real biological relationship. Species richness was derived by summing the overlapping range maps for six families of marine fishes (see Fig. 2) and will have smoothed the species richness pattern, providing species richness estimates at a coarse resolution (100 km grid cells). These estimates almost certainly do not reflect the *alpha* diversity (species richness) on the reef patch from which the genetic data were collected, although they should represent the richness of the species pool. However, because many marine animals may be capable of dispersing long distances especially when considering time scales relevant for range expansion and gene flow (Mora et al. 2011; Treml et al. 2012), the coarse spatial resolution employed here may be an appropriate resolution representing relevant broad-scale biological phenomena. Recently published work using high-resolution reef species survey data offers excellent scope for future work to explore finer spatial scales (Stuart-Smith et al. 2013).

In order to synthesise data from multiple sources genetic diversity for marine Indo-Pacific taxa we needed an approach that would allow combining of these data in a robust framework. The analysis had to accommodate the proportional nature of genetic diversity indices and allow us to account for the differences among species that may arise due to differences in ecology, life history and mutation rate. Mixed-effects *beta* regression offers solutions to both of these problems. Further, mixed-effects *beta* regression can allow explicit incorporation of species characteristics (e.g. traits influencing dispersal (Luiz et al. 2013)) which will allow future studies to understand the factors underlying species' deviations from the SGDC. Furthermore, the framework allows multiple predictor variables to be compared and their importance ranked with model selection (Burnham and Anderson 2002). This allows assessment of the relative importance of multiple spatial and environmental factors, such as the area and isolation of habitat patches, in determining patterns of genetic diversity. We have demonstrated that this approach offers a flexible and robust method for interrogating haplotype diversity and heterozygosity data from multiple species simultaeously

Limitations of approach

While a true mismatch may exist between genetic and species richness in the marine environment, it is also possible that our method did not compare measures of these properties at comparable scales. The estimate used to determine species richness (overlapping range maps for six families of marine animals) is likely to have smoothed the species richness pattern, providing species richness estimates at a coarse resolution. Given this possible scale mismatch between the two estimates, a positive correlation might be obscured that could be revealed were species richness estimated at a finer resolution. These data were not available at the broad spatial extent of the Indo-Pacific at the time of our analyses, but recently published work using high-resolution reef survey data offers excellent scope for future work (Stuart-Smith et al. 2013). These data also offer the opportunity to factor in the abundance of species, rather than just species richness. Abundant species are likely to show higher genetic diversity (McCusker and Bentzen 2010) than rare species, so if communities have high evenness we might see less variation among species in their genetic diversity.

The genetic diversity data we used took the form of diversity indices reported in the literature, as has been a common tactic in analyses of the SGDC (Vellend et al. 2014). These summary statistics do not allow any quality control of the data; if poor quality DNA sequences were used in original analyses there is no way of controlling for their effect. To maximise the quality of these estimates we limited our data set to population samples greater than 15 individuals, but if a single poor quality sequence had been included this would have affected estimates of genetic diversity. In particular, nucleotide diversity is the most sensitive to this type of error. Incorrectly called bases in a sequence will inflate estimates of nucleotide diversity far more than haplotype diversity. The geographic resolution of these summary data is also unknown. Different collection strategies may mean that some genetic diversity statistics could be derived from samples collected across a wide area, while others may have been collected within a very small area. These issues point to raw, georeferenced and properly accessioned genetic data being the gold standard for molecular genetic research. These high quality data are accummulating rapidly through the work of the NESCent working group mentioned in the first chapter (DIPnet: Diversity in the Indo-Pacific network: http://indopacificnetwork.wikispaces.com/).

Statistical model

In order to synthesise data from multiple sources genetic diversity for marine Indo-Pacific taxa we sought an approach that would allow us to combine these data in a robust framework. The analysis had to accommodate the proportional nature of genetic diversity indices and account for the differences among species in ecology, life history and mutation rate. Mixed *beta* regression offers solutions to both of these problems.

The *beta* regression model performed well for haplotype diversity and heterozygosity data, however models predicting nucleotide diversity were a poor fit. This lack of fit could be due to the sensitivity of nucleotide diversity to population structure within species. Unlike haplotype diversity, nucleotide

diversity measures the magnitude of the differences between individual DNA sequences making it sensitive to overlap of divergent clades where haplotype diversity is not. This latter explanation seems unlikely, however, as the two diversity indices showed a positive correlation where calculated for the same individuals. Alternatively, the lack of fit could be due to the upper bound of nucleotide diversity being poorly defined, meaning that, while theoretically it scales from zero to one, it is unlikely to ever reach unity. In this sense, nucleotide diversity data may not be appropriate for modelling with *beta* regression where the assumption is that data are continuous between known upper and lower bounds.

Mixed *beta* regression allows explicit incorporation of species characteristics (e.g. traits influencing dispersal (Luiz et al. 2013)) which will allow future studies to understand the factors underlying species' deviations from the SGDC. Furthermore, the framework allows multiple predictor variables to be compared and their importance ranked with model selection (Burnham and Anderson 2002). This allows assessment of the relative importance of multiple spatial and environmental factors, such as the area and isolation of habitat patches, in determining patterns of genetic diversity. We have demonstrated that this approach offers a flexible and robust method for interrogating haplotype diversity and heterozygosity data from multiple species.

Conclusions

Genetic diversity is increasingly recognized as an important property of resilient, productive populations (Hughes and Stachowicz 2004; Wimp et al. 2004; Crutsinger et al. 2006; Crutsinger et al. 2008). Overwhelmingly, though, biodiversity is measured in the units of species and decisions regarding conservation priorities are made based on species distribution or habitat data (Ball 2009; Drummond et al. 2009; Watts et al. 2009). This is understandable, given the historical expense of producing genetic data, but the cost of molecular work is rapidly falling. Additionally, access to published genetic data is improving with the increased stringency with which journals and funding bodies are requiring data to be made openly accessible. Already, managers targeting the conservation of a particular species may have access to data on the genetic structure of its meta-populations or the levels of standing genetic variation in a population of interest (Beger et al. 2014). However, when management targets the continued resilience of whole communities it may not feasible to assess the levels of genetic diversity. In this situation, being able to predict genetic diversity would be highly desirable. While our results suggest that targeting areas with greater species richness should result in capturing greater genetic diversity as well, this benefit may be highly variable among species. Where particular species are of interest, focusing at that scale will be of greater relevance. Future work taking advantage of higher resolution species richness and environmental information, and integrating life history characteristics into models, may shed further light on the generality and strength of the correlation between species and genetic diversity.

Literature Cited

Antonovics, J. 1976. The Input from Population Genetics:" The New Ecological Genetics". Systematic Botany:233-245.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics:489-522.

Ball, I. R., Possingham, H.P., Watts, M. 2009. Marxan and relatives: Software for spatial conservation prioritisation. . Pages 185-195 in A. Moilanen, Wilson, K.A., Possingham, H.P., editor. Spatial conservation prioritisation: Quantitative methods and computational tools. Oxford University Press, Oxford, UK., Oxford.

Beger, M., K. A. Selkoe, E. Treml, P. H. Barber, S. von der Heyden, E. D. Crandall, R. J. Toonen, and C. Riginos. 2014. Evolving coral reef conservation with genetic information. Bulletin of Marine Science 90:159-185.

Bellwood, D. R., W. Renema, and B. R. Rosen. 2012. Biodiversity hotspots, evolution and coral reef biogeography: a review.in D. Gower, Johnson, K., Richardson, R., Rosen, B., Rüber, L., Williams, S., editor. Biotic Evolution and Environmental Change in Southeast Asia Cambridge University Press., Cambridge, UK.

Bird, C. E., I. Fernandez-Silva, D. J. Skillings, and R. J. Toonen. 2012. Sympatric speciation in the post "Modern Synthesis" era of evolutionary biology. Evolutionary Biology 39:158-180.

Bjornstad, O. 2009. *ncf*: spatial nonparametric covariance functions. R package version 1.1-3. See http://onb. Ent. Psu. Edu/onb1/r.

Blum, M. J., M. J. Bagley, D. M. Walters, S. A. Jackson, F. B. Daniel, D. J. Chaloud, and B. S. Cade. 2012. Genetic diversity and species diversity of stream fishes covary across a land-use gradient. Oecologia 168:83-95.

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution 24:127-135.

Branscum, A. J., W. O. Johnson, and M. C. Thurmond. 2007. Bayesian *beta* regression: applications to household expenditure data and genetic distance between foot-and-mouth disease viruses. Australian & New Zealand Journal of Statistics 49:287-301.

Buckley, J. 2002. Estimation of models with *beta*-distributed dependent variables: A replication and extension of Paolino (2001). Political Analysis 11:1-12.

Burnham, K. P. and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd edition. Springer-Verlag, New York.

Carpenter, K. E., P. H. Barber, E. D. Crandall, B. M. Manjaji-Matsumoto, M. A. Juinio-Menez, M. D. Santos, C. J. Starger, and A. H. A. Toha10. 2011. Comparative phylogeography of the Coral Triangle and implications for marine management. Journal of Marine Biology.

Christie, M. R., D. W. Johnson, C. D. Stallings, and M. A. Hixon. 2010. Self-recruitment and sweepstakes reproduction amid extensive gene flow in a coral-reef fish. Molecular Ecology 19:1042-1057.

Cleary, D. F., C. Fauvelot, M. J. Genner, S. B. Menken, and A. O. Mooers. 2006. Parallel responses of species and genetic diversity to El Niño Southern Oscillation-induced environmental destruction. Ecology Letters 9:304-310.

Connolly, S. R., D. R. Bellwood, and T. P. Hughes. 2003. Indo-Pacific biodiversity of coral reefs: deviations from a mid-domain model. Ecology 84:2178-2190.

Crutsinger, G. M., M. D. Collins, J. A. Fordyce, Z. Gompert, C. C. Nice, and N. J. Sanders. 2006. Plant genotypic diversity predicts community structure and governs an ecosystem process. Science 313:966-968.

Crutsinger, G. M., L. Souza, and N. J. Sanders. 2008. Intraspecific diversity and dominant genotypes resist plant invasions. Ecology Letters 11:16-23.

Dawson, M. N. 2012. Parallel phylogeographic structure in ecologically similar sympatric sister taxa. Molecular Ecology 21:987-1004.

DeWoody, J. A. and J. C. Avise. 2000. Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. Journal of Fish Biology 56:461-473.

Drummond, S. P., K. A. Wilson, E. Meijaard, M. Watts, R. Dennis, L. Christy, and H. P. Possingham. 2009. Influence of a threatened species focus on conservation planning. Conservation Biology 24:441-449.

Endler, J. A. 1986. Natural selection in the wild. Princeton University Press.

Evanno, G., E. Castella, C. Antoine, G. Paillat, and J. Goudet. 2009. Parallel changes in genetic diversity and species diversity following a natural disturbance. Molecular Ecology 18:1137-1144.

Ferrari, S. and F. Cribari-Neto. 2004. *Beta* regression for modelling rates and proportions. Journal of Applied Statistics 31:799-815.

Figueroa-Zúñiga, J. I., R. B. Arellano-Valle, and S. L. P. Ferrari. 2013. Mixed *beta* regression: A Bayesian perspective. Computational Statistics & Data Analysis 61:137-147.

Finn, D. S. and N. L. Poff. 2011. Examining spatial concordance of genetic and species diversity patterns to evaluate the role of dispersal limitation in structuring headwater metacommunities. Journal of the North American Benthological Society 30:273-283.

Gelman, A. 1996. Inference and monitoring convergence. Pages 131-143 Markov chain Monte Carlo in practice. Springer.

Gilks, W. R. 2005. Markov chain Monte Carlo. Encyclopedia of Biostatistics. Wiley Online Library.
Hedgecock, D. 1994. Temporal and spatial genetic structure of marine animal populations in the California Current. California Cooperative Oceanic Fisheries Investigations Reports 35:73-81.

Hughes, A. R. and J. J. Stachowicz. 2004. Genetic diversity enhances the resistance of a seagrass ecosystem to disturbance. Proceedings of the National Academy of Sciences of the United States of America 101:8998.

IUCN. 2014. The IUCN Red List of Threatened Species.

Kahilainen, A., M. Puurtinen, and J. S. Kotiaho. 2014. Conservation implications of species–genetic diversity correlations. Global Ecology and Conservation 2:315-323.

Keyse, J., E. D. Crandall, R. J. Toonen, C. P. Meyer, E. A. Treml, and C. Riginos. 2014. The scope of published population genetic data for Indo-Pacific marine fauna and future research opportunities in the region. Bulletin of Marine Science 90:47-78.

Klanten, O. S., J. H. Choat, and L. van Herwerden. 2007. Extreme genetic diversity and temporal rather than spatial partitioning in a widely distributed coral reef fish. Marine Biology 150:659-670.

Lamy, T., P. Jarne, F. Laroche, J. Ä. Pointier, G. Huth, A. Segard, and P. David. 2013. Variation in habitat connectivity generates positive correlations between species and genetic diversity in a metacommunity. Molecular Ecology 22:4445-4456.

Laroche, F., P. Jarne, T. Lamy, P. David, and F. Massol. 2015. A neutral theory for interpreting

correlations between species and genetic diversity in communities. The American Naturalist 185:59-69.

Luiz, O. J., A. P. Allen, D. R. Robertson, S. R. Floeter, M. Kulbicki, L. Vigliola, R. Becheler, and J. S. Madin. 2013. Adult and larval traits as determinants of geographic range size among tropical reef fishes. Proceedings of the National Academy of Sciences 110:16498-16502.

MacArthur, R. H. and E. O. Wilson. 1967. The theory of island biogeography. Princeton Univ Pr.

Messmer, V., G. P. Jones, P. L. Munday, and S. Planes. 2012. Concordance between genetic and species diversity in coral reef fishes across the Pacific Ocean biodiversity gradient. Evolution 66:3902-3917.

Moulton, B. R. 1986. Random group effects and the precision of regression estimates. Journal of Econometrics 32:385-397.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press.

Nei, M. and W. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Science 76:5269-5273.

Nei, M. and F. Tajima. 1981. DNA polymorphism detectable by restriction endonucleases. Genetics 97:145-163.

Noyer, C. and M. A. Becerro. 2012. Relationship between genetic, chemical, and bacterial diversity in the Atlanto-Mediterranean bath sponge *Spongia lamella*. Hydrobiologia 687:85-99.

Papadopoulou, A., I. Anastasiou, F. Spagopoulou, M. Stalimerou, S. Terzopoulou, A. Legakis, and A. P. Vogler. 2011. Testing the species-genetic diversity correlation in the Aegean archipelago: towards a haplotype-based macroecology? The American Naturalist 178:241-255.

Planes, S. and P. Lenfant. 2002. Temporal change in the genetic structure between and within cohorts of a marine fish, *Diplodus sargus*, induced by a large variance in individual reproductive success. Molecular Ecology 11:1515-1524.

Quinn, G. P. and M. J. Keough. 2002. Experimental design and data analysis for biologists. Cambridge University Press.

Renema, W., D. Bellwood, J. Braga, K. Bromfield, R. Hall, K. Johnson, P. Lunt, C. Meyer, L. McMonagle, and R. Morley. 2008. Hopping hotspots: global shifts in marine biodiversity. Science 321:654-657.

Reusch, T. B. H., A. Ehlers, A. Hämmerli, and B. Worm. 2005. Ecosystem recovery after climatic extremes enhanced by genotypic diversity. Proceedings of the National Academy of Sciences of the United States of America 102:2826.

Robinson, J. D., E. Diaz-Ferguson, M. F. Poelchau, S. Pennings, T. D. Bishop, and J. Wares. 2010. Multiscale diversity in the marshes of the Georgia coastal ecosystems LTER. Estuaries and Coasts 33:865-877.

Rundle, H. and P. Nosil. 2005. Ecological speciation. Ecol Letters 8:336–352.

Smithson, M. and J. Verkuilen. 2006. A better lemon squeezer? Maximum-likelihood regression with *beta*-distributed dependent variables. Psychological Methods 11:54-71.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64:583-639.

Struebig, M. J., T. Kingston, E. J. Petit, S. C. Le Comber, A. Zubaid, A. Mohd-Adnan, and S. J. Rossiter. 2011. Parallel declines in species and genetic diversity in tropical forest fragments. Ecology Letters 14:582-590.

Stuart-Smith, R. D., A. E. Bates, J. S. Lefcheck, J. E. Duffy, S. C. Baker, R. J. Thomson, J. F. Stuart-Smith, N. A. Hill, S. J. Kininmonth, L. Airoldi, M. A. Becerro, S. J. Campbell, T. P. Dawson, S. A. Navarrete, G. A. Soler, E. M. Strain, T. J. Willis, and G. J. Edgar. 2013. Integrating abundance and functional traits reveals new global hotspots of fish diversity. Nature 501:539-542.

Sturtz, S., U. Ligges, and A. E. Gelman. 2005. R2WinBUGS: a package for running WinBUGS from R. Journal of Statistical software 12:1-16.

Vellend, M. 2004. Parallel effects of land-use history on species diversity and genetic diversity of forest herbs. Ecology 85:3043-3055.

Vellend, M. and M. A. Geber. 2005. Connections between species diversity and genetic diversity. Ecology Letters 8:767-781.

Vellend, M., G. Lajoie, A. Bourret, C. Murria, S. W. Kembel, and D. Garant. 2014. Drawing ecological inferences from coincident patterns of population- and community-level biodiversity.

Molecular Ecology 23:2890-2901.

Via, S. 2001. Sympatric speciation in animals:the ugly duckling grows up. TREE 16:10.

Ward, R. D., T. S. Zemlak, B. H. Innes, P. R. Last, and P. D. Hebert. 2005. DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 360:1847-1857.

Watts, M. E., I. R. Ball, R. S. Stewart, C. J. Klein, K. Wilson, C. Steinback, R. Lourival, L. Kircher, and H. P. Possingham. 2009. Marxan with Zones: software for optimal conservation based land-and sea-use zoning. Environmental Modelling & Software 24:1513-1521.

Wehenkel, C., F. Bergmann, and H.-R. Gregorius. 2006. Is there a trade-off between species diversity and genetic diversity in forest tree communities? Plant Ecology 185:151-161.

Wei, X. and M. Jiang. 2012. Contrasting relationships between species diversity and genetic diversity in natural and disturbed forest tree communities. The New Phytologist 193:779-786.

Whitlock, R. 2014. Relationships between adaptive and neutral genetic diversity and ecological structure and functioning: a meta-analysis. The Journal of ecology 102:857-872.

Whittaker, R. H. 1972. Evolution and measurement of species diversity. Taxon 21:213-251.

Wimp, G. M., W. P. Young, S. A. Woolbright, G. D. Martinsen, P. Keim, and T. G. Whitham. 2004. Conserving plant genetic diversity for dependent animal communities. Ecology Letters 7:776-780.

Wright, S. 1922. Coefficients of inbreeding and relationship. The American Naturalist 56:330-338.

Wright, S. G. 1931. Evolution in Mendelian Populations. Genetics 16:0097-0159.

Wright, S. G. 1940. Breeding Structure of Populations in Relation to Speciation. The American Naturalist 74:232-248.

Supplementary Table 1. Species used in haplotype diversity analyses with number of locations sampled (N) and spatial autocorrelation of haplotype diversity within species (Moran's I, p and Z score). Citation numbers refer to Supp. Table 7. Shaded rows indicate species part of species richness data.

Species	Moran's i	p	Z score	Phylum	Family	Locus	N	Citation (s)
Acanthaster planci	2.213539	0.0004	3.515949	Echinodermata	Acanthasteridae	CR	44	151,153,170
Acanthurus nigrofuscus	0.429098	0.0262	2.223754	Chordata	Acanthuridae	Cyt. B	17	99
Acanthurus nigroris	0.01149	0.4697	0.722896	Chordata	Acanthuridae	Cyt. B	15	132
Aipysurus laevis	0.286588	0.217	1.23468	Chordata	Elapidae	ND4	9	51
Amphiprion ocellaris	0.473937	0.1144	1.57872	Chordata	Pomacentridae	CR	14	68
Apogon doederleini	-0.286383	1	0.0000	Chordata	Apogonidae	CR	3	124
Caesio cuning	0.169819	0.3884	0.862446	Chordata	Caesionidae	CR	23	167
Centropyge flavissima	-0.136483	0.9797	0.025471	Chordata	Pomacanthidae	Cyt. B	8	131
Cephalopholis argus	0.753305	0.0003	3.610183	Chordata	Epinephelidae	Cyt. B	19	125,164
Chaetodon citrinellus	NA	NA	NA	Chordata	Chaetodontidae	CR	3	149
Chaetodon lunulatus	NA	NA	NA	Chordata	Chaetodontidae	CR	3	149
Chaetodon meyeri	-0.259852	0.8558	0.181734	Chordata	Chaetodontidae	Cyt. B	4	165
Chaetodon ornatissimus	-1.037734	0.2196	-1.22751	Chordata	Chaetodontidae	Cyt. B	5	165
Chaetodon trifascialis	NA	NA	NA	Chordata	Chaetodontidae	CR	3	149
Chaetodon vagabundus	NA	NA	NA	Chordata	Chaetodontidae	CR	3	149
Chlorurus sordidus	-0.847041	0.1122	-1.58852	Chordata	Labridae	COI	8	13
Chromis atripectoralis	NA	NA	NA	Chordata	Pomacentridae	CR	3	149
Chromis viridis	NA	NA	NA	Chordata	Pomacentridae	CR	3	149
Chthamalus malayensis	0.01111	0.8162	0.232402	Arthropoda	Chthamalidae	COI	25	159
Conus chaldaeus	NA	NA	NA	Mollusca	Conidae	COI	3	146
Conus ebraeus	NA	NA	NA	Mollusca	Conidae	COI	3	31
Gobiodon quinquestrigatus	NA	NA	NA	Chordata	Gobiidae	CR	3	149
Halichoeres claudia	-0.17586	0.4853	0.697729	Chordata	Labridae	COI	4	148
Halichoeres ornatissimus	-0.328636	0.4026	-0.837	Chordata	Labridae	COI	9	148
Halichoeres trimaculatus	-0.159781	0.8884	0.140388	Chordata	Labridae	COI	6	148
Haliotis asinina	0.212235	0.2242	0.212235	Mollusca	Haliotidae	CO2	6	39
Haptosquilla pulchella	-0.25779	0.9053	-0.11893	Arthropoda	Protosquillidae	COI	6	12
Hippocampus kuda	-0.414318	1	0.0000	Chordata	Syngnathidae	CR	3	67
Holothuria (Halodeima) atra	0.057463	0.577	0.557728	Echinodermata	Holothuriidae	COI	18	174
Holothuria (Microthele) nobilis	-0.233255	0.619	-0.49723	Echinodermata	Holothuriidae	COI	12	70
Hyporthodus quernus	-0.202394	0.8203	-0.22711	Chordata	Serranidae	CR	9	173
Lethrinus miniatus	0.778659	0.1568	1.416007	Chordata	Lethrinidae	CR	4	142
Linckia laevigata	0.282063	0.1098	1.599195	Echinodermata	Ophidiasteridae	COI	21	7,43,150
Lutjanus erythropterus	-0.335906	0.3011	-1.03413	Chordata	Lutjanidae	CR	11	79
Lutjanus fulvus	0.229445	0.121	1.550676	Chordata	Lutjanidae	Cyt. B	9	35,164
Lutjanus kasmira	-0.246182	0.6301	-0.48162	Chordata	Lutjanidae	Cyt. B	10	35
Lutjanus sebae	-1.143771	1	0.0000	Chordata	Lutjanidae	CR	3	142
Myripristis berndti	0.784778	0	4.357373	Chordata	Holocentridae	Cyt. B	17	20,129
Naso brevirostris	NA	NA	NA	Chordata	Acanthuridae	CR	3	38
Naso unicornis	-0.457613	0.7701	-0.29219	Chordata	Acanthuridae	CR	4	38
Naso vlamingii	NA	NA	NA	Chordata	Acanthuridae	CR	4	41
Nerita albicilla	-0.122826	0.9483	-0.06486	Mollusca	Neritidae	COI	11	21,169
Nerita plicata	-0.099011	0.9129	-0.10943	Mollusca	Neritidae	COI	16	21,169
Paracirrhites arcatus	NA	NA	NA	Chordata	Cirrhitidae	CR	3	149
Paracirrhites forsteri	NA	NA	NA	Chordata	Cirrhitidae	CR	3	149
Periclimenes soror	-0.441858	0.1271	-1.52547	Arthropoda	Palaemonidae	COI	9	7
Pinctada maxima	-0.220386	0.8483	-0.19133	Mollusca	Pteriidae	COI	8	163
Plectropomus leopardus	-1.0000	0.1264	-1.52861	Chordata	Serranidae	CR	5	74
Pomacentrus coelestis	-0.395618	0.6613	-0.43807	Chordata	Pomacentridae	CR	6	109,124
Scarus psittacus	0.173552	0.3773	0.882839	Chordata	Labridae	CR	6	82
Siganus fuscescens	0.135382	0.0045	2.843976	Chordata	Siganidae	CR	19	118
Siganus guttatus	-0.569203	1	0.0000	Chordata	Siganidae	CR	3	113
Sphyraena barracuda	NA	NA	NA	Chordata	Sphyraenidae	Cyt. B	3	145
Stegastes fasciolatus	-0.201636	0.911	-0.1118	Chordata	Pomacentridae	CR	7	114
Thyca crystallina	-0.064642	0.9776	-0.02811	Mollusca	Eulimidae	COI	18	7,43
Triaenodon obesus	-0.110341	0.7942	0.260842	Chordata	Carcharhinidae	CR	7	166
Tridacna crocea	3.021912	0.0016	3.155412	Mollusca	Cardiidae	COI	26	22,44
Tridacna maxima	0.148007	0.3859	0.867078	Mollusca	Cardiidae	COI	7	58
Zebrasoma flavescens	0.081019	0.4022	0.837762	Chordata	Acanthuridae	Cyt. B	19	128
Zebrasoma scopas	NA	NA	NA	Chordata	Acanthuridae	CR	3	149
Zebrasoma veliferum	NA	NA	NA	Chordata	Acanthuridae	CR	3	149

Supplementary Table 2. Species used in nucleotide diversity analyses with number of locations sampled (N) and details of the spatial autocorrelation among nucleotide diversity within species (Moran's I, p value). Citation numbers refer to Supp. Table 7.
Shaded rows indicate species included in species richness estimates

Species	Family	Moran's i	p	Locus	N	Citation
Acanthaster planci	Acanthasteridae	0.876127	0.1541	CR	45	151, 153, 170
Acanthurus nigrofuscus	Acanthuridae	0.017369	0.6984	Cyt. B	17	99
Acanthurus nigroris	Acanthuridae	0.161608	0.02044	Cyt. B	15	132
Aipysurus laevis	Elapidae	0.201653	0.31744	ND4	9	51
Amphiprion ocellaris	Pomacentridae	0.651042	0.0424	CR	14	68
Apogon doederleini	Apogonidae	-1.04426	1	CR	3	124
Caesio cuning	Caesionidae	0.095228	0.56688	CR	23	167
Carijoa riisei	Clavulariidae	0.497831	0.01093	SRP54	6	19
Centropyge flavissima	Pomacanthidae	-0.31237	0.62834	Cyt. B	8	131
Cephalopholis argus	Serranidae	1.13484	0	Cyt. B	19	125, 164
Chaetodon citrinellus	Chaetodontidae	NA	NA	CR	3	149
Chaetodon lunulatus	Chaetodontidae	NA	NA	CR	3	149
Chaetodon meyeri	Chaetodontidae	-0.33205	0.99737	Cyt. B	4	165
Chaetodon ornatissimus	Chaetodontidae	0.875263	0.05725	Cyt. B	5	165
Chaetodon trifascialis	Chaetodontidae	NA	NA	CR	3	149
Chaetodon vagabundus	Chaetodontidae	NA	NA	CR	3	149
Chlorurus sordidus	Scaridae	-0.65247	0.21001	COI	8	13
Chromis atripectoralis	Pomacentridae	NA	NA	CR	3	149
Chromis viridis	Pomacentridae	NA	NA	CR	3	149
Chthamalus malayensis	Chthamalidae	0.011097	0.82782	COI	25	159
Halichoeres claudia	Labridae	-0.723	0.57243	COI	4	148
Halichoeres ornatissimus	Labridae	-0.43861	0.09182	COI	9	148
Halichoeres trimaculatus	Labridae	0.100853	0.27574	COI	6	148
Haliotis asinina	Haliotidae	0.274191	0.2255	CO2	6	39
Haptosquilla pulchella	Protosquillidae	-0.33538	0.71509	COI	7	12
Hippocampus kuda	Syngnathidae	-0.01247	1	CR	3	67
Holothuria (Microthele) nobilis	Holothuriidae	NA	NA	COI	12	70
Holothuria (Halodeima) atra	Holothuriidae	0.667712	0.00132	COI	18	174
Hyporthodus quernus	Serranidae	0.003179	0.72047	CR	9	173
Lethrinus miniatus	Lethrinidae	0.957524	0.15749	CR	4	142
Linckia laevigata	Ophidiasteridae	0.113953	0.49096	COI	21	7, 43, 150
Lutjanus erythropterus	Lutjanidae	0.165147	0.23889	CR	11	79
Lutjanus fulvus	Lutjanidae	0.591729	0.0028	Cyt. B	9	35, 164
Lutjanus kasmira	Lutjanidae	0.08565	0.46577	Cyt. B	10	35
Lutjanus sebae	Lutjanidae	NA	NA	CR	3	142
Myripristis berndti	Holocentridae	0.361117	0.02733	Cyt. B	17	20, 129
Naso brevirostris	Acanthuridae	NA	NA	CR	3	38
Naso unicornis	Acanthuridae	-0.58756	0.41059	CR	4	38
Nerita albicilla	Neritidae	1.143089	0.00018	COI	11	21, 169
Nerita plicata	Neritidae	0.175066	0.43718	COI	16	21, 169
Paracirrhites arcatus	Cirrhitidae	NA	NA	CR	3	149
Paracirrhites forsteri	Cirrhitidae	NA	NA	CR	3	149
Periclimenes soror	Palaemonidae	0.180522	0.31454	COI	9	7
Pinctada maxima	Pteriidae	0.60953	0.07691	COI	8	163
Plectropomus leopardus	Serranidae	-0.70791	0.21221	CR	5	74
Pomacentrus coelestis	Pomacentridae	0.002133	0.65012	CR	6	109, 124
Scarus psittacus	Scaridae	0.614544	0.07634	CR	6	82
Siganus fuscescens	Siganidae	0.259387	0.00004	CR	19	118
Siganus guttatus	Siganidae	0.054364	1	CR	3	113
Sphyraena barracuda	Sphyraenidae	NA	NA	Cyt. B	3	145
Thyca crystallina	Eulimidae	-0.11195	0.758	COI	18	7, 43
Triaenodon obesus	Carcharhinidae	-0.06399	0.70244	CR	7	166
Tridacna crocea	Cardiidae	0.313298	0.69583	COI	26	22, 44
Tridacna maxima	Cardiidae	0.059893	0.429	COI	7	58
Zebrasoma flavescens	Acanthuridae	-0.08423	0.85917	Cyt. B	19	128
Zebrasoma scopas	Acanthuridae	NA	NA	CR	3	149

Supplementary Table 3. Species used in expected heterozygosity analyses with number of populations sampled (N) and details of the spatial autocorrelation among heterozygosity within species (Moran's I, p value and Z score).
Citation reference numbers refer to Supp. Table 7. Shaded rows indicate species included in species richness estimates.

Species	Family	Moran's I	p	No. Loci	N	Citation
Acanthaster planci	Acanthasteridae	0.163152	0.371863	7	16	78
Acanthochromis polyacanthus	Pomacentridae	0.614313	0.077431	5	7	56
Acropora microphthalma	Acroporidae	NA	NA	7	3	160
Acropora papillare	Acroporidae	-0.393339	1	7	3	160
Acropora valida	Acroporidae	NA	NA	7	3	160
Aipysurus laevis	Elapidae	0.603452	0.035535	5	9	4
Amphiprion ocellaris	Pomacentridae	0.274608	0.241416	6	15	138
Epinephelus coioides	Epinephelidae	-0.259994	0.890468	4	6	8
Epinephelus polyphekadion	Epinephelidae	0.0419	0.268343	3	4	63
Gnatholepis anjerensis	Gobiidae	-0.320872	0.974953	7	4	106
Gnatholepis cauerensis	Gobiidae	-0.010623	0.760851	7	5	106
Hyporthodus quernus	Epinephelidae	0.231184	0.287161	10	9	172
Lethrinus miniatus	Lethrinidae	-0.512385	0.508102	9	6	104
Perna viridis	Mytilidae	0.049798	0.139905	5	6	119
Pinctada maxima	Pteriidae	0.301483	0.295265	6	8	50
Pocillopora damicornis	Pocilloporidae	-0.035526	0.98255	6	26	171
Pterapogon kauderni	Apogonidae	-0.00309	0.552616	11	6	111
Rachycentron canadum	Rachycentridae	-0.472652	0.561961	7	6	161
Scarus rubroviolaceus	Labridae	1.624136	0.00011	15	10	120
Seriatopora hystrix	Pocilloporidae	0.08255	0.636413	6	21	171
Tridacna crocea	Cardiidae	0.165023	0.042901	10	7	152
Zebrasoma flavescens	Acanthuridae	0.181742	0.12464	14	19	128

Supplementary Table 4. Parameter values (mean, median, upper and lower credible limits and standard deviation) for 61 species from the SGDC beta regression models predicting haplotype diversity. Shaded rows indicate species analysed in Messmer et al (2012).

	Species	Mean	Median	2.5%	97.5%	SD
1	*Acanthaster planci*	0.2034	0.2035	-0.0678	0.4756	0.1383
2	*Acanthurus nigrofuscus*	0.3957	0.3929	0.0462	0.7561	0.1814
3	*Acanthurus nigroris*	1.1880	1.1800	0.7278	1.6980	0.2473
4	*Aipysurus laevis*	-0.4102	-0.4079	-0.7621	-0.0704	0.1751
5	*Amphiprion ocellaris*	0.3142	0.3153	-0.4180	1.0270	0.3669
6	*Apogon doederleini*	0.0534	0.0577	-0.8034	0.8888	0.4290
7	*Caesio cuning*	0.5731	0.5786	-0.0860	1.1970	0.3265
8	*Centropyge flavissima*	1.3880	1.3750	0.6791	2.1680	0.3802
9	*Cephalopholis argus*	0.5083	0.5031	0.1479	0.8991	0.1904
10	*Chaetodon citrinellus*	0.5551	0.5541	-0.3012	1.4160	0.4334
11	*Chaetodon lunulatus*	0.5666	0.5652	-0.2392	1.3790	0.4116
12	*Chaetodon meyeri*	0.0976	0.1089	-0.8641	1.0120	0.4761
13	*Chaetodon ornatissimus*	0.2268	0.2204	-0.4245	0.9053	0.3378
14	*Chaetodon trifascialis*	0.3020	0.3043	-0.5129	1.1140	0.4143
15	*Chaetodon vagabundus*	0.5636	0.5603	-0.2789	1.4140	0.4292
16	*Chlorurus sordidus*	0.1908	0.1936	-0.4147	0.7866	0.3059
17	*Chromis atripectoralis*	0.8636	0.8601	-0.0054	1.7520	0.4469
18	*Chromis viridis*	0.7824	0.7808	-0.0789	1.6530	0.4394
19	*Chthamalus malayensis*	-0.0913	-0.0909	-0.4131	0.2281	0.1633
20	*Conus chaldaeus*	0.2304	0.2341	-0.6191	1.0450	0.4227
21	*Conus ebraeus*	0.0590	0.0585	-0.6344	0.7551	0.3522
22	*Gobiodon quinquestrigatus*	0.4253	0.4248	-0.4089	1.2630	0.4248
23	*Halichoeres claudia*	0.2057	0.2061	-0.5485	0.9633	0.3833
24	*Halichoeres ornatissimus*	0.5693	0.5662	-0.2707	1.4260	0.4291
25	*Halichoeres trimaculatus*	0.5981	0.5906	-0.0080	1.2420	0.3181
26	*Haliotis asinina*	-0.0710	-0.0660	-0.7769	0.6051	0.3515
27	*Haptosquilla pulchella*	0.1270	0.1328	-0.6050	0.8292	0.3632
28	*Hippocampus kuda*	-0.1786	-0.1842	-0.8737	0.5541	0.3612
29	*Holothuria Halodeima atra*	-0.4076	-0.4102	-0.7456	-0.0612	0.1747
30	*Holothuria Microthele nob*	0.1930	0.2015	-0.5811	0.9222	0.3820
31	*Hyporthodus quernus*	-0.1869	-0.1757	-0.9370	0.5041	0.3676
32	*Lethrinus miniatus*	0.9611	0.9543	0.2361	1.7120	0.3746
33	*Linckia laevigata*	0.7214	0.7218	0.2052	1.2350	0.2630
34	*Lutjanus erythropterus*	0.0973	0.0954	-0.3837	0.5889	0.2490
35	*Lutjanus fulvus*	-0.1275	-0.1282	-0.5492	0.2950	0.2148
36	*Lutjanus kasmira*	-0.2338	-0.2291	-0.8808	0.3831	0.3214
37	*Lutjanus sebae*	0.3988	0.3927	-0.4357	1.2610	0.4267
38	*Myripristis berndti*	0.9998	0.9979	0.5362	1.4680	0.2367
39	*Naso brevirostris*	0.0068	0.0065	-0.8878	0.8850	0.4497
40	*Naso unicornis*	0.0308	0.0316	-0.8774	0.9295	0.4578
41	*Naso vlamingii*	0.2435	0.2422	-0.5494	1.0370	0.4047
42	*Nerita albicilla*	0.3028	0.3044	-0.2561	0.8453	0.2808
43	*Nerita plicata*	0.2889	0.2928	-0.2266	0.7844	0.2574
44	*Paracirrhites arcatus*	0.2786	0.2795	-0.5495	1.1050	0.4197
45	*Paracirrhites forsteri*	0.3435	0.3469	-0.4969	1.1870	0.4257
46	*Periclimenes soror*	0.1900	0.1960	-0.5857	0.9248	0.3852
47	*Pinctada maxima*	0.4103	0.4089	0.0289	0.8029	0.1963
48	*Plectropomus leopardus*	0.0887	0.0952	-0.5487	0.6894	0.3147
49	*Pomacentrus coelestis*	0.0326	0.0347	-0.6465	0.6949	0.3410
50	*Scarus psittacus*	0.4997	0.4990	-0.2934	1.3000	0.4048
51	*Siganus fuscescens*	0.1918	0.2078	-0.3333	0.6270	0.2436
52	*Siganus guttatus*	0.2375	0.2368	-0.6929	1.1650	0.4711
53	*Sphyraena barracuda*	0.1285	0.1301	-0.6376	0.8855	0.3866
54	*Stegastes fasciolatus*	0.0015	0.0078	-0.8887	0.8538	0.4408
55	*Thyca crystallina*	0.1469	0.1523	-0.6821	0.9604	0.4167
56	*Triaenodon obesus*	0.4680	0.4648	-0.1017	1.0670	0.2979
57	*Tridacna crocea*	0.1617	0.1680	-0.4317	0.7161	0.2914
58	*Tridacna maxima*	0.3473	0.3485	-0.4181	1.1040	0.3857
59	*Zebrasoma flavescens*	0.2773	0.2721	-0.1681	0.7538	0.2355
60	*Zebrasoma scopas*	0.3596	0.3575	-0.4616	1.1900	0.4207
61	*Zebrasoma veliferum*	0.4673	0.4677	-0.3730	1.3140	0.4277

Supplementary Table 5. Parameter values (mean, median, upper and lower credible limits and standard deviation) for 57 species from the SGDC beta regression models predicting nucleotide diversity. Shaded rows indicate species analysed in Messmer et al (2012).

	Species	Mean	Median	2.5%	97.5%	SD
1	*Acanthaster planci*	0.1210	0.1206	0.0382	0.2053	0.0427
2	*Acanthurus nigrofuscus*	0.0908	0.0913	-0.2113	0.3860	0.1524
3	*Acanthurus nigroris*	0.4147	0.4036	-0.0171	0.9048	0.2343
4	*Aipysurus laevis*	0.0491	0.0438	-0.3368	0.4642	0.2019
5	*Amphiprion ocellaris*	0.3630	0.3619	0.1664	0.5655	0.1023
6	*Apogon doederleini*	0.1016	0.1000	-0.2672	0.4871	0.1907
7	*Caesio cuning*	0.2114	0.2081	0.0093	0.4271	0.1066
8	*Carijoa riisei*	0.2021	0.2022	-0.0010	0.3962	0.1021
9	*Centropyge flavissima*	0.2967	0.2861	-0.1180	0.7485	0.2206
10	*Cephalopholis argus*	0.3164	0.3165	0.0417	0.5857	0.1396
11	*Chaetodon citrinellus*	0.3906	0.3841	0.0689	0.7478	0.1728
12	*Chaetodon lunulatus*	0.3356	0.3306	0.0260	0.6872	0.1679
13	*Chaetodon meyeri*	0.1376	0.1355	-0.3468	0.6377	0.2476
14	*Chaetodon ornatissimus*	0.1134	0.1119	-0.2681	0.4985	0.1930
15	*Chaetodon trifascialis*	0.1510	0.1474	-0.2211	0.5407	0.1927
16	*Chaetodon vagabundus*	0.1943	0.1891	-0.1944	0.6131	0.2037
17	*Chlorurus sordidus*	0.0284	0.0272	-0.1567	0.2174	0.0954
18	*Chromis atripectoralis*	0.5198	0.5090	0.1320	0.9625	0.2133
19	*Chromis viridis*	0.3726	0.3650	0.0363	0.7558	0.1844
20	*Chthamalus malayensis*	0.0148	0.0138	-0.1574	0.1880	0.0885
21	*Halichoeres claudia*	0.1480	0.1416	-0.3161	0.6387	0.2384
22	*Halichoeres ornatissimus*	0.2253	0.2134	-0.2343	0.7394	0.2423
23	*Halichoeres trimaculatus*	0.1574	0.1545	-0.2609	0.5948	0.2175
24	*Haliotis asinina*	0.1078	0.1064	-0.3427	0.5689	0.2287
25	*Haptosquilla pulchella*	0.1642	0.1607	-0.2076	0.5615	0.1969
26	*Hippocampus kuda*	-0.0138	-0.0069	-0.4461	0.3898	0.2092
27	*Holothuria (Microthele)*	0.0696	0.0685	-0.3247	0.4684	0.2004
28	*Holothuria Halodeima a*	-0.0131	-0.0097	-0.2226	0.1836	0.1040
29	*Hyporthodus quernus*	-0.3513	-0.3523	-0.6511	-0.0460	0.1542
30	*Lethrinus miniatus*	0.4454	0.4384	0.0667	0.8686	0.2072
31	*Linckia laevigata*	0.0959	0.0942	-0.1119	0.3146	0.1090
32	*Lutjanus erythropterus*	0.0626	0.0639	-0.0876	0.2055	0.0745
33	*Lutjanus fulvus*	-0.0292	-0.0185	-0.4220	0.3147	0.1883
34	*Lutjanus kasmira*	0.0127	0.0176	-0.4182	0.4262	0.2138
35	*Lutjanus sebae*	0.1710	0.1679	-0.2481	0.6124	0.2187
36	*Myripristis berndti*	0.2475	0.2419	-0.1168	0.6470	0.1925
37	*Naso brevirostris*	-0.0825	-0.0798	-0.3401	0.1597	0.1275
38	*Naso unicornis*	0.0230	0.0262	-0.2111	0.2422	0.1160
39	*Nerita albicilla*	0.0923	0.0910	-0.1387	0.3284	0.1182
40	*Nerita plicata*	-0.1568	-0.1567	-0.3435	0.0331	0.0952
41	*Paracirrhites arcatus*	-0.0254	-0.0247	-0.3431	0.2945	0.1607
42	*Paracirrhites forsteri*	0.1794	0.1752	-0.1338	0.5161	0.1665
43	*Periclimenes soror*	0.0192	0.0195	-0.3863	0.4290	0.2055
44	*Pinctada maxima*	0.2142	0.2106	-0.1068	0.5504	0.1669
45	*Plectropomus leopardus*	0.0593	0.0554	-0.1879	0.3255	0.1305
46	*Pomacentrus coelestis*	-0.0003	0.0006	-0.3423	0.3384	0.1708
47	*Scarus psittacus*	0.1960	0.1908	-0.1996	0.6053	0.2042
48	*Siganus fuscescens*	0.0562	0.0500	-0.2473	0.4022	0.1644
49	*Siganus guttatus*	0.1326	0.1314	-0.3355	0.6232	0.2401
50	*Sphyraena barracuda*	0.1766	0.1700	-0.2764	0.6712	0.2386
51	*Thyca crystallina*	0.1674	0.1628	-0.2359	0.5999	0.2102
52	*Triaenodon obesus*	0.2036	0.1988	-0.2186	0.6603	0.2195
53	*Tridacna crocea*	0.1062	0.1024	-0.1951	0.4254	0.1557
54	*Tridacna maxima*	0.0814	0.0808	-0.3172	0.4827	0.2013
55	*Zebrasoma flavescens*	0.0194	0.0238	-0.3438	0.3501	0.1758
56	*Zebrasoma scopas*	-0.2851	-0.2850	-0.5990	0.0284	0.1595
57	*Zebrasoma veliferum*	0.1346	0.1303	-0.1450	0.4264	0.1458

Supplementary Table 6. Parameter values (mean, median, upper and lower credible limits and standard deviation) for 22 species from the SGDC beta regression models predicting expected heterozygosity from microsatellites with species richness.

	Species	**Mean**	**Median**	**2.5%**	**97.5%**	**SD**
1	*Acanthaster planci*	0.1082	0.1119	-0.0717	0.2721	0.0847
2	*Acanthochromis polyacanthus*	0.0957	0.1056	-0.2048	0.3451	0.1309
3	*Acropora microphthalma*	0.0559	0.0754	-0.2301	0.2512	0.1197
4	*Acropora papillare*	0.0599	0.0748	-0.1911	0.2352	0.1070
5	*Acropora valida*	0.1648	0.1475	-0.0621	0.4817	0.1304
6	*Aipysurus laevis*	0.0456	0.0503	-0.1073	0.1795	0.0745
7	*Amphiprion ocellaris*	0.1963	0.1829	0.0300	0.4256	0.1018
8	*Epinephelus coioides*	0.1327	0.1304	-0.0418	0.3215	0.0886
9	*Epinephelus polyphekadion*	0.1439	0.1353	-0.0815	0.4123	0.1187
10	*Gnatholepis anjerensis*	0.1179	0.1190	-0.1393	0.3669	0.1194
11	*Gnatholepis cauerensis*	0.1387	0.1337	-0.0510	0.3526	0.0978
12	*Hyporthodus quernus*	0.1108	0.1156	-0.1161	0.3150	0.1025
13	*Lethrinus miniatus*	0.1181	0.1197	-0.1445	0.3723	0.1220
14	*Perna viridis*	0.0683	0.0822	-0.1870	0.2608	0.1107
15	*Pinctada maxima*	0.1032	0.1057	-0.0447	0.2357	0.0702
16	*Pocillopora damicornis*	0.1778	0.1654	0.0058	0.4105	0.1003
17	*Pterapogon kauderni*	0.1242	0.1233	-0.1319	0.3831	0.1219
18	*Rachycentron canadum*	0.0858	0.0972	-0.1623	0.2837	0.1088
19	*Scarus rubroviolaceus*	0.2082	0.1947	0.0540	0.4192	0.0953
20	*Seriatopora hystrix*	0.1066	0.1109	-0.0811	0.2769	0.0878
21	*Tridacna crocea*	0.1574	0.1446	-0.0613	0.4488	0.1234
22	*Zebrasoma flavescens*	0.1527	0.1474	-0.0007	0.3389	0.0839

Supplementary Table 7. Full bibliography of studies from which genetic diversity data were extracted for use in this study. Ref. No. refers to study number in Supp. Tables 1-3.

Ref.	Citation
4	Lukoschek, V., M. Waycott, and J. S. Keogh. 2008. Relative information content of polymorphic microsatellites and mitochondrial DNA for inferring dispersal and population genetic structure in the olive sea snake, *Aipysurus laevis*. Molecular Ecology 17:3062-3077.
7	Crandall, E. D., M. E. Jones, M. M. Munoz, et al. 2008. Comparative phylogeography of two seastars and their ectosymbionts within the Coral Triangle. Molecular Ecology 17:5276-5290.
8	Antoro, S., U. Na-Nakorn, and W. Koedprang. 2006. Study of genetic diversity of orange-spotted grouper, *Epinephelus coioides*, from Thailand and Indonesia using microsatellite markers. Marine Biotechnology 8:17-26.
12	Barber, P. H., M. K. Moosa, and S. R. Palumbi. 2002. Rapid recovery of genetic populations on Krakatau: diversity of stomatopod temporal and spatial scales of marine larval dispersal. Proceedings of the Royal Society of London Series B-Biological Sciences 269:1591-1597.
13	Bay, L. K., J. H. Choat, L. van Herwerden et al. 2004. High genetic diversities and complex genetic structure in an Indo-Pacific tropical reef fish (*Chlorurus sordidus*): evidence of an unstable evolutionary past? Marine Biology 144:757-767.
19	Concepcion, G. T., S. E. Kahng, M. W. Crepeau et al. 2010. Resolving natural ranges and marine invasions in a globally distributed octocoral (genus *Carijoa*). Marine Ecology-Progress Series 401:113-127.
20	Craig, M. T., J. A. Eble, B. W. Bowen et al. 2007. High genetic connectivity across the Indian and Pacific Oceans in the reef fish *Myripristis berndti* (Holocentridae). Marine Ecology-Progress Series 334:245-254.
21	Crandall, E. D., M. A. Frey, R. K. Grosberg et al. 2008. Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. Molecular Ecology 17:611-626.
22	DeBoer, T. S., M. D. Subia, Ambariyanto, M. V. Erdmann et al. 2008. Phylogeography and Limited Genetic Connectivity in the Endangered Boring Giant Clam across the Coral Triangle. Conservation Biology 22:1255-1266.
31	Duda, T. F. and H. A. Lessios. 2009. Connectivity of populations within and between major biogeographic regions of the tropical Pacific in *Conus ebraeus*, a widespread marine gastropod. Coral Reefs 28:651-659.
35	Gaither, M. R., R. J. Toonen, D. R. Robertson et al. 2010. Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira* and *Lutjanus fulvus*). Journal of Biogeography 37:133-147.
38	Horne, J. B., L. van Herwerden, J. H. Choat et al. 2008. High population connectivity across the Indo-Pacific: Congruent lack of phylogeographic structure in three reef fish congeners. Molecular Phylogenetics and Evolution 49:629-638.
39	Imron, B. Jeffrey, P. Hale et al. 2007. Pleistocene isolation and recent gene flow in *Haliotis asinina*, an Indo-Pacific vetigastropod with limited dispersal capacity. Molecular Ecology 16:289-304.
41	Klanten, O. S., J. H. Choat, and L. van Herwerden. 2007. Extreme genetic diversity and temporal rather than spatial partitioning in a widely distributed coral reef fish. Marine Biology 150:659-670.

43 Kochzius, M., C. Seidel, J. Hauschild et al. 2009. Genetic population structures of the blue starfish *Linckia laevigata* and its gastropod ectoparasite *Thyca crystallina*. Marine Ecology-Progress Series 396:211-219.

44 Kochzius, M. and A. Nuryanto. 2008. Strong genetic population structure in the boring giant clam, *Tridacna crocea*, across the Indo-Malay Archipelago: implications related to evolutionary processes and connectivity. Molecular Ecology 17:3775-3787.

50 Lind, C. E., B. S. Evans, J. J. U. Taylor et al. 2007. Population genetics of a marine bivalve, *Pinctada maxima*, throughout the Indo-Australian Archipelago shows differentiation and decreased diversity at range limits. Molecular Ecology 16:5193-5203.

51 Lukoschek, V., M. Waycott, and H. Marsh. 2007. Phylogeography of the olive sea snake, *Aipysurus laevis* (Hydrophiinae) indicates Pleistocene range expansion around northern Australia but low contemporary gene flow. Molecular Ecology 16:3406-3422.

56 Miller-Sims, V. C., G. Gerlach, M. J. Kingsford et al. 2008. Dispersal in the spiny damselfish, *Acanthochromis polyacanthus*, a coral reef fish species without a larval pelagic stage. Molecular Ecology 17:5036-5048.

58 Nuryanto, A. and M. Kochzius. 2009. Highly restricted gene flow and deep evolutionary lineages in the giant clam *Tridacna maxima*. Coral Reefs 28:607-619.

63 Rhodes, K. L., R. I. Lewis, R. W. Chapman et al. 2003. Genetic structure of camouflage grouper, *Epinephelus polyphekadion* (Pisces : Serranidae), in the western central Pacific. Marine Biology 142:771-776.

67 Teske, P. R., H. Hamilton, P. J. Palsboll et al. 2005. Molecular evidence for long-distance colonization in an Indo-Pacific seahorse lineage. Marine Ecology-Progress Series 286:249-260.

68 Timm, J. and M. Kochzius. 2008. Geological history and oceanography of the Indo-Malay Archipelago shape the genetic population structure in the false clown anemonefish (*Amphiprion ocellaris*). Molecular Ecology 17:3999-4014.

70 Uthicke, S. and J. A. H. Benzie. 2003. Gene flow and population history in high dispersal marine invertebrates: mitochondrial DNA analysis of *Holothuria nobilis* (Echinodermata : Holothuroidea) populations from the Indo-Pacific. Molecular Ecology 12:2635-2648.

74 Van Herwerden, L., J. H. Choat, S. J. Newman et al. 2009. Complex patterns of population structure and recruitment of *Plectropomus leopardus* (Pisces: Epinephelidae) in the Indo-West Pacific: implications for fisheries management. Marine Biology 156:1595-1607.

78 Yasuda N, Nagai S, Hamaguchi M et al. 2009. Gene flow of Acanthaster planci (L.) in relation to ocean currents revealed by microsatellite analysis. Molecular Ecology 18(8):1574-1590.

79 Zhang JB, Cai ZP, Huang LM. 2006. Population genetic structure of crimson snapper *Lutjanus erythropterus* in East Asia, revealed by analysis of the mitochondrial control region. Ices Journal of Marine Science 63(4):693-704.

82 Winters, K. L., L. van Herwerden, J. H. Choat et al. 2010. Phylogeography of the Indo-Pacific parrotfish Scarus psittacus: isolation generates distinctive peripheral populations in two oceans. Marine Biology 157:1679-1691.

99 Eble, J. A., L. A. Rocha, M. T. Craig et al. 2011. Not All Larvae Stay Close to Home: Insights into Marine Population Connectivity with a Focus on the Brown Surgeonfish (*Acanthurus nigrofuscus*). Journal of Marine Biology 2011:1-12.

104 van Herwerden, L., J. Benzie, and C. Davies. 2003. Microsatellite variation and population genetic structure of the red throat emperor on the Great Barrier Reef. Journal of Fish Biology 62:987-999.

106 Thacker, C. E., A. R. Thompson, D. M. Roje et al. 2008. New expansions in old clades: population genetics and phylogeny of *Gnatholepis*

species (Teleostei: Gobioidei) in the Pacific. Marine Biology 153:375-385.

109 Liu, S.-Y. V., T. Kokita, and C.-F. Dai. 2008. Population genetic structure of the neon damselfish (*Pomacentrus coelestis*) in the northwestern Pacific Ocean. Marine Biology 154:745-753.

111 Hoffman, E. A., N. Kolm, A. Berglund et al. 2005. Genetic structure in the coral-reef-associated Banggai cardinalfish, *Pterapogon kauderni*. Molecular Ecology 14:1367-1375.

113 Iwamoto, K., A. Takemura, T. Yoshino et al. 2009. Molecular ecological study of *Siganus spinus* and *S. guttatus* from Okinawan waters based on mitochondrial DNA control region sequences. Journal of Oceanography 65:103-112.

114 Ramon, M. L., P. A. Nelson, E. Martini et al. 2008. Phylogeography, historical demography, and the role of post-settlement ecology in two Hawaiian damselfish species. Marine Biology 153:1207-1217.

118 Ravago-Gotanco, R. G. and M. A. Juinio-Menez. 2010. Phylogeography of the mottled spinefoot *Siganus fuscescens*: Pleistocene divergence and limited genetic connectivity across the Philippine archipelago. Molecular Ecology 19:4520-4534.

119 Prakoon, W., S. Tunkijjanukij, T. T. T. Nguyen et al. 2010. Spatial and Temporal Genetic Variation of Green Mussel, *Perna viridis* in the Gulf of Thailand and Implication for Aquaculture. Marine Biotechnology 12:506-515.

120 Fitzpatrick, J. M., D. B. Carlon, C. Lippe et al. 2011. The West Pacific diversity hotspot as a source or sink for new species? Population genetic insights from the Indo-Pacific parrotfish *Scarus rubroviolaceus*. Molecular Ecology 20:219-234.

124 Mirams, A. G. K., E. A. Treml, J. L. Shields et al. 2011. Vicariance and dispersal across an intermittent barrier: population genetic structure of marine animals across the Torres Strait land bridge. Coral Reefs 30:937-949.

125 Gaither, M. R., B. W. Bowen, T. R. Bordenave et al. 2011. Phylogeography of the reef fish *Cephalopholis argus* (Epinephelidae) indicates Pleistocene isolation across the indo-pacific barrier with contemporary overlap in the coral triangle. BMC Evolutionary Biology 11.

128 Eble, J. A., R. J. Toonen, L. Sorenson et al. 2011. Escaping paradise: larval export from Hawaii in an Indo-Pacific reef fish, the yellow tang *Zebrasoma flavescens*. Marine Ecology-Progress Series 428:245-258.

129 Muths, D., E. Tessier, G. Gouws et al. 2011. Restricted dispersal of the reef fish *Myripristis berndti* at the scale of the SW Indian Ocean. Marine Ecology-Progress Series 443:167-U180.

131 DiBattista, J. D., E. Waldrop, B. W. Bowen et al. 2012. Twisted sister species of pygmy angelfishes: discordance between taxonomy, coloration, and phylogenetics. Coral Reefs.

132 DiBattista, J. D., C. Wilcox, M. T. Craig et al. 2011. Phylogeography of the Pacific Blueline Surgeonfish, *Acanthurus nigroris*, Reveals High Genetic Connectivity and a Cryptic Endemic Species in the Hawaiian Archipelago. Journal of Marine Biology 2011:1-17.

138 Timm, J., S. Planes, and M. Kochzius. 2012. High similarity of genetic population structure in the false clown anemonefish (*Amphiprion ocellaris*) found in microsatellite and mitochondrial control region analysis. Conservation Genetics 13:693-706.

142 van Herwerden, L., W. J. Aspden, S. J. Newman et al. 2009. A comparison of the population genetics of *Lethrinus miniatus* and *Lutjanus sebae* from the east and west coasts of Australia: Evidence for panmixia and isolation. Fisheries Research 100:148-155.

145 Daly-Engel, T. S., J. E. Randall, and B. W. Bowen. 2012. Is the Great Barracuda (*Sphyraena barracuda*) a reef fish or a pelagic fish? The phylogeographic perspective. Marine Biology 159:975-985.

146 Duda, T. F., M. Terbio, G. Chen et al. 2012. Patterns of population structure and historical demography of *Conus* species in the tropical Pacific. American Malacological Bulletin 30:175-187.

148 Ludt, W. B., M. A. Bernal, B. W. Bowen et al. 2012. Living in the Past: Phylogeography and Population Histories of Indo-Pacific Wrasses (Genus *Halichoeres*) in Shallow Lagoons versus Outer Reef Slopes. Plos One 7:e38042.

149 Messmer, V., G. P. Jones, P. L. Munday et al. 2012. Concordance between Genetic and Species Diversity in Coral Reef Fishes across the Pacific Ocean Biodiversity Gradient. Evolution:no-no.

150 Otwoma LM. 2012. Genetic population structure of the coral reef starfish *Linckia laevigata* in the Indo-West Pacific. [Brussels]: Free University of Brussels. p. 73

151 Timmers MA, Bird CE, Skillings DJ et al. 2012. There's No Place Like Home: Crown-of-Thorns Outbreaks in the Central Pacific Are Regionally Derived and Independent Events. PloS One 7(2):e31159

152 Hui M. 2012. Connectivity and evolution of giant clams (Tridacnidae): a molecular genetic approach [PhD]. [Bremen]: University of Bremen.

153 Vogler C, Benzie J, Barber PH et al. 2012. Phylogeography of the Crown-of-Thorns Starfish in the Indian Ocean. PloS ONE 7(8):e43499

159 Tsang LM, Wu TH, Shih HT et al. 2012. Genetic and Morphological Differentiation of the Indo-West Pacific Intertidal Barnacle *Chthamalus malayensis*. Integrative and Comparative Biology 52(3):388-409

160 Richards ZT, van Oppen MJH. 2012. Rarity and genetic diversity in Indo-Pacific *Acropora* corals. Ecology and Evolution 2(8):1867-1888.

161 Phinchongsakuldit J, Chaipakdee P, Collins JF et al. 2013. Population genetics of cobia (*Rachycentron canadum*) in the Gulf of Thailand and Andaman Sea: fisheries management implications. Aquaculture International 21(1):197-217

163 Lind CE, Evans BS, Elphinstone MS et al. 2012. Phylogeography of a pearl oyster (*Pinctada maxima*) across the Indo-Australian Archipelago: evidence of strong regional structure and population expansions but no phylogenetic breaks. Biological Journal of the Linnean Society 107(3):632-646.

164 Gaither MR, Toonen RJ, Bowen BW. 2012. Coming out of the starting blocks: extended lag time rearranges genetic diversity in introduced marine fishes of Hawai'i. Proceedings of the Royal Society B-Biological Sciences 279(1744):3948-3957

165 DiBattista JD, Rocha LA, Craig MT et al. 2012a. Phylogeography of Two Closely Related Indo-Pacific Butterflyfishes Reveals Divergent Evolutionary Histories and Discordant Results from mtDNA and Microsatellites. Journal of Heredity 103(5):617-629

166 Whitney NM, Robbins WD, Schultz JK et al. 2012. Oceanic dispersal in a sedentary reef shark (*Triaenodon obesus*): genetic evidence for extensive connectivity without a pelagic larval stage. Journal of Biogeography 39(6):1144-1156

167 Ackiss AS, Pardede S, Crandall ED et al. 2013. Pronounced genetic structure in a highly mobile coral reef fish, *Caesio cuning*, in the Coral Triangle. Marine Ecology Progress Series 480:185-197

169 Crandall ED, Treml EA, Barber PH. 2012. Coalescent and biophysical models of stepping-stone gene flow in neritid snails. Molecular Ecology 21(22):5579-98.

170 Vogler C, Benzie JAH, Tenggardjaja K et al. 2013. Phylogeography of the crown-of-thorns starfish: genetic structure within the Pacific species. Coral Reefs 32(2):515-525.

171 Starger C, Barber P, Baker A. 2010. The recovery of coral genetic diversity in the Sunda Strait following the 1883 eruption of Krakatau. Coral

Reefs 29(3):547-565

172 Rivera, M. A. J., K. R. Andrews, D. R. Kobayashi et al. 2011. Genetic Analyses and Simulations of Larval Dispersal Reveal Distinct Populations and Directional Connectivity across the Range of the Hawaiian Grouper (*Epinephelus quernus*). Journal of Marine Biology 2011:1-11.

173 Rivera, M. A. J., C. D. Kelley, and G. K. Roderick. 2004. Subtle population genetic structure in the Hawaiian grouper, *Epinephelus quernus* (Serranidae) as revealed by mitochondrial DNA analyses. Biological Journal of the Linnean Society 81:449-468.

174 Skillings, D. J., C. E. Bird, and R. J. Toonen. 2011. Gateways to Hawai'i: Genetic Population Structure of the Tropical Sea Cucumber *Holothuria atra*. Journal of Marine Biology 2011:1-16.

CHAPTER 4

Strong signature of the Torres Strait landbridge on population structure in three codistributed species of *Tridacna* giant clams.

ABSTRACT

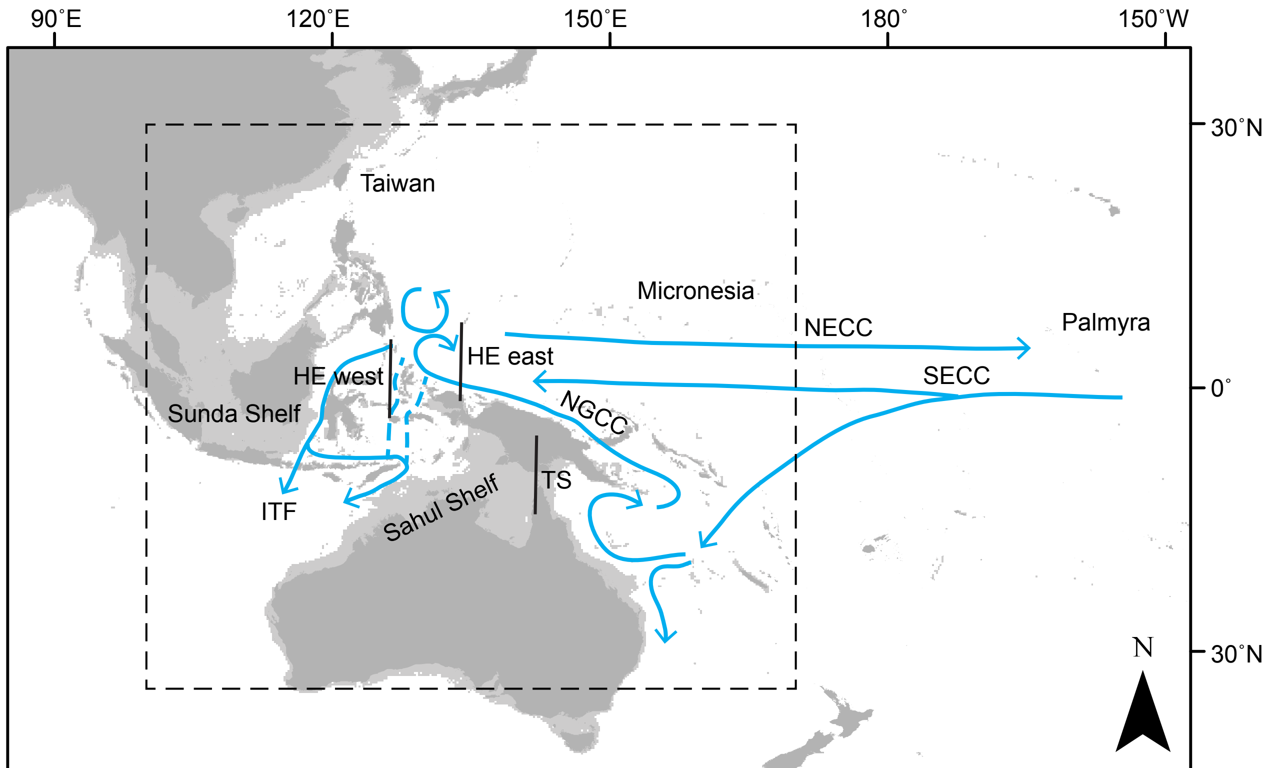
Making broad inference about processes underlying patterns of biodiversity requires data from multiple species. Molecular genetic data for many marine Indo-Pacific species exist, but sampling can be patchy and rarely fully represents a species range. Combining data sets collected for multiple studies into a single analysis can increase the value of each individual study and allow new insight into patterns among regions. The genus *Tridacna* has been the target of multiple genetic studies in the Indo-Pacific over the past four decades. Here we combine recently sampled COI sequences with existing data for three codistributed species (*Tridacna maxima, T. crocea* and a cryptic species, *T.* sp.) to unite sampling in the Coral Triangle biodiversity hotspot with samples from Micronesia, Egypt, Taiwan, Australia, Papua New Guinea and the Solomon Islands. For the first time, the signature of the Torres Strait can be seen to have strongly shaped population structure in two of these species. The persistence of a strong signal of historical barriers on genetic differentiation here argues against the assumption of *Tridacna* populations in the Indian Ocean, Coral Triangle, Micronesia and east Australia as open meta-populations.

INTRODUCTION

Defining disjunctions between populations can be troublesome in the marine environment, given the lack of obvious barriers and the presence of a highly dispersive larval phase in the lives of many marine species. In general, marine species tend to have larger range sizes than terrestrial species (Rapoport 1994) but, despite the apparent ease of dispersal in the marine environment, genetic studies of marine fauna often reveal substantial subdivisions within species that imply little gene flow between populations (for a review, see Hellberg 2009). Physical barriers to gene flow in the marine environment can range in porosity from relatively hard barriers such as land bridges (Knowlton 1993) through to softer barriers such as swift currents crossing between populations (e.g. the Eastern Australian current (Miller et al. 2013)), upwellings (e.g. the Benguela system of South Africa (Teske et al. 2011)) or river outflows (e.g. the Amazon-Orinoco (Lessios et al. 2003)).

The Coral Triangle, at the interface between the Indian and Pacific Oceans, has undergone intermittent periods of restricted flow due to sea level change, particularly during the Pleistocene Epoch (Voris 2000). The presence of large areas of continental shelf in the Coral Triangle has led to glacial cycles having a significant effect on habitat availability (Fig. 1). During the later part of the Pleistocene, sea levels fluctuated with glacial periods, reaching approximately 120 metres below

Figure 1. The Indo-Australian Archipelago and central Pacific; dominant currents are in blue, (adapted from Benzie and Williams 1997): NECC (North Equatorial Counter Current), SECC (South Equatorial Counter Current), NGCC (New Guinea Coastal Current), ITF (Indonesian Throughflow). Putative barriers to gene flow are in black lines: HE (Halmahera Eddy), SS (Sunda Strait), TS (Torres Strait). The light grey outline shows the 120 metre isobath: the approximate coastline during the lowest Pleistocene sea levels. The dashed square marks the extent of the biophysical model used to estimate larval dispersal.



present levels a number of times, most recently ~18,000 years ago (Voris 2000). The signature of population expansion in marine species following the re-flooding of shelf areas can be found in the patterns of genetic differentiation within species (Lukoschek et al. 2007; Crandall et al. 2012). Several regions of genetic discontinuity exist in the Coral Triangle that appear to be concordant across species (reviewed in (Carpenter et al. 2010; Briggs et al. 2013), discussed below.

Here we consider the influence of historical versus contemporary factors on the genetic structure within three species of *Tridacna* giant clams. For two of these species, we also consider the relative effect of three potential barriers to gene flow. *Tridacna maxima* and *T. crocea* have been sampled intensely in this region, from the Red Sea (Nuryanto and Kochzius 2009b) to the Central Pacific (Gardner et al. 2012), though significant gaps in sampling remain to be filled (Fig. 2 & 3). By guiding our sampling to complement these previous efforts we assemble a data set for three codistributed congeners across a very broad spatial extent. When attempting to use the patterns of genetic diversity within species to infer the influence of external forces, each species forms a single replicate. Choosing closely related species reduces the amount of variation between replicates due to factors not under consideration. Finding concordant phylogeographic patterns across

Figure 2. Sampling of *Tridacna maxima* for this study is shown in red and existing data are shown in black dots. The species range (light blue) is adapted from (Lucas 1988). Lighter grey outline indicates the 120m isobath: the approximate coastline during the lowest Pleistocene sea levels. Sampling location names can be found in Supp. Table 1, along with sources for existing data.

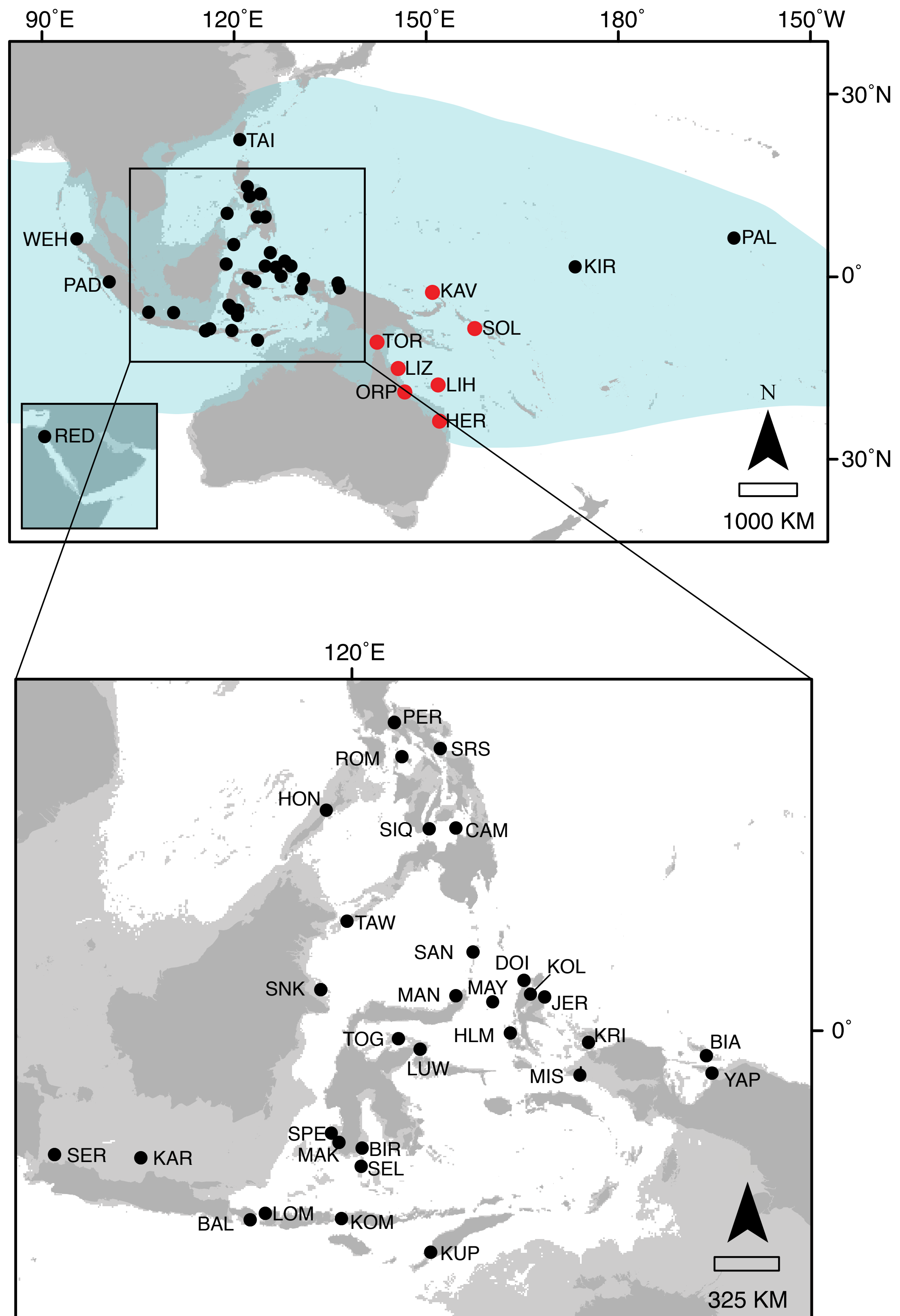
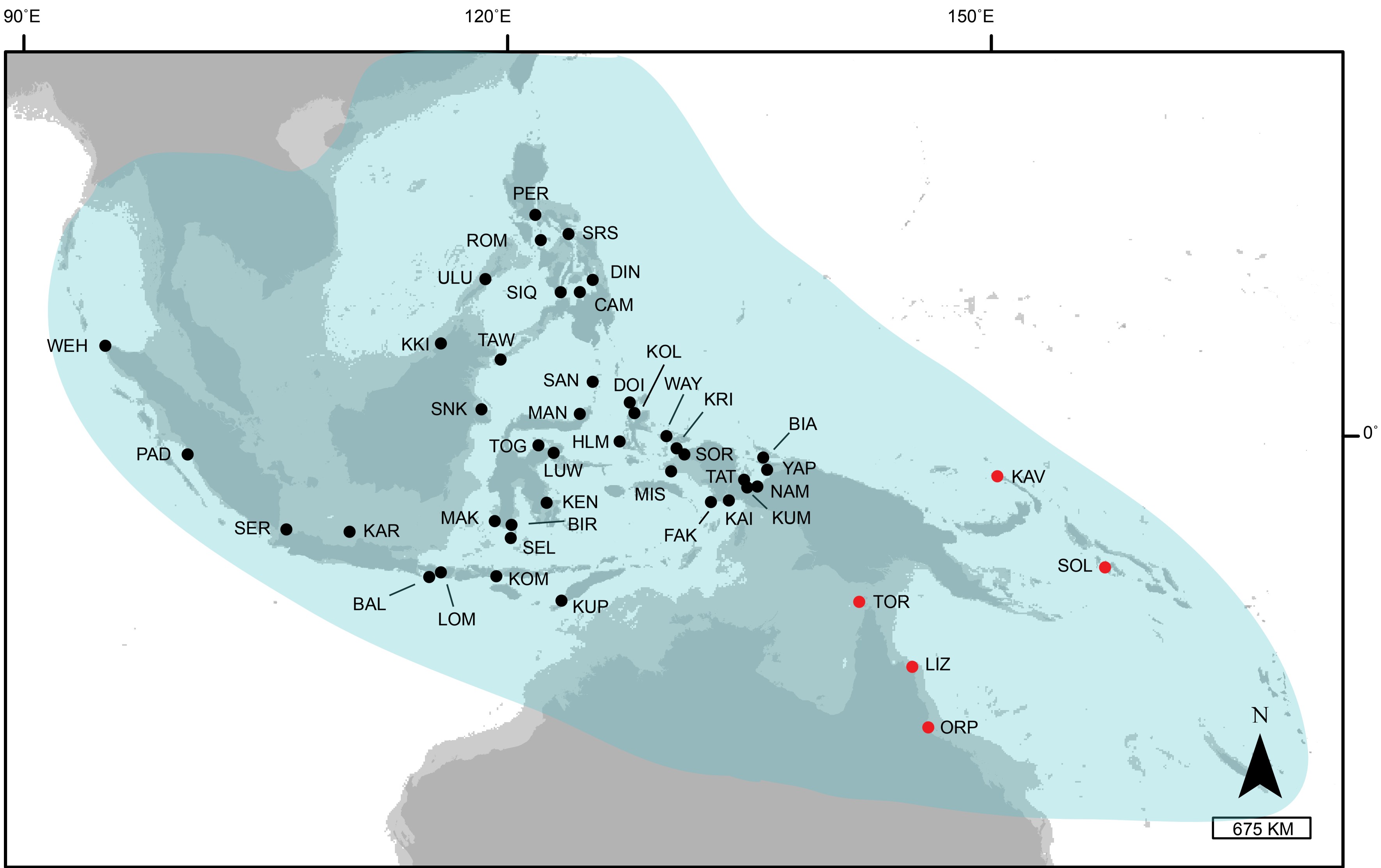


Figure 3. Sampling of *Tridacna crocea* for this study is shown in red and existing data are shown in black dots. The species range (light blue) is adapted from (Lucas 1988). Lighter grey outline indicates the 120 m. isobath: the approximate coastline during the lowest Pleistocene sea levels. Sampling location names can be found in Supplementary Table 1, along with sources for existing data.



codistributed species can suggest shared process (Dawson 2012). Finding non-concordant patterns in codistributed and closely related species leads to examination of the life history and ecological differences among species (Crandall et al. 2008a).

Species richness for the Tridacninae peaks in the Coral Triangle with eight of the 11 species coexisting in eastern Indonesia, making this region an excellent target for considering the factors influencing genetic structure. The most widespread species, *T. maxima*, ranges from the western Indian Ocean and Red Sea to the Pitcairn Islands in the central Pacific (Lucas 1988). *T. crocea* is the smallest extant species and has a range restricted to the central Indo-Pacific, stretching from Northern Australia to southern Japan and east to Vanuatu (IUCN Redlist Mollusc Specialist Group 1996). *Tridacna* species reproduce by broadcast spawning and possess planktonic larvae that are competent to settle within a week (Beckvar 1981; Fitt et al. 1984; Lucas 1988); hence successful fertilisation is heavily dependent on the density of spawning adults. Larvae can spend up to 19 days (Jameson 1976) drifting with the currents before settling to suitable hard substrate. This planktonic life stage, common in sessile and sedentary marine fauna, allows long distance dispersal to occur between populations, potentially connecting populations separated by habitat unfavourable for the adult phase. Below we describe potential barriers to the flow of these planktonic larvae through the region.

The Torres Strait landbridge

The Torres Strait, between northern Queensland, Australia, and Papua New Guinea (PNG), is a shallow strait between the Gulf of Carpenteria and the Coral Sea. This area formed a land bridge during lower sea levels of the Pleistocene and remained closed for approximately 80% of the time during the past quarter of a million years (Voris 2000). Previous work has discovered genetic divergence consistent with vicariant separation of populations either side of the Torres Strait landbridge in marine fishes (Elliott 1996; Begg et al. 1998; Chenoweth et al. 1998; Van Herwerden et al. 2009; Mirams 2011), molluscs (Imron et al. 2007), sea turtles (Dethmers et al. 2006) and sea snakes (Lukoschek et al. 2007) with a lack of structure found in some other species (Gopurenko et al. 1999; Uthicke and Benzie 2003; Duncan et al. 2006; Klanten et al. 2007; Dudgeon et al. 2009). The Torres Strait remained closed for much of the late Pleistocene, sea levels last rose enough to permit migration of marine species through the Torres Strait approximately 7,000 years ago (Reeves et al. 2008). Consequently, populations of giant clams either side of the Torres Strait would have been evolving in isolation from each other for over 200,000 years. Modelling of other marine species larval dispersal guided by modern oceanographic data indicates a low probability of gene flow across the Torres Strait (Treml et al. 2008) suggesting that this historically impermeable barrier is not as porous as we might expect today. Given the influence of the similarly aged Sunda Strait previously shown for *T. maxima* and *T. crocea* (DeBoer et al. 2008; Kochzius and Nuryanto 2008; Nuryanto and Kochzius 2009a; DeBoer et al. 2014) we expect to find a signature of the Torres Strait on genetic differentiation within populations of these species.

The Eastern and Western boundaries of the Halmahera Eddy

As the easterly New Guinea Coastal Current that flows along the northern shore of Papua reaches the island of Halmahera, it turns back towards the Pacific and becomes the source for the North Equatorial Counter Current (Fig. 1). This results in a seasonally strong eddy just above Halmahera and the Bird's Head Peninsula region of west Papua. Previous studies have shown the congruence of this eddy with genetic divergence between populations of damselfish (Liu et al. 2012), mantis shrimps (Barber et al. 2006; Crandall et al. 2008b), seastars (Crandall et al. 2008b) and of both *T. crocea* (DeBoer et al. 2008; Kochzius and Nuryanto 2008; DeBoer et al. 2014) and *T. maxima* (Nuryanto and Kochzius 2009a; DeBoer et al. 2014), although the position of the disjunction shifts slightly between species. For this reason, we test barriers at both the western and eastern boundary of the eddy (Schiller et al. 2008). Our western boundary falls where most of the abovementioned authors designate the Halmahera Eddy (Eastern Barrier in Barber et al. 2006), west of Halmahera and east of northern Sulawesi and the southern Philippines. This boundary has been shown to be an area of disjunction for *T. crocea* (DeBoer et al. 2008; Kochzius and Nuryanto 2008; DeBoer et al. 2014). Our eastern boundary extends north approximately 500km from the northern shore of the Bird's Head Peninsula, guided by the position and diameter of the Hamahera Eddy (Kashino et al. 1999). This eastern boundary concords approximately with the boundary to the west of

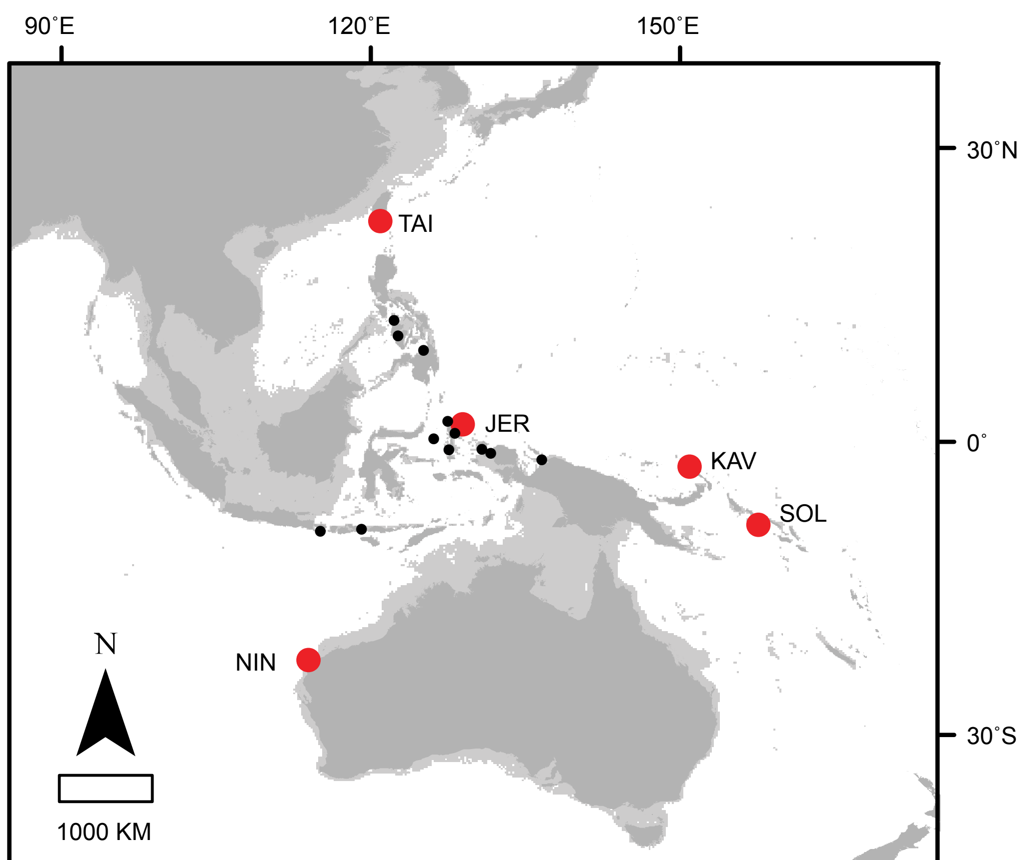


Figure 4. Sampling of *Tridacna sp*. for this study is shown in red and populations where fewer than 6 individuals were sampled are shown as small black dots. Lighter grey outline indicates the 120 m. isobath: the approximate coastline during the lowest Pleistocene sea levels. Sampling location names can be found in Supplementary Table 1, along with sources for existing data.

Cenderawasih Bay reported for *T. crocea, T. maxima* and many of the taxa listed above. We expect to recover the effect of the Halmahera Eddy on genetic differentiation in our analyses, but we do not expect its effect to rival that of the Torres Strait, given its seasonal nature.

There has been extensive effort expended on elucidating the patterns of phylogeographic structure within *Tridacna* by a number of research groups in Indo-Pacific, however most studies have focussed on one or a few regions within the species' ranges. Range-wide sampling is desirable to ensure that the patterns revealed through phylogeographic analyses are not artefacts of incomplete sampling. *T. crocea* and *T. maxima* have been densely sampled in the Coral Triangle region which has allowed researchers to pinpoint the locations representing genetic discontinuities (DeBoer et al. 2008; Kochzius and Nuryanto 2008; Nuryanto and Kochzius 2009; DeBoer et al. 2014). Recent work has gone some way towards linking this dense sampling to regions outside the Coral Triangle (Gardner et al. 2012; Huelsken et al. 2013) but eastern Australia, PNG and Solomon Island populations of *T. maxima* have not been considered alongside Taiwan and the central Pacific in previous phylogeographic studies. This study combines COI sequence data from these regions in a single analysis for the first time, allowing us to place the findings of earlier studies into the fullest context. While this single-marker analysis does have limited scope compared to multi-locus techniques, the ability to increase geographic coverage greatly by combining numerous existing data is valuable. To our knowledge, this is the first consideration of the effects of historical versus contemporary levels of isolation on genetic structure among populations of giant clams. It is also the first comparison of the influence of hard and soft barriers to gene flow within this species group. Our combination of novel COI sequence data with numerous existing data allows the simultaneous consideration of the relative effects of the Torres Strait barrier compared to the previously-investigated effects of the Halmahera Eddy.

METHODS

Sampling and DNA sequencing

Mitochondrial cytochrome oxidase I DNA sequences for three species of giant clam (*Tridacna maxima, T. crocea* and *T*. sp.) were obtained from previous studies in the region (DeBoer et al. 2008; Kochzius and Nuryanto 2008; Nuryanto and Kochzius 2009; Gardner et al. 2012; Huelsken et al. 2013; DeBoer et al. 2014; Su et al. 2014) and from recently sampled tissues. Tissue samples were collected from the mantle of living giant clams in situ using a non-lethal biopsy technique. New samples for this study were collected as allowed under scientific research permits from PNG, Australia and the Solomon Islands. We combine these data with published and unpublished sequences from the Red Sea, Indonesia, the Philippines, Taiwan, Kiribati and Palmyra to yield a data set spanning the majority of the species range for *T. maxima* and *T. crocea*, with the range for *T*. sp. still undefined.

DNA for new sequences was extracted using a Chelex protocol (Walsh et al. 1991) using 200µl Chelex 100 (20%) and 8µl Proteinase K (20mg/ml). Chelex extraction was followed with a CTAB protocol to remove PCR inhibitors (Huelsken et al. 2011). An approximately 485 basepair fragment of the COI barcoding gene was isolated using *Tridacna* – specific primers (TriCroF: 5' – GGGTGATAATTCGAACAGAA – 3' and TriCroR – 5'-TAGTTAAAGCCCCAGCTAAA – 3' (Kochzius and Nuryanto 2008) for *Tridacna crocea* or *T*. sp., and *Tridacna* 1F: 5'-ACCCTTTAYTTTTTATTAGCAY-3' (DeBoer et al. 2008) and *Tridacna* 3R: 5'-GTCATTGGCGATTACAGCATTG-3' or for *T. maxima* and *T*. sp. Some individuals of *T. maxima* and *T*.sp. required PCR with different primers due to poor amplification; for these we used Maxima F3: 5'-GTTTAGRGTRATAATYCGAACAG-3' (DeBoer et al. 2014) and universal vertebrate primers HCO2198 (5'- TAAACTTCAGGGTGACCAAAAAATCA-3') and LCO1490 (5'- GGTCAACAAATCATAAAGATATTGG-3' (Folmer et al. 1994). PCR thermocycler reactions followed a two step protocol with the two different annealing temperatures (50˚C for 20 cycles followed by 44˚C for the final set of 20 cycles.). Each cycle consisted of initial denaturation (95˚C) ran for 30 seconds, followed by annealing at either 50˚C or 44˚C, and a 45 second extension at 72˚C for Biotaq or 68˚C for Titanium taq. A final 5 minute extention stage was performed at the end of the 40 cycles. PCR reactions were made to a total volume of 25µl and contained 0.1 units of Biotaq (Bioline), 3mM of MgCl2, 0.8 mM of dNTPs, 0.3 M of each primer and 2.5µl template DNA. PCR product was prepared for capillary sequencing at Macrogen (Korea) with an Exo-Sap (New England Biolabs) protocol to remove unused primers and other small fragments of DNA.

Newly produced DNA sequences for each species were aligned and checked for ambiguities by eye in CodonCode Aligner and those with double peaks (characteristic of nuclear pseudo-genes) were discarded, before being aligned with published sequences of that species in SeAl (Rambaut 1996). Sequences were trimmed to the same length, and a custom Perl script was used to eliminate sequences containing greater than 5% bases called as ambiguous (R,S,W,Y etc) or un-defined (N) reducing the useable size of the gene fragment to 332 bp. for *T. crocea*, 385 bp. for *T. maxima* and 353bp. for *T*. sp. Sequences were checked for stop codons in Geneious using translation table 5 invertebrate mtDNA with the appropriate reading frame determined by comparison to Genbank sequences.

Genetic data analyses

Combination of data from multiple sources raises the possibility of samples from different lab groups coming from the same population. Populations were defined based on geographic proximity: if sampling sites were within 100 km. overwater distance from each other (following predictions Treml in al. (2012)), their samples were combined, provided pairwise ΦST was not significant between them. This applied to Solomon Islands [includes Roviana and Marovo Lagoons (Huelsken et al. 2013), this study]; Padang [includes Cubadak (DeBoer et al. 2008, 2014]; Pulau Seribu [includes Alam Kotok, Belat, Pramuka (DeBoer et al. 2014), Karang Congkak (DeBoer et al. 2014) and Semak Daun (DeBoer unpublished)]; Karimunjava [includes Java (DeBoer et al. 2014)], Bira [includes Pulau Sembilan (Kochzius and Nuryanto 2008, Nuryanto and Kochzius 2009)], Manado

[includes Lembeh Strait, Bunaken (DeBoer et al. 2014) and Bangka Batu (DeBoer unpublished)], Komodo [includes Flores (DeBoer et al. 2008, 2014) and Sebayur (DeBoer et al. 2014)], Bali [includes Nusa Penida (DeBoer et al. 2008, 2014)] and Yapen [includes Ambai and Serui (DeBoer et al. 2014)]. Locations with fewer than six individuals sampled were discarded from analyses to improve confidence in our estimates of diversity and differentiation statistics.

We obtained DNA sequences of *Tridacna* sp. from 17 sites in the Indo-Pacific (Fig. 4). Of these, five populations (Ningaloo Reef in West Australia, Tanjung Jerawai in eastern Indonesia, southern Taiwan, the Western Province of the Solomon Islands and Kavieng in PNG) were sampled in high enough numbers ($n \geq 6$) to be included in population-level analyses. Taiwanese samples of this cryptic species were collected as part of a phylogenetic study to describe the species ((Su et al. 2014).

Molecular diversity indices were calculated in Arlequin v3.5.1.3, including haplotype diversity (h), the likelihood of two haplotypes drawn from the population being the same, and nucleotide diversity (π), the average number of differences in DNA sequences between two individuals from the same population. Tajima's D was calculated to assess the conformity of Tridacna populations to neutral expectations. Genetic differentiation was measured using Φ_{ST}, based on Tamura-Nei distance, calculated with 10,100 permutations. F_{ST} was calculated based on haplotype identities.

The mitochondrial COI divergence between populations of giant clams was visualised using haplotype networks to show shared haplotypes among populations and the mutational steps separating haplotypes from each other. Median joining networks were created in Network 4.611 (Bandelt et al. 1999) using maximum parsimony to reduce the presence of non-parsimonious edges in the network. We activated the frequency >1 parameter which removes singleton haplotypes from the diagram to ease viewing. We checked these reduced networks against full median joining networks for singleton haplotypes whose deletion would alter the conclusions drawn, but found none. Epsilon value was set to 0 to reduce complexity introduced by reticulation in the network and the resulting network was manually arranged in Network Publisher 2.0 (fluxus-engineering.com). Colours in the networks define clades so that clade membership for sampling locations can be shown on maps in Fig. 5-7.

Predictors of genetic differentiation

Below we describe the historical and contemporary factors we tested for their influence on genetic differentiation (measured as F_{ST} and Φ_{ST}). These predictors were chosen to inform us whether divergence in *Tridacna* species has been most strongly influenced by historical or contemporary factors. First we describe our calculation of continuous overwater distances among populations, one historical and one contemporary. We then describe a model predicting contemporary larval dispersal for giant clams, likely to give the most contemporary estimate of dispersal distance. Finally, we describe our derivation of both hard (the Torres Strait) and soft (the Halmahera Eddy's east and west boundaries) barrier presence between populations.

Modern and last glacial maximum overwater distance

When dispersal is limited and in the absence of selection, gene flow between populations is expected to be proportional to the geographic distance (Wright 1943), so we expect a positive relationship between distance and the level of genetic differentiation if these neutral dynamics exist. To investigate the effects of geographic separation on populations of giant clams, we assessed the explanatory power of a number of measures of distance. The contemporary over-water distance separating each pair of sampling locations was calculated using a cost distance analysis in ESRI ArcMap 10 (Redlands, USA) where the least cost path between populations was forced around land. Overwater distance during the last glacial maximum of the Pleistocene Epoch (hereafter LGM overwater distance) was estimated using the same technique but with the land boundary shifted to the 120 metre isobath. Where present-day populations would have been dry land in this time period, we moved the population to the closest cell connected to the ocean, provided moving to this cell did not put the population on the other side of an existing landmass. Least cost paths were chosen over more sophisticated methods such as Circuitscape (Shah and McRae 2008) because we had little information on the porosity of the matrix through which migrants travel, other than whether it was land or sea, and these methods are equivalent in the absence of this information (Liggins et al. 2013). These three distance measures span a gradient from historical (LGM overwater distance), through recent (modern overwater distance) to contemporary (larval dispersal distance).

Biophysical model of larval dispersal

The probability of contemporary larval dispersal from one population to another was modelled using an Eulerian advection–diffusion approach (Treml et al. 2008). This model incorporated ocean current and wind patterns alongside information on the biology of species larvae to generate a probability of successful dispersal from one population to another. Simulated larvae were released from cells containing sampling locations in a matrix of 10 km grid squares and the probability of their successful transport into the next cell depended on the strength and direction of currents and the life history parameters we set (described below). Simulated larvae that escape the daily mortality rate and arrive in a cell with suitable habitat (coral or rocky reef) during the time that they are competent to settle will have successfully dispersed to that location. Best estimates from the literature state a maximum pelagic larval duration (PLD) of 19 days for *T. maxima* and 17 days for *T. crocea* (Jameson 1976) with larvae becoming competent to settle after approximately three to four days (Lucas 1988). We modelled *T.* sp. dispersal using the parameters for *T. maxima*, given the lack of data for the larval characteristics of *T.* sp. and its morphological and ecological similarity to *T. maxima* (pers. obs). The literature on seasonality of spawning of giant clams in the literature revealed no consistent pattern so we assumed spawning throughout the year. Larval mortality was modelled using a Weibold function parameterised on reported larval mortality for *T. crocea* in (Mies et al. 2012) and for *Tridacna* spp. in (Fitt et al. 1984). Reports of adult density of *T. maxima* ranged from 17.9 clams per square hectare in Papua New Guinea (Kinch 2002) to 77,050 in French Polynesia (Gilbert et al. 2006), making estimation of adult density difficult. Consequently, adult density was assumed to scale linearly with area of reef habitat in a cell. Simulated larvae were

released proportional to the area of reef habitat. The resulting pairwise matrix among sampling locations presents the corrected dispersal distance based on the probability of successful larval dispersal. This matrix is non-symmetrical because ocean currents and wind are directional. In order to compare the predictions of this larval dispersal model with symmetrical distances we needed to derive the maximum, minimum and mean dispersal distances from the biophysical model.

Given that the spatial extent of the biophysical model (dashed square in Fig. 1) encompasses all three barriers, and that only four *T. maxima* and one *T. crocea* sampling location lay outside this, we conducted our analyses at the scale of this model. Sampling locations outside of this extent: the Red Sea, Palmyra, Kiribati and Aceh, Indonesia, were excluded from analyses. Due to the strong population structure within *T. crocea* and *T. maxima* (Fig. 6 and 7), we compared the effects of the continuous predictors (modern and LGM overwater distances and the predictions of the larval dispersal model) within the two clades sampled at the most locations: clade 2 of *T. crocea* and clade 3 of *T. maxima*.

Putative barriers to gene flow

Three putative barriers or locations of genetic discontinuities identified in previous studies of the region were tested for their effects on genetic divergence within *Tridacna maxima, T. crocea* and *T.* sp. The closure of the Torres Strait which accompanied lower sea levels at the last glacial maximum was considered alongside two 'softer' barriers: the western and eastern limits of the Halmahera Eddy, the locations of genetic discontinuity identified in other marine invertebrates. The presence or absence of a barrier separating populations was determined by whether this barrier intersected a least cost path between each pair of populations. Least cost paths were chosen over more sophisticated methods such as Circuitscape (Shah and McRae 2008) because we had little information on the porosity of the matrix through which migrants travel other than whether it was land or sea; these methods are comparable when this information is missing (Liggins et al. 2013). Least cost paths between all pairs of populations were calculated in ArcMap 10 (ESRI, Redlands, USA). Least cost paths were constrained to travel over water only, using the modern coastline. Barrier presence between a pair of populations was coded as 1 and absence as 0 in a separate triangular matrix for each barrier. The effects of barriers could only be assessed for *T. maxima* and *T. crocea* due to a paucity of sampling either side of these barriers for *T*. sp.

Statistical models

The relative effects of contemporary and LGM overwater distance and the probability of larval dispersal between populations on the level of genetic divergence were assessed using distance-based redundancy analysis (dbRDA) using the *capscale* function in the R package *vegan* (Oksanen et al. 2007). Multiple, correlated predictors can be combined in a single analysis by calculating the principle coordinates of each neighbourhood (pairwise distance) matrix (PCNM) and using the significant axes of these as predictor variables. Further ordination in the dbRDA reduces the effects of collinearity. This is in contrast to multiple regression on distance matrices (MRDM, Legendre et al. 1994) where correlations among predictors preclude their co-analysis. DbRDA allows the

effects of each predictor variable to be partialled out from each other (Balkenhol et al. 2009). This ability to test multiple barriers and continuous distances in a single analysis strengthens the field of phylogeography in that it allows rigorous hypothesis testing rather than *post-hoc* correlative approaches (Legendre and Fortin 2010). Modern overwater distance, LGM overwater distance and the maximum, minimum and mean larval dispersal probabilities were used to predict Φ_{ST} and F_{ST}, along with Rousset's transformation of these statistics against log transformed overwater distances. F_{ST} / (1 - F_{ST}) is expected to show a linear relationship with the logarithm of geographic distance in two-dimensional habitats (Rousset 1997).

We ran separate dbRDA analyses for each species. Firstly, due to the non-symmetrical nature of the larval dispersal predictions we derived the minimum, maximum and mean dispersal distances between each pair of populations as predicted by the biophysical model. Before comparing these distances to the other predictors we conducted a dbRDA predicting Φ_{ST} with the minimum, maximum and mean values of larval dispersal distance and performed backward stepwise model reduction with p-to-remove set to 0.05 to determine the best representative of these variables. Models were compared based on the Akaike Information Criterion and on the individual p-values associated with each variable, as calculated by the *ordistep* method in *vegan* (Oksanen et al. 2007).

Prior to analysis, all predictor variables were checked for collinearity. Strong correlations between the continuous overwater distances (modern and LGM) and among the results of the larval dispersal model were found. For *T. crocea*, the effect of the Torres Strait barrier also correlated with LGM overwater distance. Following PCNM transformation, however, these correlations all fell to below 0.8 except for a correlation of 0.89 between the mean and maximum dispersal probabilities from the larval dispersal model.

We tested the individual effect of explanatory variables on Φ_{ST} by running univariate dbRDA analyses for the three species (Table 1) and within clade 2 of *T. crocea* and clade 3 of *T. maxima* (Table 2). Only those predictors returning significant relationships with Φ_{ST} were used in models to compare the relative effects of multiple predictors. We found the best combination of modern overwater distance, LGM overwater distance and the larval dispersal distance by setting up models with these three continuous predictors and performing backward stepwise model selection. We then compared the relative effects of the remaining variables by testing the marginal significance of each variable in the analysis (Table 3). This tests the effect of each variable after the other variables have been fitted (Oksanen et al. 2007).

To identify the relative strength of the Torres Strait, east Halmahera Eddy and west Halmahera Eddy we conditioned dbRDA models with each continuous distance variable in turn before fitting the barriers and testing the marginal significance of each. This tests for variation explained by the barriers above that explained by overwater distance or larval dispersal distance (Table 4).

RESULTS

Sampling and DNA sequencing

114 novel DNA sequences of the COI barcoding gene were produced for *Tridacna maxima*, 78 for *T. crocea* and 19 for *T.* sp. for combination with previously published sequence data. Additional unpublished sequences for *T.* sp. were obtained from Dr. Timery DeBoer. The combined data set comprises the most extensive survey to date of the phylogeography within these three species (see Supplementary Tables 1 a-c for details of data sources). We assembled 614 sequences for *T. maxima* containing 283 unique haplotypes across 43 populations, 721 for *T. crocea* with 311 unique haplotypes from 46 populations and 58 for *T.* sp. with 28 unique haplotypes from five populations.

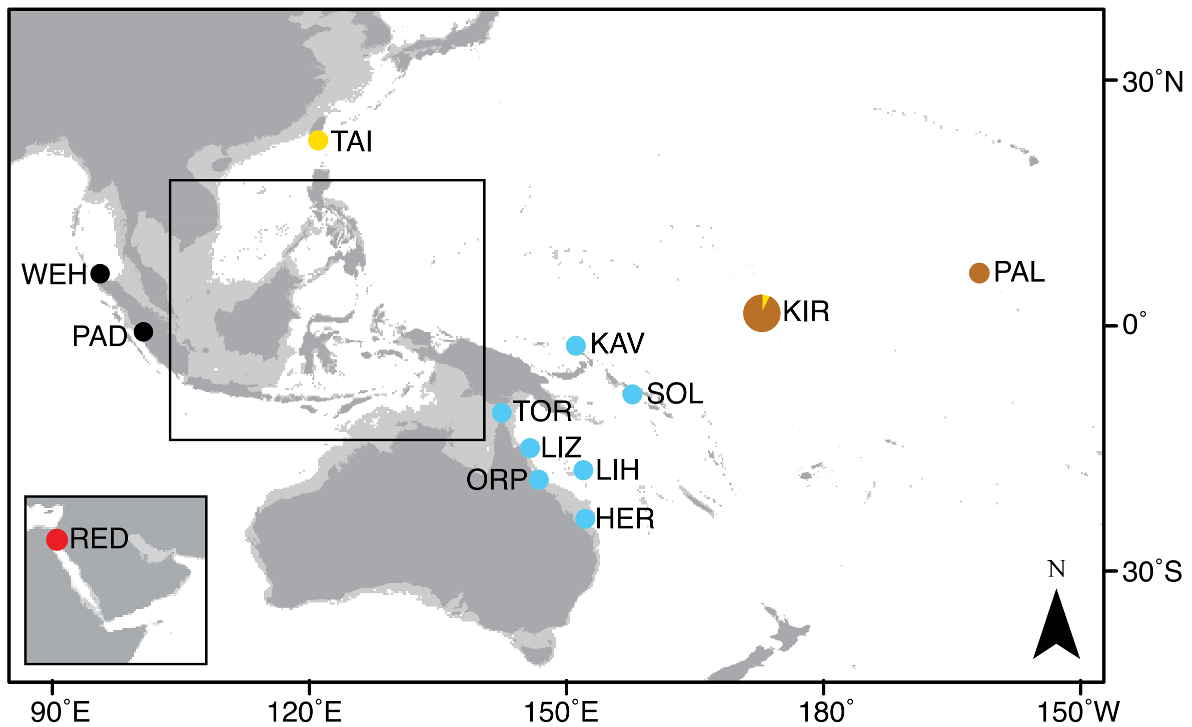
Geography of haplotypes

Unsurprisingly, we recovered the same deep divergence across the Sunda Strait and in the region of eastern Indonesia reported in previous studies for *T. maxima* (Nuryanto and Kochzius 2009a; DeBoer et al. 2014) and *T. crocea* (DeBoer et al. 2008a; Kochzius and Nuryanto 2008; DeBoer et al. 2014), meaning our treatment of these data did not alter previous results.

Figure 5 shows the frequencies and geographic distribution of haplotypes within *T. maxima*. The sampled populations are represented by five clades separated by at least six mutational steps from each other. To the west, the Red Sea population is populated solely by clade 1 (red in Fig. 5) and shares no haplotypes with any other sampled population. Palmyra and Kiribati in the east are mostly populated by members of a divergent clade (clade 5, brown in Fig. 5), differentiated by 23 steps from other clades, but Kiribati is home to two individuals with a clade 3 (yellow) haplotype common in the Coral Triangle that also occurs in Taiwan. West Pacific sites of *T. maxima* are solely populated by members of a second distinct clade (clade 4, blue in Fig. 5), separated by 20 mutational steps. West Pacific populations share no haplotypes with Kiribati or Palmyra. East Indonesia contains seven haplotypes belonging to clade 4, haplotypes occur as far west as Jerawai as well as in Cenderawasih Bay but are not present in Misool or Kri. Indian Ocean populations only contain members of clade 2, separated from the rest of the network by six steps, but this clade also extends east through Java and the central Coral Triangle to east Indonesia and Cenderawasih Bay. Clade 2 also occurs in low densities in central Indonesian populations and as far east as Cenderawasih Bay. The most common haplotype in the central clade (clade 3) occurs throughout the range of *T. maxima* with the exception of the West Pacific and Red Sea.

T. crocea haplotypes (Fig. 6) fall into four main clades, a central, diverse clade (clade 2, yellow) and three satellite clades separated by at least nine mutations. Populations in Indonesia, Malaysia and the Philippines contain members of clade 2, which links the Indian Ocean clade to clades occurring further to the east. Indian Ocean populations contain only clade 1 (also clade 1 in Kochzius and Nuryanto (2008), white clade in DeBoer et al. (2008)), haplotypes of which do not occur in any populations further east. Similarly, West Pacific populations (blue in Huelsken et al.

Figure 5a. Map showing *Tridacna maxima* sampling locations, coloured by clade identity with reference to (5b). Lighter grey outline indicates the 120m isobath: the approximate coastline during the lowest Pleistocene sea levels.



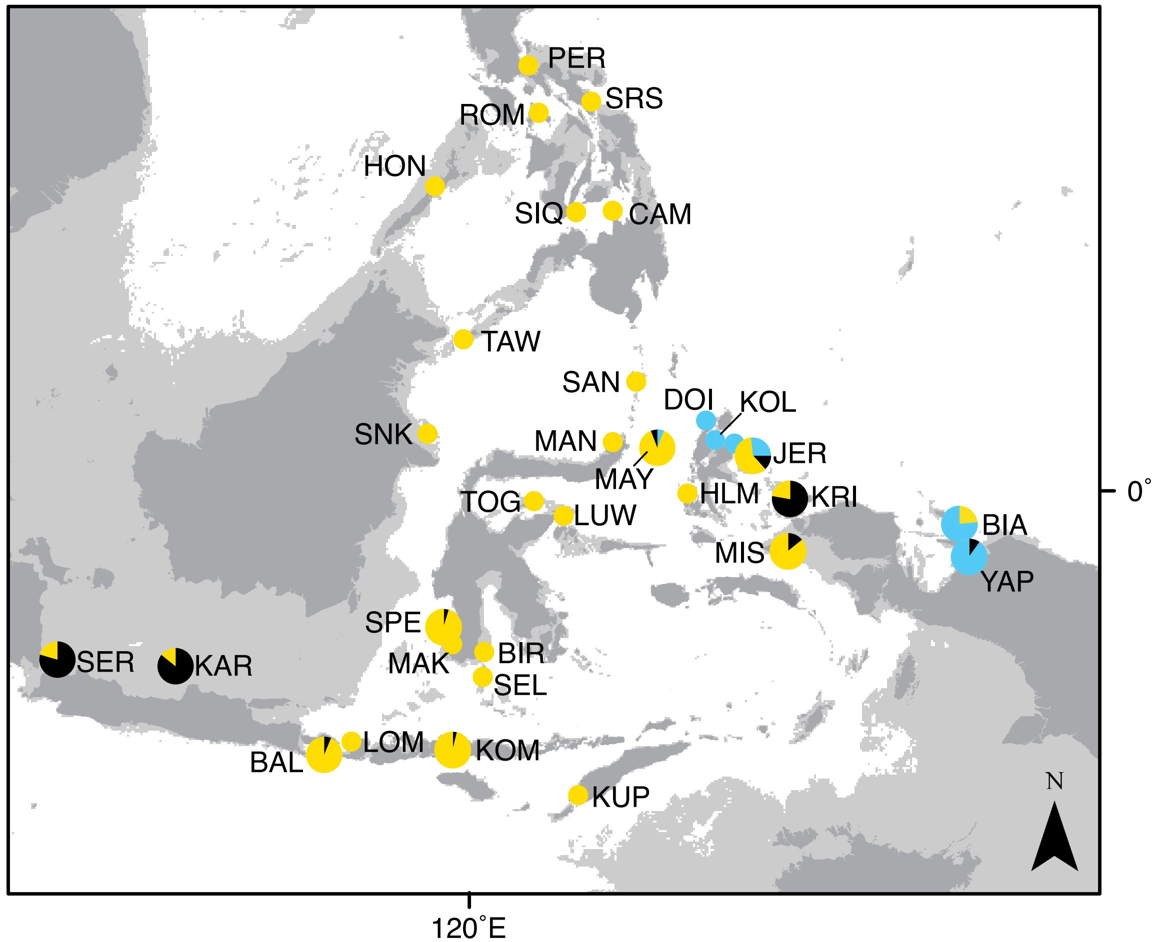
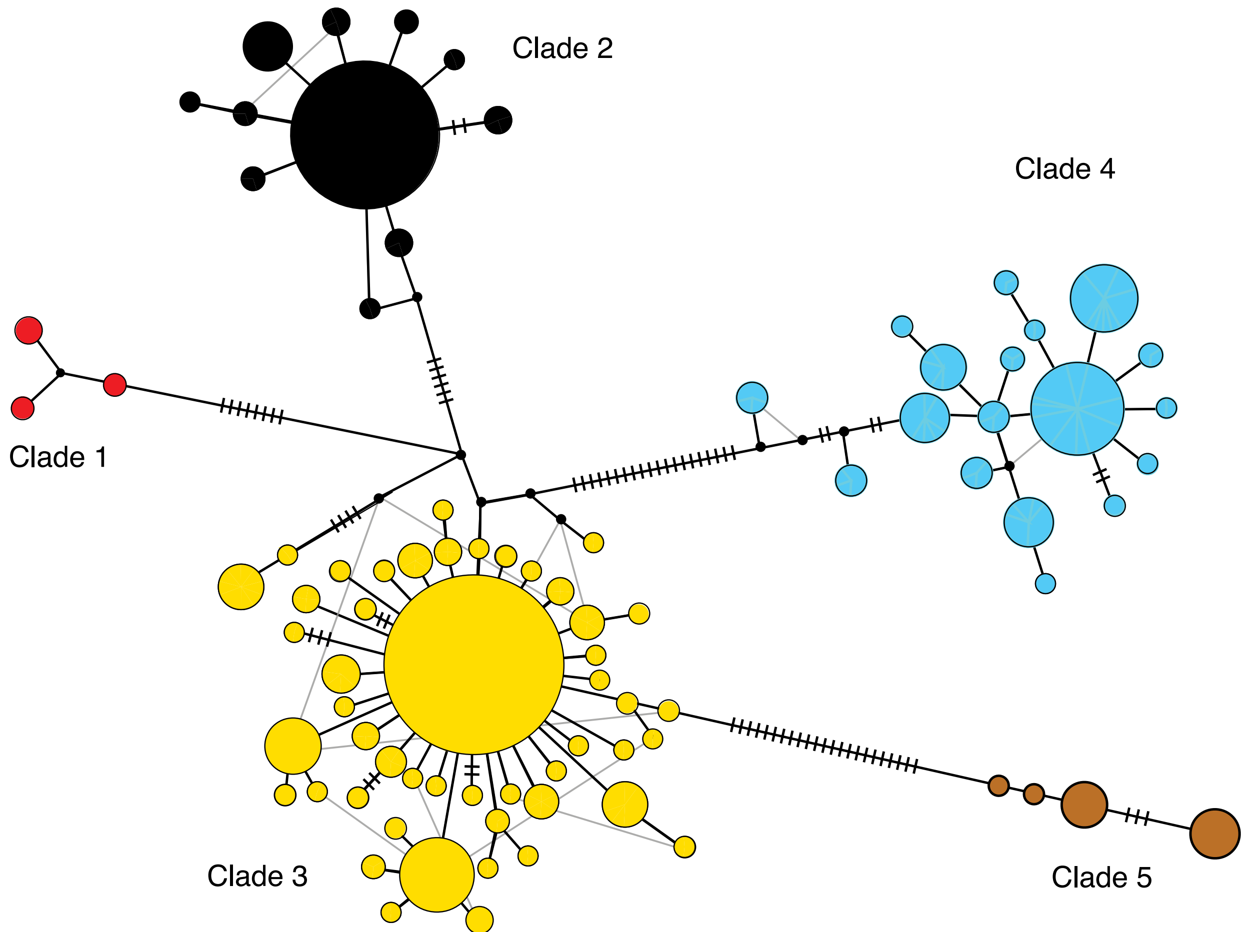


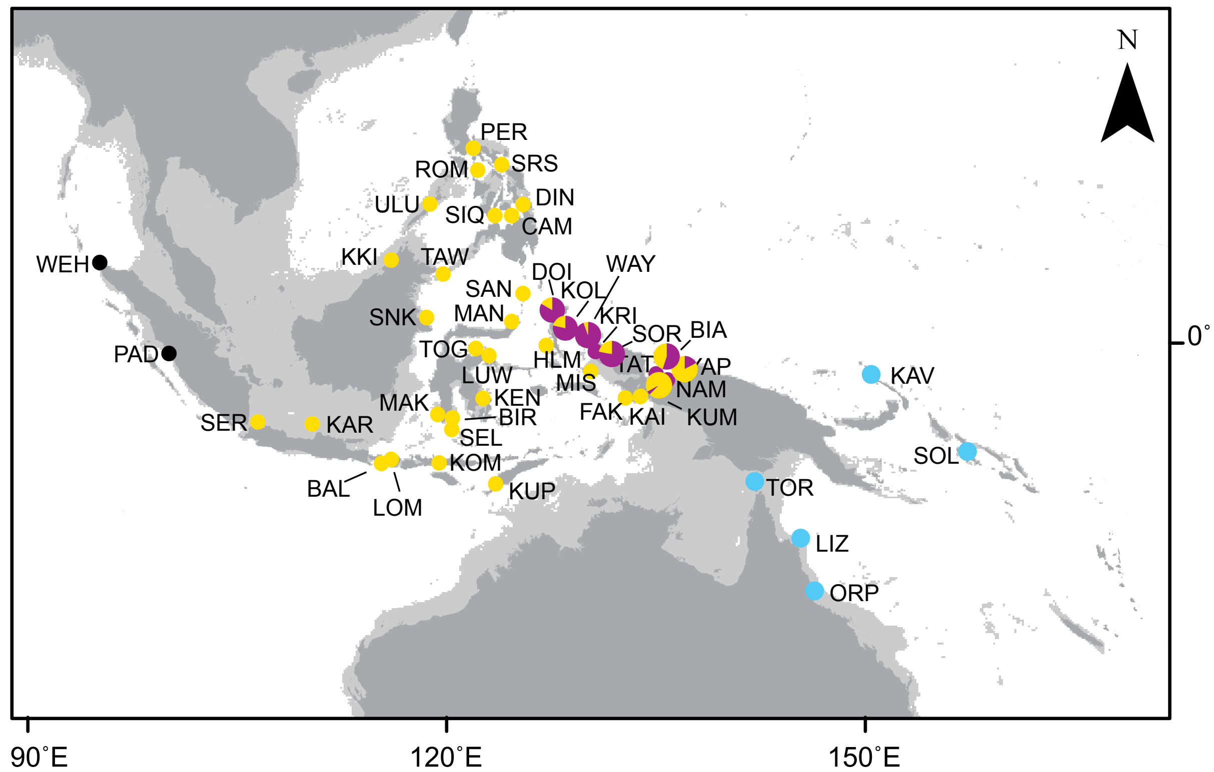
Figure 5b. Median joining haplotype network from a 385bp fragment of COI for *T. maxima*. Nodes represent unique haplotypes, sized according to their frequency. Singleton haplotypes were removed to ease viewing. Perpendicular marks on network edges represent single nucleotide polymorphisms separating haplotypes from each other.



(2013)) are exclusively populated by members of Clade 4, separated from the central clade 2 by ten mutational steps. clade 3 (clade 5 in Kochzius and Nuryanto (2008), grey in DeBoer et al. (2008) haplotypes are only found in eastern Indonesia (green in Huelsken et al. (2013)). Populations in the southern part of Cenderawasih Bay are populated exclusively by clade 3 but populations in northern Halmahera, the north shore of the Bird's Head Peninsula and northern Cenderawasih contain haplotypes from clades 2 and 3.

Within the five populations for which we performed population level analyses of *T.* sp. we found two distinct clades separated by seven polymorphic positions. The Ningaloo population from west Australia shares haplotypes with populations in Taiwan and Tanjung Jerawai in eastern Indonesia. West Pacific populations in PNG and the Solomon Islands share haplotypes with Tanjung Jerawai in eastern Indonesia (Fig. 7).

Figure 6. Map showing sampling locations for *T. crocea*, coloured by clade identity of individuals collected there. Lighter grey outline is the coastline during the last glacial maximum. Below, median joining haplotype network from a 332 bp fragment of COI. Unique haplotypes are represented by nodes, the size of which is proportional to haplotype frequency. Singleton haplotypes were removed to ease viewing. Perpendicular marks on network edges indicate the number of mutational steps separating haplotypes.



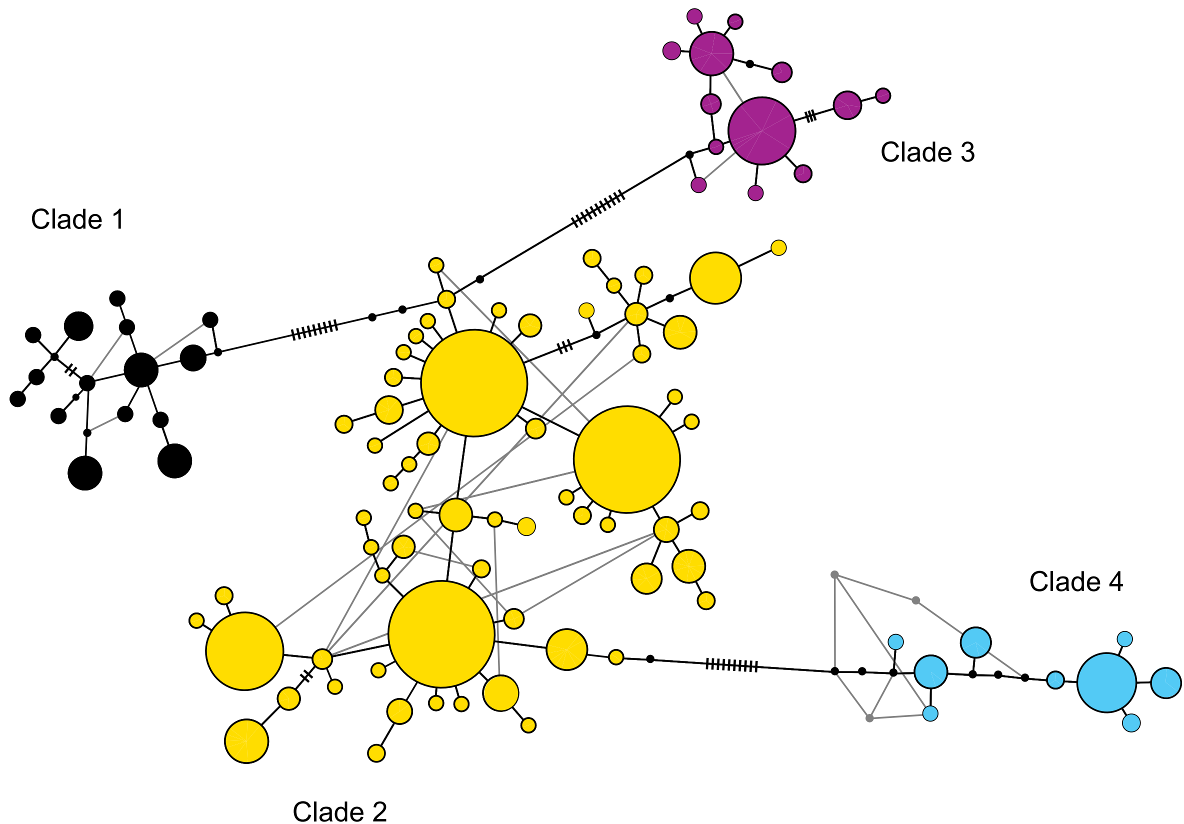


Figure 7. Map shows sampling locations for *Tridacna* sp., coloured to denote clade identity of individuals found at that location. Lighter grey outline is the coastline during the last glacial maximum. Below, median joining haplotype network from a 353 bp fragment of COI for *T. sp*. Unique haplotypes are represented by nodes, the size of which is proportional to haplotype frequency. Perpendicular marks on network edges indicate the number of mutational steps separating haplotypes.

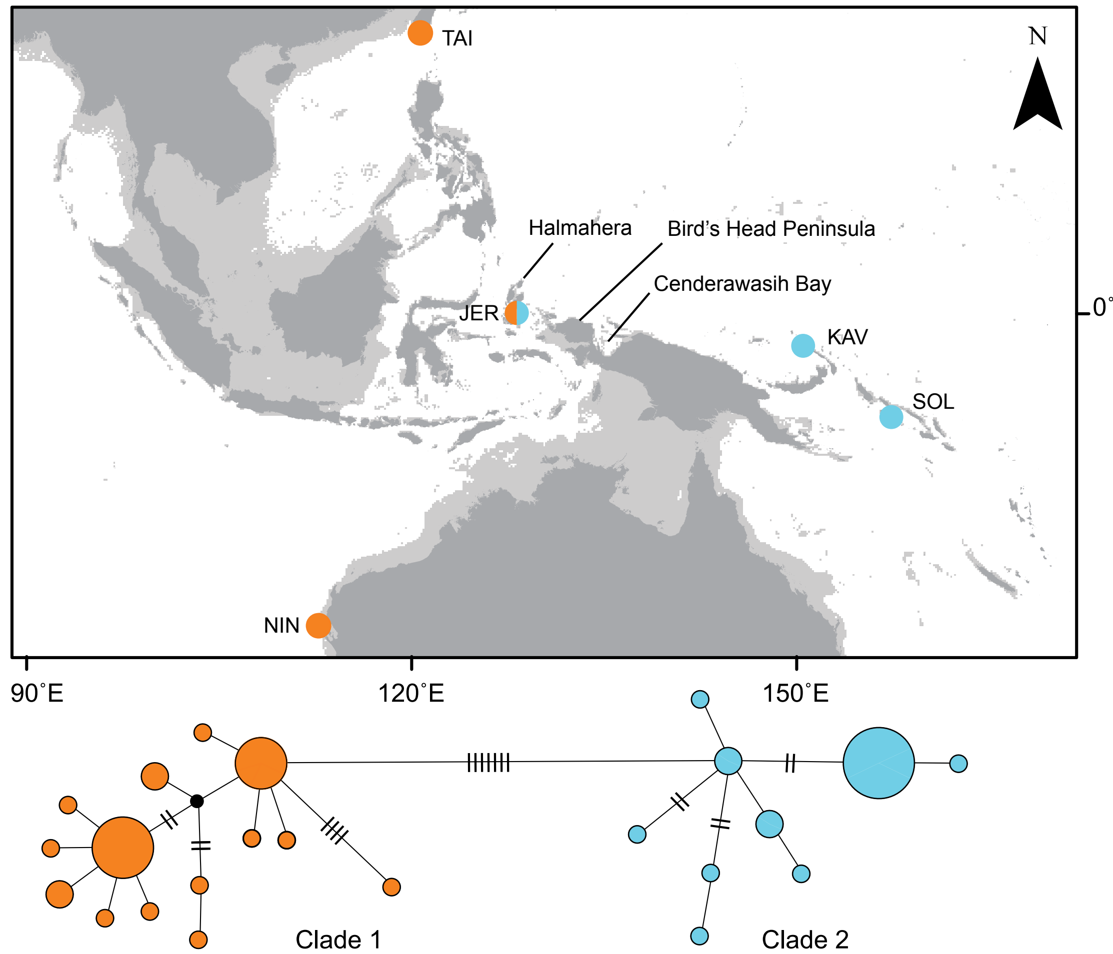


Table 1. Marginal tests of the effects of continuous predictors within *T. crocea*, *T. maxima* and *T*. sp. in the context of other, competing predictors. Degrees of freedom indicate the number of significant axes in the PCNM of the distance matrix. Marg. Var. is the variance in Φ_{ST} explained by the predictor after all others have been fitted.

T. crocea		df	Marg. Var	Pseudo-F	Prob.	
Larval Dispersal	Mean	5	0.3281	1.6863	0.096	
Overwater Distance	LGM	5	0.5975	3.071	0.002	*
	Modern	6	0.53791	2.3039	0.010	*

T. maxima		df	Marg. Var	Pseudo-F	Prob.	
Larval Dispersal	Mean	3	0.51412	2.6561	0.044	*
Overwater Distance	LGM	3	0.14096	0.7282	0.610	
	Modern	3	0.28031	1.4482	0.232	

***T.* sp.**		df	Marg. Var	Pseudo-F	Prob.	
Larval Dispersal	Mean	1	0.34124	9.6273	0.033	*
	Max	1	0.39341	11.0992	0.017	*

Performance of F_{ST} versus Φ_{ST}

Pairwise F_{ST} did not result in any significant relationships with any of the predictor variables ($p > 0.1$ in all cases, results not shown). Genetic distance measured as Φ_{ST} showed a much clearer response to overwater distance, both contemporary and historical, so hereafter we focus on results concerning Φ_{ST}. Pairwise population F_{ST} and Φ_{ST} values for all three species were transformed according to Rousset (Rousset 1997) but the results of dbRDA models predicting these transformed values with log-transformed LGM and modern overwater distances were very similar to those using untransformed data (results not shown) so, to reduce complexity and allow simple combination with the larval dispersal model, these transformations were not pursued further.

Comparisons among continuous distances

For *T. crocea*, the mean larval dispersal distance was chosen over maximum or minimum dispersal distances by backward model selection and was used for further analyses. Stepwise model selection found that a model containing both modern and LGM overwater distances was the best predictor of Φ_{ST} in *T. crocea*, mean larval dispersal distance was dropped from this model (Table 2). This simplified model explained 79% of the variation ($F_{11,32} = 10.024$, $p = 0.001$) in Φ_{ST}. Tests of marginal significance showed that both modern and LGM overwater distances explain some variation above the other, neither could be removed from the model.

Table 2. Marginal tests on best models of the effects of continuous distances on Φ_{ST} in *T. crocea* and *T. maxima*. Degrees of freedom indicate the number of significant axes in the PCNM of the distance matrix. Marg. Var. is the variance in Φ_{ST} explained by the predictor after others in the model have been fitted.

T. crocea

Simplified model: Φ_{ST} ~ LGM + Modern Overwater Distance

		df	Marg. Var	Pseudo-F	Prob.	
Overwater Distance	LGM	5	0.70819	3.2874	0.002	*
	Modern	6	0.79973	3.0936	0.002	*

T. maxima

Simplified model: Φ_{ST} ~ Mean Larval Dispersal + Modern Overwater Distance

		df	Marg. Var	Pseudo-F	Prob.	
Larval Dispersal	Mean	3	0.57076	3.0284	0.014	*
Overwater Distance	Modern	3	0.47911	2.5421	0.035	*

For *T. maxima*, the best predictors after backwards model selection were the mean and minimum larval dispersal distances. Both modern and LGM overwater distances were removed from the models predicting Φ_{ST} during stepwise selection (Table 2). Tests of marginal significance indicated that both minimum and mean larval dispersal distances explain some variation above that explained by the other, but mean larval dispersal distance explained more variance in *T. maxima* Φ_{ST} than minimum (Table 2).

For *T*. sp., the cryptic species, we had only five populations with enough data to analyse so the results of these analyses should be interpreted in light of this. In univariate dbRDA analyses, neither LGM or modern overwater distances significantly predicted Φ_{ST} (Supp. Table 1). Mean and maximum larval dispersal distances both showed significant relationships with Φ_{ST} and neither could be reduced by stepwise model selection, though maximum larval dispersal distance explained slightly more variance than mean (Table 1).

Within-clade analyses

Univariate dbRDA models showed that modern overwater distance and larval dispersal distances significantly predicted Φ_{ST} within clade 2 of *T. crocea* but none of the continuous predictors showed

a significant relationship with Φ_{ST} within clade 3 of *T. maxima*. LGM overwater distance was not a significant predictor of Φ_{ST} within clade 2 of *T. crocea*. When we compared the effects of modern overwater distance with the maximum predictions of the larval dispersal model on pairwise Φ_{ST} within clade 2 of *T. crocea* we found that the maximum larval dispersal distance outperformed overwater distance (Table 3).

Effects of barriers

Tridacna crocea

Conditioning our dbRDA models of barrier effects on each continuous distance in turn, we found that, for *T. crocea*, a combination of the barriers (Torres Strait, Halmahera east and Halmahera west) explained approximately 12% of the variation in addition to the 66% explained by modern overwater distance, 10% above the 64% by LGM overwater distance and 10% over the 63% explained by the mean larval dispersal distance. Tests of marginal significance for each barrier after the other predictors had been fitted showed that the Torres Strait always had significant effects on Φ_{ST} above the effects of continuous distance, whether LGM ($F_{2,36} = 4.1686$, $p = 0.003$), Modern ($F_{2,35} = 6.166$, $p = 0.001$) or larval dispersal distance ($F_{2,36} = 4.3641$, $p = 0.001$) (Table 4). Backwards model selection never removed the Torres Strait from analyses predicting *T. crocea* Φ_{ST}, whereas the Halmahera Eddy east and west were both removed from models conditioned on LGM overwater distance and mean larval dispersal distance. The Halmahera Eddy west remained in the model conditioned on modern overwater distance, and, together with the Torres Strait, explained 12% of the variation in Φ_{ST} in *T. crocea*.

Table 3. Marginal tests of the effects of continuous predictors within clade 2 of *T. crocea* and clade 3 of *T. maxima* in the context of other, competing predictors. Degrees of freedom indicate the number of significant axes in the PCNM of the distance matrix. Marg. Var. is the variance in Φ_{ST} explained by the predictor after all others have been fitted.

Clade 2 *T. crocea*		df	Marg. Var	Pseudo-F	Prob.	
Larval Dispersal	Max	5	0.09141	1.7956	0.049	*
Overwater Distance	LGM	3	0.042143	1.3797	0.198	
	Modern	5	0.05191	1.0197	0.461	

Clade 3 *T. maxima*		df	Marg. Var	Pseudo-F	Prob.
Larval Dispersal	Max	1	0.001385	0.3365	0.833
Overwater Distance	LGM	1	0.000848	0.206	0.951
	Modern	2	0.002382	0.2895	0.962

Table 4. Marginal tests of the relative effects of the Torres Strait and Halmahera barriers on Φ_{ST} in *T. crocea* and *T. maxima* when conditioned on (considered in the context of) each continuous distance measure separately. Degrees of freedom indicate the number of significant axes in the PCNM of the distance matrix. Marg. Var. is the variance in Φ_{ST} explained by the barrier after others in the model have been fitted.

T. crocea

Barrier model: Φ_{ST} ~ Torres + HLM West + HLM East

		df	Marg. Var	Pseudo-F	Prob.	
Conditioned on Mean Larval Dispersal Distance	Torres	2	0.42254	4.3641	0.001	*
	HLM West	1	0.0562	1.1609	0.311	
	HLM East	1	0.12065	2.4922	0.032	*
Conditioned on Modern Overwater Distance	Torres	2	0.48558	6.166	0.001	*
	HLM West	1	0.16794	4.265	0.004	*
	HLM East	1	0.02599	0.6601	0.653	
Conditioned on LGM Overwater Distance	Torres	2	0.38754	4.1686	0.003	*
	HLM West	1	0.08102	1.7429	0.118	
	HLM East	1	0.12613	2.7133	0.033	*

T. maxima

Barrier model: Φ_{ST} ~ Torres + HLM West + HLM East

		df	Marg. Var	Pseudo-F	Prob.	
Conditioned on Mean Larval Dispersal Distance	Torres	1	0.35694	6.9616	0.001	*
	HLM West	1	0.15856	3.0924	0.043	*
	HLM East	1	0.29075	5.6705	0.004	*
Conditioned on Modern Overwater Distance	Torres	1	0.55135	10.4867	0.001	*
	HLM West	1	0.05174	0.9841	0.392	
	HLM East	1	0.31413	5.9747	0.01	*
Conditioned on LGM Overwater Distance	Torres	1	0.54754	9.7679	0.001	*
	HLM West	1	0.08992	1.6041	0.209	
	HLM East	1	0.30592	5.4574	0.007	*

Tridacna maxima

The Torres Strait, Halmahera Eddy East and Halmahera Eddy West barriers together explained approximately 13% of variation in *T. maxima* pairwise Φ_{ST} in addition to the 64% explained by modern overwater distance, 14% over the 62% explained by LGM overwater distance and 12% over the 65% explained by mean larval dispersal distance. Model reduction consistently removed the Halmahera Eddy West from models predicting *T. maxima* Φ_{ST} with barriers, regardless of the continuous distance held as a condition (Table 4). The Torres Strait and Halmahera Eddy East remained in all *T. maxima* Φ_{ST} models, whether conditioned on modern or LGM overwater distances, the mean or minimum larval dispersal distances, and together explained approximately 10% of variation in *T. maxima* pairwise Φ_{ST} (Table 4).

DISCUSSION

Here we synthesised new and existing COI sequence data for three species of *Tridacna* to assess the patterns of genetic diversity in the central Indo-Pacific. We evaluated the relative strengths of three continuous measures of distance in predicting genetic differentiation and found that the best predictor depended on the species under consideration. The filling of a significant sampling gap in the Coral Sea, Solomon Islands and PNG enabled us to assess the impact of the Torres Strait landbridge relative to the effects of other barriers in the region for the first time, and indeed the Torres Strait emerges as a substantial predictor of genetic differentiation in two of these species, consistently explaining more variance than the effects of the Halmahera Eddy. Below we discuss our findings for *T. maxima, T. crocea* and *T*. sp. within the context of previous work in the region, the implications for our understanding of contemporary versus historical routes of dispersal and management implications.

Low genetic connectivity among regions in the Indo-Pacific

Patterns of genetic structure among the three species were generally concordant, with evidence of very low genetic connectivity across the region, consistent with previous work (Ravago-Gotanco et al. 2007; DeBoer et al. 2008; Kochzius and Nuryanto 2008; Nuryanto and Kochzius 2009a; Tisera et al. 2011; Huelsken et al. 2013). Coexistence of divergent clades in eastern Indonesia within all three species, adds to considerable evidence for this as a common region of overlap between clades of other marine species (Barber et al. 2006; Crandall et al. 2008b; Liu et al. 2012). However, our addition of data from PNG allows us to refine placement of genetic discontinuities within the three species. Huelsken et al. (2013) reported deep divergence between *T. maxima* and *T. crocea* populations in the Solomon Islands and Cenderawasih Bay. With samples from Kavieng in PNG, we are able to shift this boundary ~1000km to the northwest and clarify the porosity of this disjunction. In agreement with Huelsken et al. (2013) we found that Coral Triangle populations of *T. crocea* do not share haplotypes with populations to the east and that these regions are populated by different clades (Fig. 6).

For *T. maxima*, however, there is an overlap of clades 2, 3 and 4 in eastern Indonesia and, contrary to Huelsken et al. (2013), we found shared haplotypes between Halmahera and Cenderawasih Bay and populations in the Coral Sea, Solomon Islands and PNG, possibly due to increased sample size and the shortened fragment length of our combined data. The fact that several haplotypes from clade 4 are restricted to populations outside the Coral Triangle, while Indonesian populations generally share only the more common haplotypes, suggests that gene flow has progressed east to west. This is concordant with the New Guinea Coastal Current, a route that would not have been blocked during Pleistocene lower sealevels. Previous work on *T. maxima* identified deep divergence between eastern Micronesia and the Coral Triangle (Gardner et al. 2012). Given the geography, one might assume that the Coral Sea, PNG and the Solomon Islands might link the divergent clade in Micronesia to clades found in the Coral Triangle. However, by combining all existing data, it is clear that these populations form their own distinct clade.

Contrary to Gardner et al. (2012), following combination of these data, we found two individuals of *T. maxima* from Kiribati with a haplotype shared throughout the Coral Triangle and Taiwan (Fig. 5). This suggestion of a long-range connection between the Coral Triangle and Kiribati that bypasses PNG and the Solomon Islands is concordant with larval transport by the North Equatorial Counter Current (Fig. 1). The distance between the Coral Triangle and Kiribati is over 5000 km. which, given the maximum speed of the NECC (Hsin and Qiu 2012), would take over 4 months to traverse, far outside the capabilities of a giant clam larva. The possibility of stepping stone dispersal via Micronesia could be tested with the inclusion of samples from western Micronesia. We encourage researchers considering such work to sequence COI to allow combination with these data for *T. maxima* and expect to find rare haplotypes with intense sampling of these stepping stone populations.

The signatures of historical *versus* contemporary routes of dispersal

We found support for a stronger influence of historical rather than contemporary dispersal on the genetic structure within *T. crocea* (Table 1). Within *T. maxima* and *T*. sp., however, the reverse was true: genetic differentiation was better explained by a model of contemporary larval dispersal than by LGM or modern overwater distance (Tables 1 and 2). These differences between species, particularly the two species with the best sampling, *T. maxima* and *T. crocea,* is concordant with the patterns of haplotype distribution seen in Fig. 5-6. Haplotypes of *T. maxima*'s clade 2 are spread throughout the Coral Triangle and clade 3 spreads as far as the central Pacific. Clades within *T. crocea*, on the other hand, only mix in northwest Papua where the highly restricted clade 3 mixes with wide-ranging clade 2. The strong performance of contemporary dispersal distances as predictors of genetic structure in *T. maxima* and *T*. sp. suggests that dispersal may be higher in these species. While difference in pelagic larval duration is minimal between these species (19 days max. for *T. maxima*, 17 days max. for *T. crocea* (Jameson 1976)), there may be other differences in life history or ecology that influence successful dispersal or survival of migrants. These differences deserve further investigation, but there are few data on life history or ecology of giant clams from the wild, most data coming from aquaculture studies (Jameson 1976; Lucas 1988; Mies et al. 2012).

The within-clade analyses restricted the spatial scale of analysis to the Coral Triangle only, an area strongly affected by swift, complex currents (Gordon et al. 2003) but also where modern and LGM overwater distances are likely to be fairly similar (Fig.1 and Supp. Fig. 2). The stronger signal of contemporary predictors (modern overwater and larval dispersal distance) over that of LGM overwater distance within clade 2 of *T. crocea* could indicate a stronger effect of contemporary dispersal processes at smaller spatial scales. That *T. maxima* clade 3 data revealed no significant links between any of the distance predictors and genetic differentiation may suggest that dispersal at these small spatial scales is relatively unrestricted.

The effect of hard and soft barriers

Throughout the late Pleistocene, sea levels were significantly lower than today. Landbridges formed across the Sunda and Torres Straits, separating populations on either side for thousands of generations. Populations separated by these land bridges are likely to have accumulated different suites of genetic mutations. We expected to find a strong signature of the Torres Strait landbridge, similar to the reported effects of the Sunda Strait found by previous studies on *Tridacna maxima* and *T. crocea* (DeBoer et al. 2008; Kochzius and Nuryanto 2008; Nuryanto and Kochzius 2009a; DeBoer et al. 2014). In agreement with these predictions, novel samples from the west Pacific allowed us to demonstrate the strong effect of the Torres Strait landbridge on genetic structure within these two species for the first time. This finding is concordant with there being limited water flow across the Torres Strait today (Wolanski et al. 2013).

Unsurprisingly, the effect of the Halmahera Eddy was not as strong as that of the Torres Strait in either *T. maxima* or *T. crocea*. We found a consistent pattern of the eastern edge of this eddy having an effect above and beyond the effects of continous distance (whether larval dispersal distance or either overwater distance) in both species, with the exception of models conditioning on modern overwater distance for *T. crocea.* This reinforces previous suggestions that Cenderawasih Bay, set back from the open ocean and characterised by strong environmental gradients (DeBoer et al. 2012), encloses a relatively isolated set of populations (DeBoer et al. 2008; Kochzius and Nuryanto 2008; Nuryanto and Kochzius 2009a; DeBoer et al. 2014).

Conservation Implications

The signatures of historical barriers can be quickly erased by contemporary gene flow (Avise et al. 1987), but we see high levels of genetic differentiation in Tridacna. This implies that, for example, *T. maxima* and *T. crocea* populations in the region surrounding the Coral Sea are relatively isolated from surrounding regions. It should be remembered that we used a single mtDNA locus here (COI) so we are not able to estimate gene flow (Marko and Hart 2011). While we do not have multiple, independent loci with which to test gene flow for this data set, microsatellite data for *T. crocea* show that discrete clades remain separated in the Coral Triangle (DeBoer et al. 2014) so high levels of gene flow among the clades here seem unlikely.

Cryptic diversity within marine species is an increasingly common finding in the Indo-Pacific (Palumbi 1996; Colborn et al. 2001; Holland et al. 2004; Meyer et al. 2005; Barber and Boyce 2006), where past vicariant events and continuing diversity of habitats support a huge diversity of life. Here we conducted the first population-level analyses within the cryptic species reported by Huelsken et al. (2013) and subsequently described as *T. noae* by Su et al (2014) and *T. ningaloo* by Penny and Willan (in press). Evidence from this species generally supported our findings of low genetic connectivity within the Indo-Pacific, with the exception of shared haplotypes between Taiwan and the Indian Ocean coast of Australia (Fig. 7). While we had a low number of populations to assess connectivity, this pattern is not entirely surprising, given the route of the Indonesian Throughflow. Populations with sample sizes less than six were excluded, but these locations can further characterise the geographic range of *T.* sp. (Fig. 4). Huelsken et al. (2013) could not confirm whether this species' range was continuous between Australia, the Solomon Islands and Taiwan, but we can now see that the range seems continuous through central and eastern Indonesia and the Philippines. There is no evidence to suggest that this species occurs in Micronesia, Sumatra or the Coral Sea but work remains.

Given the scale and logistical difficulties associated with working in the Indo-Pacific, most research includes a limited set of locations, either densely packed or widely spaced (Keyse et al. 2014). Of *Tridacna* in particular, there have been 15 studies published since 1992 examining genetic connectivity in the Indo-Pacific. This study, in the vein of recent studies taking a synthetic approach to existing DNA sequence data (Vogler et al. 2012; Vogler et al. 2013; Crandall et al. 2014; Selkoe et al. 2014), would have been impossible without the cooperation of multiple researchers. Despite the restrictions imposed on inference by use of a single locus, we have been able to span a significant portion of the ranges of three codistributed species by combining data from multiple sources. Genomic data allow consideration of fine scale processes and independent corroboration of genetic patterns across loci, but they are currently expensive to produce and difficult to combine across research groups. The ubiquity of mtDNA data, and their relative ease of production, argues for their continued inclusion in phylogeographic studies even as we move further into the genomic era (Bowen et al. 2014).

LITERATURE CITED

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics:489-522.

Balkenhol, N., L. P. Waits, and R. J. Dezzani. 2009. Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. Ecography 32:818-830.

Bandelt, H. J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific

phylogenies. Molecular Biology and Evolution 16:37-48.

Barber, P. and S. L. Boyce. 2006. Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. Proceedings of the Royal Society B: Biological Sciences 273:2053-2061.

Barber, P. H., M. V. Erdmann, and S. R. Palumbi. 2006. Comparative phylogeography of three codistributed stomatopods: origins and timing of regional lineage diversification in the Coral Triangle. Evolution 60:1825-1839.

Beckvar, N. 1981. Cultivation, spawning, and growth of the giant clams *Tridacna gigas, T. derasa*, and *T. squamosa* in Palau, Caroline Islands. Aquaculture 24:21-30.

Begg, G., C. Keenan, and M. Sellin. 1998. Genetic variation and stock structure of school mackerel and spotted mackerel in northern Australian waters. Journal of Fish Biology 53:543-559.

Bowen, B. W., K. Shanker, N. Yasuda, M. Celia, M. C. D. Malay, S. von der Heyden, G. Paulay, L. A. Rocha, K. A. Selkoe, P. H. Barber, S. T. Williams, H. A. Lessios, E. D. Crandall, G. Bernardi, C. P. Meyer, K. E. Carpenter, and R. J. Toonen. 2014. Phylogeography unplugged: comparative surveys in the genomic era. Bulletin of Marine Science 90:13-46.

Briggs, J. C., B. W. Bowen, and C. McClain. 2013. Marine shelf habitat: biogeography and evolution. Journal of Biogeography 40:1023-1035.

Carpenter, K. E., P. H. Barber, E. D. Crandall, M. C. A. Ablan-Lagman, Ambariyanto, G. N. Mahardika, B. M. Manjaji-Matsumoto, M. A. Juinio-Meñez, M. D. Santos, C. J. Starger, and A. H. A. Toha. 2010. Comparative phylogeography of the Coral Triangle and implications for marine management. Journal of Marine Biology 2011:1-14.

Chenoweth, S. F., J. M. Hughes, C. P. Keenan, and S. Lavery. 1998. When oceans meet: a teleost shows secondary intergradation at an Indian-Pacific interface. Proceedings of the Royal Society of London Series B-Biological Sciences 265:415-420.

Colborn, J., R. E. Crabtree, J. B. Shaklee, E. Pfeiler, and B. W. Bowen. 2001. The evolutionary enigma of bonefishes (*Albula* spp.): cryptic species and ancient separations in a globally distributed shorefish. Evolution 55:807-820.

Crandall, E. D., M. A. Frey, R. K. Grosberg, and P. H. Barber. 2008a. Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. Molecular Ecology 17:611-626.

Crandall, E. D., M. E. Jones, M. M. MuÑOz, B. Akinronbi, M. V. Erdmann, and P. H. Barber. 2008b. Comparative phylogeography of two seastars and their ectosymbionts within the Coral Triangle. Molecular Ecology 17:5276-5290.

Crandall, E. D., E. J. Sbrocco, T. S. Deboer, P. H. Barber, and K. E. Carpenter. 2012. Expansion dating: calibrating molecular clocks in marine species from expansions onto the Sunda Shelf following the Last Glacial Maximum. Molecular Biology and Evolution 29:707-719.

Crandall, E. D., E. A. Treml, L. Liggins, L. Gleeson, N. Yasuda, P. H. Barber, G. Wörheide, and C. Riginos. 2014. Return of the ghosts of dispersal past: historical spread and contemporary gene flow in the blue sea star *Linckia laevigata*. Bulletin of Marine Science 90:399-425.

Dawson, M. N. 2012. Parallel phylogeographic structure in ecologically similar sympatric sister taxa. Molecular Ecology 21:987-1004.

DeBoer, T. S., A. C. Baker, M. V. Erdmann, Ambariyanto, P. R. Jones, and P. H. Barber. 2012. Patterns of *Symbiodinium* distribution in three giant clam species across the biodiverse Bird's Head region of Indonesia. Marine Ecology Progress Series 444:117-132.

DeBoer, T. S., M. R. Naguit, M. V. Erdmann, M. C. A. Ablan-Lagman, Ambariyanto, K. E. Carpenter, A. H. A. Toha, and P. H. Barber. 2014. Concordant phylogenetic patterns inferred from mitochondrial and microsatellite DNA in the giant clam *Tridacna crocea*. Bulletin of Marine Science 90:301 - 331.

DeBoer, T. S., M. R. A. Naguit, M. V. Erdmann, M. C. A. Ablan-Lagman, K. E. Carpenter, A. H. A. Toha, and P. H. Barber. 2014. Concordance between phylogeographic and biogeographic boundaries in the Coral Triangle: conservation implications based on comparative analyses of multiple giant clam species. Bulletin of Marine Science 90:277-300.

DeBoer, T. S., M. D. Subia, M. V. Erdmann, K. Kovitvongsa, and P. H. Barber. 2008. Phylogeography and limited genetic connectivity in the endangered boring giant clam across the Coral Triangle. Conservation Biology 22:1255-1266.

Dethmers, K. E. M., D. Broderick, C. Moritz, N. N. Fitzsimmons, C. J. Limpus, S. Lavery, S. Whiting, M. Guinea, R. I. T. Prince, and R. Kennett. 2006. The genetic structure of Australasian green turtles (*Chelonia mydas*): exploring the geographical scale of genetic exchange. Molecular Ecology 15:3931-3946.

Dudgeon, C. L., D. Broderick, and J. R. Ovenden. 2009. IUCN classification zones concord with, but underestimate, the population genetic structure of the zebra shark *Stegostoma fasciatum* in the Indo-West Pacific. Molecular Ecology 18:248-261.

Duncan, K. M., A. P. Martin, B. W. Bowen, and H. G. De Couet. 2006. Global phylogeography of the scalloped hammerhead shark (*Sphyrna lewini*). Molecular Ecology 15:2239-2251.

Elliott, N. G. 1996. Allozyme and mitochondrial DNA analysis of the tropical saddle-tail sea perch, *Lutjanus malabaricus* (Schneider), from Australian waters. Marine and Freshwater Research 47.

Fitt, W. K., C. R. Fisher, and R. K. Trench. 1984. Larval biology of Tridacnid clams. Aquaculture 39:181-195.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3:294-299.

Gardner, J. P. A., C. Boesche, J. M. Meyer, and A. R. Wood. 2012. Analyses of DNA obtained from shells and brine-preserved meat of the giant clam *Tridacna maxima* from the central Pacific Ocean. Marine Ecology Progress Series 453:297-301.

Gilbert, A., S. Andréfouét, L. Yan, and G. Remoissenet. 2006. The giant clam *Tridacna maxima* communities of three French Polynesia islands: comparison of their population sizes and structures at early stages of their exploitation. ICES Journal of Marine Science: Journal du Conseil 63:1573-1589.

Gopurenko, D., J. Hughes, and C. Keenan. 1999. Mitochondrial DNA evidence for rapid colonisation of the Indo-West Pacific by the mudcrab *Scylla serrata.* Marine Biology 134:227-233.

Gordon, A. L., C. F. Giulivi, and A. G. Ilahude. 2003. Deep topographic barriers within the Indonesian seas. Deep Sea Research Part II: Topical Studies in Oceanography 50:2205-2228.

Group, M. S. 1996. *Tridacna crocea.* The IUCN Red List of Threatened Species. .

Hellberg, M. E. 2009. Gene flow and isolation among populations of marine animals. Annual Review of Ecology and Systematics 40:291-310.

Holland, B. S., M. N. Dawson, G. L. Crow, and D. K. Hofmann. 2004. Global phylogeography of *Cassiopea* (Scyphozoa: Rhizostomeae): molecular evidence for cryptic species and multiple invasions of the Hawaiian Islands. Marine Biology 145:1119-1128.

Hsin, Y.-C. and B. Qiu. 2012. Seasonal fluctuations of the surface North Equatorial Countercurrent (NECC) across the Pacific basin. Journal of Geophysical Research 117.

Huelsken, T., J. Keyse, L. Liggins, S. Penny, E. A. Treml, and C. Riginos. 2013. A Novel Widespread Cryptic Species and Phylogeographic Patterns within Several Giant Clam Species (Cardiidae: *Tridacna*) from the Indo-Pacific Ocean. PLOS one 8:e80858.

Huelsken, T., S. Schreiber, and M. Hollmann. 2011. COI amplification success from mucus-rich marine gastropods (Gastropoda: Naticidae) depends on DNA extraction method and preserving agent. Mitteilungen der Deutschen Malakozoologischen Gesellschaft 85:17-26.

Imron, B. Jeffrey, P. Hale, B. M. Degnan, and S. M. Degnan. 2007. Pleistocene isolation and recent gene flow in *Haliotis asinina*, an Indo-Pacific vetigastropod with limited dispersal capacity. Molecular Ecology 16:289-304.

Jameson, S. C. 1976. Early Life History of the Giant Clams *Tridacna crocea* Lamarck, Tridacna maxima (Röding), and *Hippopus hippopus* (Linnaeus). Pacific Science 30:219-233.

Kashino, Y., H. Watanabe, B. Herunadi, M. Aoyama, and D. Hartoyo. 1999. Current variability at the Pacific entrance of the Indonesian Throughflow. Journal of Geophysical Research: Oceans (1978-2012) 104:11021-11035.

Keyse, J., E. D. Crandall, R. J. Toonen, C. P. Meyer, E. A. Treml, and C. Riginos. 2014. The scope of published population genetic data for Indo-Pacific marine fauna and future research opportunities in the region. Bulletin of Marine Science 90:47-78.

Kinch, J. 2002. Giant clams: their status and trade in Milne Bay Province, Papua New Guinea. .

Klanten, O. S., J. H. Choat, and L. van Herwerden. 2007. Extreme genetic diversity and temporal rather than spatial partitioning in a widely distributed coral reef fish. Marine Biology 150:659-670.

Knowlton, N. 1993. Sibling species in the sea. Annual Review of Ecology and Systematics 24:189-216.

Kochzius, M. and A. Nuryanto. 2008. Strong genetic population structure in the boring giant clam, *Tridacna crocea*, across the Indo-Malay Archipelago: implications related to evolutionary processes and connectivity. Molecular Ecology 17:3775-3787.

Legendre, P. and J. Fortin. 2010. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Molecular Ecology Resources 10:831–844.

Legendre, P., F.-J. Lapointe, and P. Casgrain. 1994. Modeling brain evolution from behavior: a permutational regression approach. Evolution 48:1487-1499.

Lessios, H. A., J. Kane, and D. R. Robertson. 2003. Phylogeography of the pantropical sea urchin *Tripneustes*: Contrasting patterns of population structure between oceans. Evolution 57:2026-2036.

Liggins, L., E. A. Treml, and C. Riginos. 2013. Taking the plunge: an introduction to undertaking seascape genetic studies and using biophysical models. Geography Compass 7:173-196.

Liu, S. Y. V., C. F. Dai, G. R. Allen, and M. V. Erdmann. 2012. Phylogeography of the neon damselfish *Pomacentrus coelestis* indicates a cryptic species and different species origins in the West Pacific Ocean. Marine Ecology Progress Series 458:155-167.

Lucas, J. S. 1988. Giant clams: description, distribution and life history. Pages 21-32 in J. W. Copland, Lucas, J.S., editor. Giant Clams in Asia and the Pacific.

Lukoschek, V., M. Waycott, and H. Marsh. 2007. Phylogeography of the olive sea snake, *Aipysurus laevis* (Hydrophiinae) indicates Pleistocene range expansion around northern Australia but low contemporary gene flow. Molecular Ecology 16:3406-3422.

Marko, P. B. and M. W. Hart. 2011. The complex analytical landscape of gene flow inference. Trends in Ecology & Evolution.

Meyer, C. P., J. B. Geller, and G. Paulay. 2005. Fine scale endemism on coral reefs: archipelagic differentiation in turbinid gastropods. Evolution 59:113-125.

Mies, M., F. Braga, M. S. Scozzafave, D. E. L. d. Lemos, and P. Y. G. Sumida. 2012. Early development, survival and growth rates of the giant clam *Tridacna crocea* (Bivalvia: Tridacnidae). Brazilian Journal Of Oceanography 60:129-135.

Miller, A. D., V. L. Versace, T. G. Matthews, S. Montgomery, and K. C. Bowie. 2013. Ocean currents influence the genetic structure of an intertidal mollusc in southeastern Australia - implications for predicting the movement of passive dispersers across a marine biogeographic barrier. Ecology and evolution 3:1248-1261.

Mirams, A. G. K., Treml, E.A., Shields, J.L., Liggins, L. & Riginos, C. 2011. Vicariance and dispersal across an intermittent barrier: population genetic structure of marine animals across the Torres Strait land bridge. Coral Reefs.

Nuryanto, A. and M. Kochzius. 2009. Highly restricted gene flow and deep evolutionary lineages in the giant clam *Tridacna maxima*. Coral Reefs 28:607-619.

Oksanen, J., R. Kindt, P. Legendre, B. O‚ÄôHara, M. H. H. Stevens, M. J. Oksanen, and M.

Suggests. 2007. The vegan package. Community ecology package:631-637.

Palumbi, S. R. 1996. What can molecular genetics contribute to marine biogeography? An urchin's tale. Journal of Experimental Marine Biology and Ecology 203:75-92.

Rambaut, A. E. 1996. Se-Al: sequence alignment editior v2.0a11. Available: http://tree.bio.ed.ac.uk/software/. Accessed 2013 October 20.

Rapoport, E. 1994. Remarks on marine and continental biogeography: an areographical viewpoint. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 343:71-78.

Ravago-Gotanco, R. G., R. M. Magsino, and M. A. Juinio-Meñez. 2007. Influence of the North Equatorial Current on the population genetic structure of *Tridacna crocea* (Mollusca: Tridacnidae) along the eastern Philippine seaboard. Marine Ecology Progress Series 336:161-168.

Reeves, J. M., A. R. Chivas, A. Garc√≠a, S. Holt, M. J. Couapel, B. G. Jones, D. I. Cendon, and D. Fink. 2008. The sedimentary record of palaeoenvironments and sea-level change in the Gulf of Carpentaria, Australia, through the last glacial cycle. Quaternary International 183:3-22.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219-1228.

Schiller, A., P. R. Oke, G. Brassington, M. Entel, R. Fiedler, D. A. Griffin, and J. V. Mansbridge. 2008. Eddy-resolving ocean circulation in the Asian–Australian region inferred from an ocean reanalysis effort. Progress in Oceanography 76:334-365.

Selkoe, K. A., O. E. Gaggiotti, B. W. Bowen, and R. J. Toonen. 2014. Emergent patterns of population genetic structure for a coral reef community. Molecular Ecology 23:3064-3079.

Shah, V. and B. McRae. 2008. Circuitscape: a tool for landscape ecology. Pages 62-66 in Proceedings of the 7th Python in Science Conference.

Su, Y., J. H. Hung, H. Kubo, and L. L. Liu. 2014. *Tridacna noae* (Röding, 1798) – a valid giant clam species separated from *T. maxima* (Röding, 1798) by morphological and genetic data. Raffles Bulletin of Zoology 62:124–135.

Teske, P. R., S. von der Heyden, C. D. McQuaid, and N. P. Barker. 2011. A review of marine phylogeography in southern Africa. South African Journal of Science 107:43-53.

Tisera, W. L., M. R. A. Naguit, B. M. Rehatta, and H. P. Calumpong. 2011. Ecology and Genetic

structure of Giant Clams around Savu Sea, East Nusa Tenggara Province, Indonesia. Asian Journal of Biodiversity 3.

Treml, E. A., P. N. Halpin, D. L. Urban, and L. F. Pratson. 2008. Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. Landscape Ecology 23:19-36.

Treml, E. A., J. J. Roberts, Y. Chao, P. N. Halpin, H. P. Possingham, and C. Riginos. 2012. Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. Integrative and Comparative Biology 52:525-537.

Uthicke, S. and J. A. H. Benzie. 2003. Gene flow and population history in high dispersal marine invertebrates: mitochondrial DNA analysis of *Holothuria nobilis* (Echinodermata : Holothuroidea) populations from the Indo-Pacific. Molecular Ecology 12:2635-2648.

Van Herwerden, L., J. H. Choat, S. J. Newman, M. Leray, and G. Hillersoy. 2009. Complex patterns of population structure and recruitment of *Plectropomus leopardus* (Pisces: Epinephelidae) in the Indo-West Pacific: implications for fisheries management. Marine Biology 156:1595-1607.

Vogler, C., J. Benzie, P. H. Barber, M. V. Erdmann, Ambariyanto, C. Sheppard, K. Tenggardjaja, K. Gérard, and G. Wörheide. 2012. Phylogeography of the Crown-of-Thorns Starfish in the Indian Ocean. PLOS one 7:e43499.

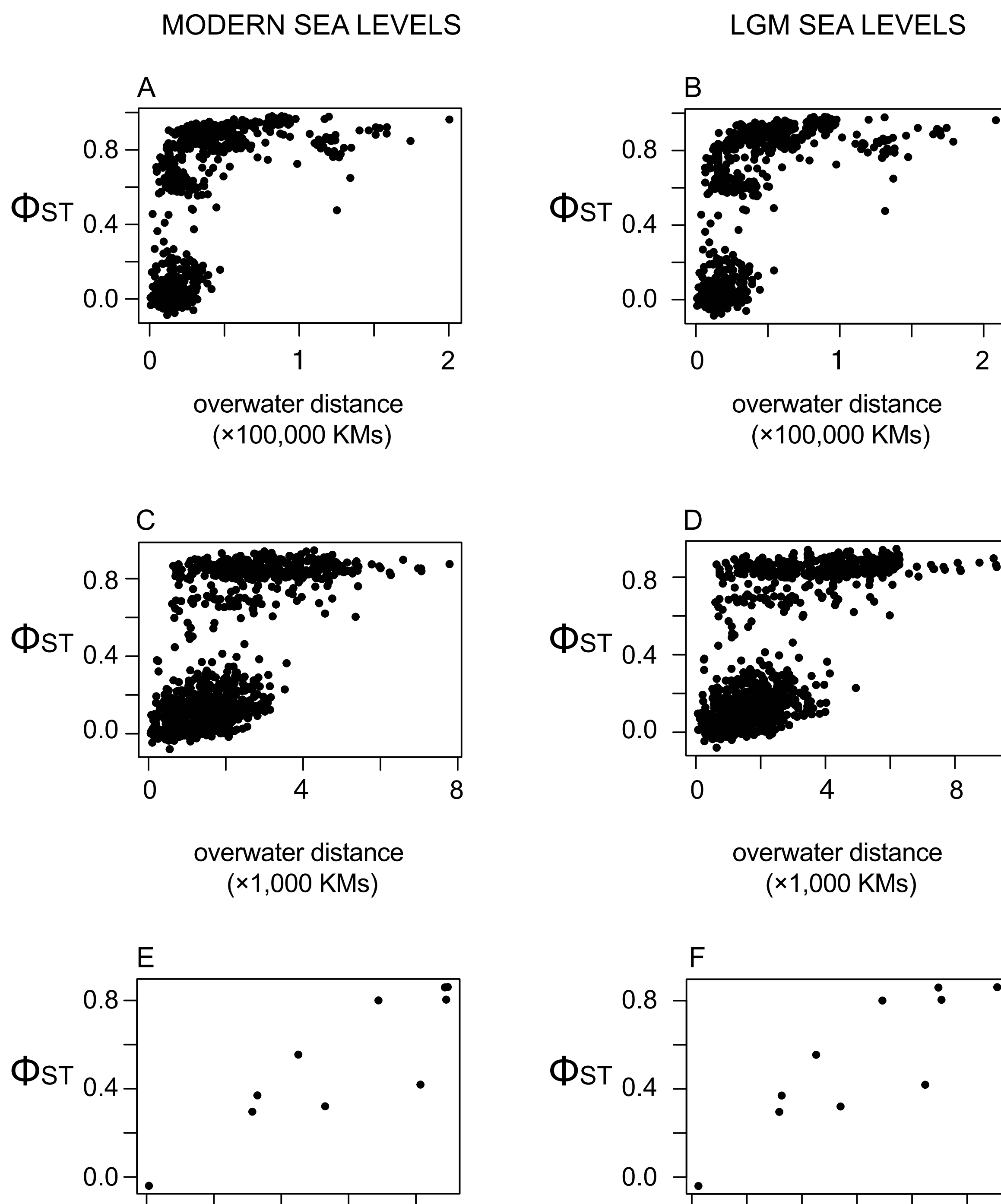
Vogler, C., J. A. H. Benzie, K. Tenggardjaja, Ambariyanto, P. H. Barber, and G. Wörheide. 2013. Phylogeography of the crown-of-thorns starfish: genetic structure within the Pacific species. Coral Reefs 32:515-525.

Voris, H. K. 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. Journal of Biogeography 27:1153-1167.

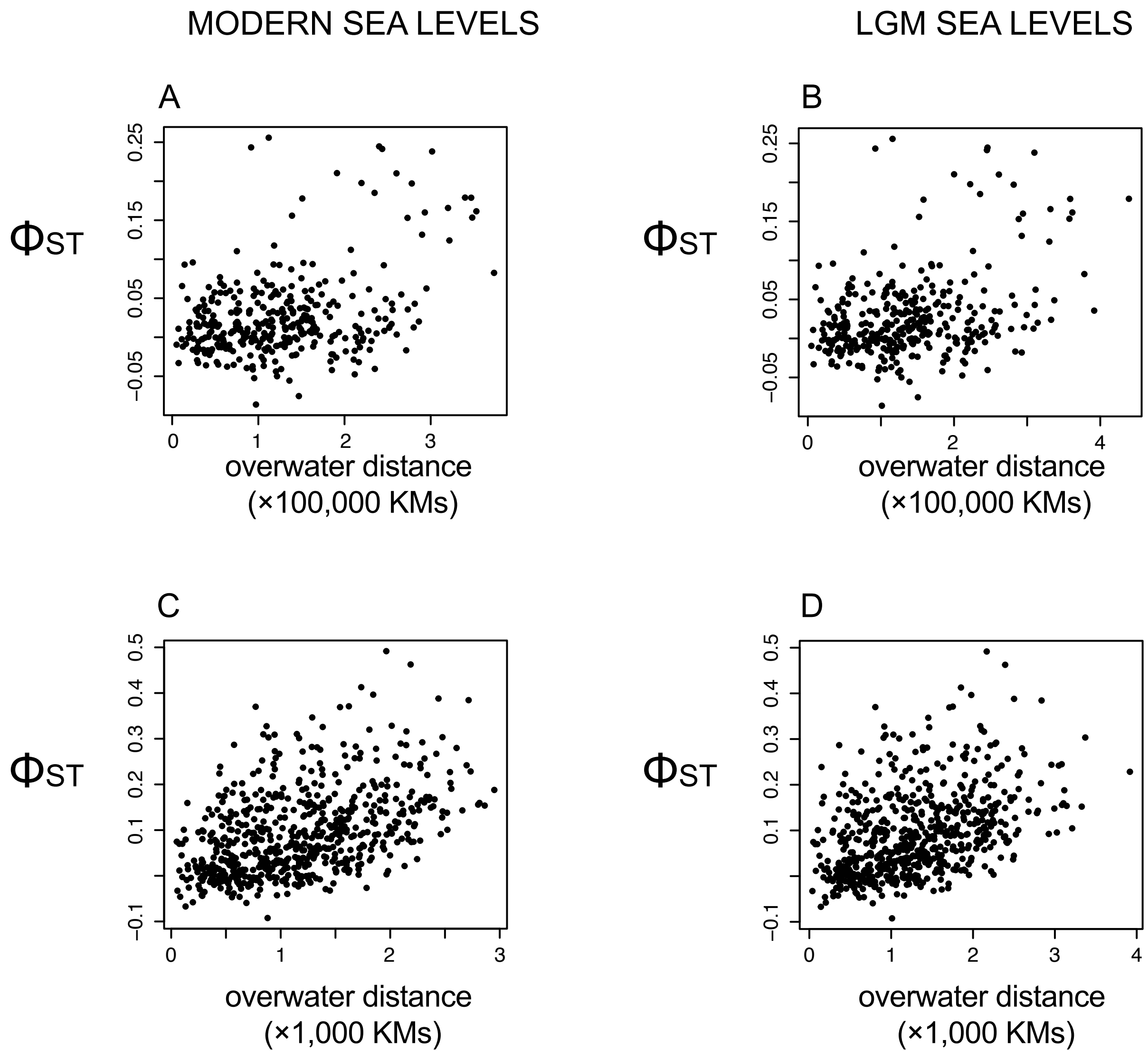
Wolanski, E., J. Lambrechts, C. Thomas, and E. Deleersnijder. 2013. The net water circulation through Torres strait. Continental Shelf Research 64:66-74.

Wright, S. 1943. Isolation by distance. Genetics 28:114-138.

Supplementary Figure 1. Isolation by distance plots of Φ_{ST} predicted by modern overwater distance (A, C, E) and overwater distance at the last glacial maximum (B, D, F) for *T. maxima* (A, B); *T. crocea* (C, D) and *T. sp*. (E, F).



Supplementary Figure 2. Isolation by distance plots of Φ_{ST} within clade 3 of *Tridacna maxima* (A, B)and clade 3 of *Tridacna crocea* (C,D) predicted by modern overwater distance (A, C) and overwater distance at the last glacial maximum (B, D).



Supplementary.Table 1 (a) Summary genetic statistics for populations of *Tridacna maxima*. Bold values indicate significant values after Bonferroni correctionto p = 0.0011

Population name	Code	Lat	Long	n	N.h	% poly	h	pi	D	Fs	Reference
Aceh	WEH	5.86129	95.28118	16	7	1.04	0.750	0.003	-0.230	-3.691	DeBoer et al 2014
Bali	BAL	-8.66604	115.3687	32	19	7.01	0.881	0.009	-1.796	**-11.525**	DeBoer et al 2014
Biak	BIA	-1.16144	136.0474	35	21	10.91	0.950	0.034	0.667	-2.349	Nuryanto & Kochzius 2009; DeBoer et al 2014
Bira	BIR	-5.39285	120.4436	22	16	5.45	0.909	0.005	**-2.390**	**-15.015**	Nuryanto & Kochzius 2009
Camiguin	CAM	9.25802	124.6918	8	7	3.38	0.964	0.009	-1.637	-2.899	DeBoer et al 2014
Heron Island	HER	-23.4443	151.9072	31	16	5.45	0.903	0.007	-1.814	**-9.465**	This study; Huelsken et al 2013
Honda	HON	9.87876	118.7619	9	4	1.30	0.583	0.003	-1.678	-0.822	DeBoer et al 2014
Karimunjava	KAR	-5.81665	110.4102	28	12	4.42	0.680	0.008	-0.988	-3.279	Nuryanto & Kochzius 2009; DeBoer et al 2014
Kavieng	KAV	-2.59457	150.7735	12	7	1.82	0.909	0.006	0.231	-1.934	This study
Kiribati	KIR	1.36452	173.0876	25	12	9.35	0.837	0.016	-1.421	-0.947	Gardner et al 2012
Komodo	KOM	-8.61634	119.5049	26	11	6.23	0.674	0.007	-2.150	-3.478	Nuryanto & Kochzius 2009; DeBoer et al 2014
Kri	KRI	-0.55652	130.6903	9	6	3.12	0.722	0.012	0.091	0.959	DeBoer et al 2014
Kupang	KUP	-10.1631	123.5431	14	10	3.90	0.890	0.008	-1.408	-4.456	Nuryanto & Kochzius 2009
Lihou Reef	LIH	-17.4167	151.6667	12	7		0.773	0.007	-1.063	-1.477	This study; Huelsken et al 2013
Lizard Island	LIZ	-14.6922	145.4558	23	12	5.71	0.889	0.009	-1.213	-3.454	This study; Huelsken et al 2013
Lombok	LOM	-8.37429	116.0544	11	8	2.60	0.927	0.005	-1.809	**-4.551**	DeBoer et al 2014
Luwuk	LUW	-0.87055	123.0793	16	9	2.34	0.858	0.005	-1.314	**-4.818**	Nuryanto & Kochzius 2009
Makassar	MAK	-5.12693	119.3965	11	7	2.08	0.873	0.005	-1.493	-3.323	DeBoer et al 2014
Manado	MAN	1.57023	124.6956	71	28	9.87	0.754	0.005	**-2.485**	**-27.316**	Nuryanto & Kochzius 2009; DeBoer et al 2014
Mayu	MAY	1.30311	126.3686	17	7	7.01	0.596	0.010	-2.236	0.130	DeBoer et al 2014
Misool	MIS	-2.03512	130.3192	8	5	4.16	0.857	0.011	-1.484	0.498	DeBoer et al 2014
Orpheus Island	ORP	-18.6077	146.4887	17	12	5.19	0.934	0.012	-0.873	-4.093	This study
Pulau Doi, Halmahera	DOI	2.27667	127.7798	10	6	1.30	0.889	0.004	-0.783	-2.781	DeBoer et al 2014
Padang, Sumatra	PAD	-0.96027	100.3464	32	14	3.64	0.859	0.004	-1.722	**-9.601**	Nuryanto & Kochzius 2009; DeBoer et al 2014
Palmyra Atoll	PAL	5.88096	-162.072	14	7	1.30	0.824	0.001	-1.986	-1.290	Gardner et al 2012
Perez, Philippines	PER	14.1858	121.9113	16	5	1.04	0.667	0.002	-1.031	-1.977	DeBoer et al 2014
Pulau Seribu, Java	SER	-5.67517	106.5185	113	43	10.91	0.726	0.012	-1.291	**-23.759**	Nuryanto & Kochzius 2009; DeBoer et al 2014
Red Sea	RED	27.8938	34.43735	13	9	2.34	0.936	0.006	-0.666	-4.430	Nuryanto & Kochzius 2009
Romblon, Philippines	ROM	12.5745	122.2482	28	15	3.90	0.820	0.004	-2.179	**-14.185**	DeBoer et al 2014
Western Solomons	SOL	-8.33582	157.4456	31	17	5.71	0.933	0.010	-0.979	-6.988	This study; Huelsken et al 2013
Sangalaki, Borneo	SNK	1.85596	118.5818	7	3	2.08	0.524	0.006	-1.576	1.598	Nuryanto & Kochzius 2009
Sangihe	SAN	3.581	125.48	10	3	0.78	0.378	0.002	-0.507	0.300	DeBoer et al 2014
Selayar	SEL	-6.20754	120.3988	12	7	5.19	0.773	0.009	-2.101	0.060	DeBoer et al 2014
Siquijor	SIQ	9.22909	123.4759	18	7	2.86	0.634	0.004	-2.031	-2.537	DeBoer et al 2014
Sorsogon	SRS	12.9343	123.9797	29	14	5.19	0.828	0.008	-1.489	-5.768	DeBoer et al 2014
Spermonde Archipelago	SPE	-4.6856	119.058	21	9	4.94	0.629	0.006	-2.222	-2.871	Nuryanto & Kochzius 2009
Tanjung Jerawai	JER	1.52029	128.7014	15	12	11.43	0.971	0.044	0.655	-0.675	DeBoer et al 2014
Taiwan	TAI	21.92	120.72	11	9	9.87	0.964	0.024	-1.493	-1.179	Su et al 2014, Tang thesis 2005
Tawi Tawi	TAW	4.97026	119.7622	18	12	3.12	0.895	0.005	-1.693	-8.998	DeBoer et al 2014
Togian Islands	TOG	-0.39444	122.0852	21	17	6.49	0.967	0.008	-2.086	**-13.833**	Nuryanto & Kochzius 2009
Torres Strait	TOR	-10.4875	142.1718	21	13	5.97	0.919	0.009	-1.837	-5.744	This study; Huelsken et al 2013
Western Halmahera	HLM	-0.12773	127.1691	13	7	1.82	0.833	0.004	-1.206	-3.156	DeBoer et al 2014
Yapen	YAP	-1.94088	136.2961	10	7	9.35	0.867	0.025	-1.441	0.612	DeBoer et al 2014

Supplementary.Table 1 (b) Summary genetic statistics for populations of *Tridacna crocea*. Values in bold are significant after Bonferroni correction to p = 0.0011

Population name	Code	Lat	Long	n	N.h	% poly	h	pi	D	Fs	Reference
Aceh	WEH	5.86129	95.28118	30	14	0.05	0.93	0.009	-0.614	-4.065	DeBoer et al 2008; 2014
Bali	BAL	-8.66604	115.3687	19	12	0.09	0.94	0.013	-1.879	-2.490	DeBoer et al 2008; 2014
Biak	BIA	-1.16144	136.0474	40	25	0.13	0.95	0.033	0.263	-5.190	DeBoer et al 2008 & 2014; Kochzius & Nuryanto 2008
Bira	BIR	-5.39285	120.4436	23	12	0.06	0.89	0.010	-1.393	-3.790	Kochzius & Nuryanto 2008
Camiguin	CAM	9.25802	124.6918	9	8	0.03	0.97	0.010	-0.894	-4.034	DeBoer et al 2014
Dinagat	DIN	9.99988	125.5008	23	12	0.04	0.81	0.008	-0.954	-3.629	DeBoer et al 2014
FakFak	FAK	-3.93647	132.8322	17	12	0.05	0.92	0.011	-1.333	-5.504	DeBoer et al 2008; 2014
Honda	HON	9.87876	118.7619	29	15	0.06	0.85	0.009	-1.513	-5.850	DeBoer et al 2014
Kaimana	KAI	-3.81558	133.9269	17	9	0.03	0.87	0.007	-1.212	-3.533	DeBoer et al 2008; 2014
Karimunjava	KAR	-5.81665	110.4102	37	12	0.06	0.70	0.008	-1.367	-3.040	DeBoer et al 2008 & 2014; Kochzius & Nuryanto 2008
Kavieng	KAV	-2.59457	150.7735	13	8	0.04	0.91	0.010	-0.766	-1.818	This study
Kendari	KEN	-3.9774	122.6348	19	14	0.06	0.91	0.011	-1.347	**-7.477**	Kochzius & Nuryanto 2008
Kolorai	KOL	1.63819	128.0623	14	12	0.10	0.98	0.034	0.533	-2.237	DeBoer et al 2014
Komodo	KOM	-8.61634	119.5049	37	25	0.11	0.90	0.012	-2.049	**-19.069**	DeBoer et al 2008 & 2014; Kochzius & Nuryanto 2008
Kota Kinabalu	KKI	5.99868	116.069	17	13	0.07	0.95	0.011	-1.717	**-6.895**	Kochzius & Nuryanto 2008
Kri	KRI	-0.55652	130.6903	23	15	0.07	0.93	0.010	-1.556	**-7.760**	DeBoer et al 2008; 2014
Kupang	KUP	-10.1631	123.5431	9	7	0.03	0.92	0.008	-1.422	-3.017	Kochzius & Nuryanto 2008
Lizard Island	LIZ	-14.6922	145.4558	27	7	0.02	0.62	0.003	-1.576	-3.445	This study
Lombok	LOM	-8.37429	116.0544	6	5	0.04	0.93	0.018	-0.599	-0.075	DeBoer et al 2008; 2014
Luwuk	LUW	-0.87055	123.0793	13	6	0.03	0.72	0.006	-1.258	-1.172	Kochzius & Nuryanto 2008
Makassar	MAK	-5.12693	119.3965	53	20	0.07	0.88	0.008	-1.500	**-11.278**	DeBoer et al 2008; 2014, Kochzius unpublished
Manado	MAN	1.57023	124.6956	24	10	0.03	0.82	0.006	-1.104	-3.912	DeBoer et al 2008 & 2014; Kochzius & Nuryanto 2008
Misool	MIS	-2.03512	130.3192	19	12	0.05	0.95	0.010	-0.906	-5.103	DeBoer et al 2008; 2014
Nambire	NAM	-2.94975	135.6707	22	10	0.04	0.91	0.009	-0.575	-2.413	DeBoer et al 2014
Orpheus Island	ORP	-18.6077	146.4887	16	7	0.04	0.79	0.008	-0.968	-0.828	This study
Pulau Doi, Halmahera	DOI	2.27667	127.7798	6	6	0.07	1.00	0.024	-1.251	-1.298	DeBoer et al 2014
Padang, Sumatra	PAD	-0.96027	100.3464	39	12	0.05	0.91	0.011	0.048	-1.510	Kochzius & Nuryanto 2008; DeBoer et al 2014
Perez, Philippines	PER	14.1858	121.9113	31	19	0.08	0.93	0.012	-1.362	**-10.278**	DeBoer et al 2014
Pulau Kumbur	KUM	-2.99778	135.0556	22	13	0.08	0.90	0.011	-1.980	-4.970	DeBoer et al 2014
Pulau Seribu, Java	SER	-5.67517	106.5185	91	22	0.07	0.79	0.011	-0.673	-5.839	DeBoer et al 2008 & 2014; Kochzius & Nuryanto 2008
Romblon, Philippines	ROM	12.5745	122.2482	34	17	0.06	0.89	0.009	-1.278	**-8.497**	DeBoer et al 2014
Western Solomons	SOL	-8.33582	157.4456	21	7	0.02	0.75	0.005	-0.181	-1.860	This study
Sangalaki, Borneo	SNK	1.85596	118.5818	15	10	0.05	0.90	0.011	-1.294	-3.174	Kochzius & Nuryanto 2008
Sangihe	SAN	3.581	125.48	14	12	0.04	0.98	0.009	-1.149	**-8.179**	DeBoer et al 2008; 2014
Selayar	SEL	-6.20754	120.3988	34	17	0.08	0.78	0.008	-2.005	**-9.628**	DeBoer et al 2008; 2014
Siquijor	SIQ	9.22909	123.4759	7	4	0.02	0.71	0.007	-1.270	0.281	DeBoer et al 2014
Sorong Bay, West Papua	SOR	-0.97081	131.1659	18	12	0.09	0.95	0.026	0.056	-1.234	DeBoer et al 2008; 2014
Sorsogon	SRS	12.9343	123.9797	12	3	0.02	0.44	0.004	-0.380	1.551	DeBoer et al 2014
Tawi Tawi	TAW	4.97026	119.7622	25	11	0.05	0.90	0.010	-0.516	-2.383	DeBoer et al 2014
Togian Islands	TOG	-0.39444	122.0852	20	13	0.07	0.88	0.011	-1.694	-5.545	Kochzius & Nuryanto 2008
Torres Strait	TOR	-10.4875	142.1718	12	6	0.03	0.76	0.005	-1.645	-1.653	This study
Tridacna Atoll	TAT	-2.54715	134.8816	22	10	0.03	0.78	0.006	-1.241	-4.636	DeBoer et al 2014
Ulugan	ULU	10.0815	118.811	25	10	0.04	0.88	0.008	-0.708	-2.349	DeBoer et al 2014
Western Halmahera	HLM	-0.12773	127.1691	12	7	0.02	0.88	0.008	-0.202	-1.862	DeBoer et al 2014
Wayag	WAY	0.17593	130.0594	18	15	0.09	0.96	0.015	-1.690	-6.074	DeBoer et al 2008; 2014
Yapen	YAP	-1.94088	136.2961	29	17	0.09	0.91	0.020	-0.552	-3.998	DeBoer et al 2008; 2014

Supplementary.Table 1 (c) Summary genetic statistics for populations of *Tridacna sp.* Bold values are significant after Bonferroni correction to $p = 0.01$

Population name	Code	Lat	Long	n	% poly	h	pi	D	Fs	Reference
Kavieng	KAV	-2.59457	150.7735	11	0.02	0.618	0.006	-0.539	-0.313	This study
Ningaloo Reef	NIN	-22.0297	113.7397	21	0.03	0.729	0.004	-1.572	**-4.976**	This study, Huelsken et al 2013
Western Solomons	SOL	-8.33582	157.4456	8	0.01	0.643	0.005	-0.251	-0.034	This study
Tanjung Jerawai	JER	1.52029	128.7014	10	0.03	0.756	0.015	1.020	1.672	DeBoer unpublished
Taiwan	TAI	21.92	120.72	7	0.02	0.952	0.006	-0.734	-3.146	Su et al 2014, Tang thesis 2005

Supplementary Table 2 Results of univariate dbRDA models predicting Φ_{ST} in *T. crocea, T. maxima* and *T.*sp using all principle coordinates of continuous predictor distance matrices. The degrees of freedom (df) give information on the number of axes of variation in the variable. Significant predictors from these tests were used in comparative models to assess the relative influence of each predictor (Table 3).

T. crocea Φ_{ST}		df	Variance	Pseudo-F	Prob.	
Larval Dispersal	Max	26	5.1107	3.2792	0.001	*
	Min	25	5.083	3.4965	0.001	*
	Mean	25	4.9578	3.046	0.001	*
Overwater Distance	LGM	27	5.083	2.8778	0.001	*
	Modern	27	5.3619	4.1382	0.001	*
Barriers	Torres	4	3.5294	13.233	0.001	*
	HLM West	8	2.4032	2.8214	0.002	*
	HLM East	4	2.3344	5.9969	0.001	*

T. maxima Φ_{ST}		df	Variance	Pseudo-F	Prob.	
Larval Dispersal	Max	24	6.1655	4.1828	0.001	*
	Min	23	6.4309	7.3442	0.001	*
	Mean	24	6.2387	4.6599	0.001	*
Overwater Distance	LGM	24	6.5058	7.6924	0.001	*
	Modern	24	6.3461	5.5637	0.001	*
Barriers	Torres	5	4.3319	10.533	0.001	*
	HLM West	8	3.427	3.5124	0.003	*
	HLM East	3	3.239	9.8551	0.001	*

***T.* sp** Φ_{ST}		df	Variance	Pseudo-F	Prob.	
Larval Dispersal	Max	2	0.65828	6.4262	0.025	*
	Min	3	0.65288	2.018	0.525	
	Mean	2	0.66323	6.8028	0.008	*
Overwater Distance	LGM	3	0.73545	9.7001	0.1833	
	Modern	3	0.73553	9.7322	0.1833	

CHAPTER 5

General Discussion

Despite the demonstrated importance of genetic diversity for maintaining productive, resilient ecosystems (Hughes et al. 2008), biodiversity is overwhelmingly measured using simple species richness (Magurran 2004). Knowledge of the patterns and processes underlying genetic diversity and a working understanding of the mechanisms generating biodiversity is necessary to engender effective protection of this natural asset. A marine biodiversity hotspot of species richness, such as the Coral Triangle at the centre of the Indo-Pacific (Hoeksema 2007; Bellwood et al. 2012), offers a tractable system to investigate the factors associated with variation within multiple species. This hyperdiverse region is characterised by strong gradients in species richness (Mora et al. 2003; Carpenter and Springer 2005; Allen and Gillooly 2006) but little is known about the generality of this pattern to genetic diversity. Why do we see such strong spatial gradients in the diversity of life? What processes are generating and maintaining such diversity? Can we predict where we will find high levels of diversity?

The Indo-Pacific has been the focus of a substantial body of molecular genetic work on marine species, resulting in a significant repository of genetic diversity data. In this thesis, I demonstrated the advantages of combining existing data to extend the geographic and taxonomic of studies in the Indo-Pacific. In doing so, these extensive data sets are able to provide much greater information to answer questions of broad evolutionary and ecological relevance. Specifically, I assessed the suitability of these data for the purpose of mapping genetic diversity in the Indo-Pacific and proposed ways to improve data utility. I demonstrated two different uses for these data: one method involving the use of published summary statistics for multiple species and the other combining raw DNA sequence data for several codistributed congeners. The first method involved development of a novel statistical approach to predict genetic diversity as a function of species richness with a view to making broad-scale predictions about the distribution of within-species diversity. The second method illustrated the benefits of accessible, georeferenced raw sequence data for allowing broad-scale inference about the historical and contemporary factors influencing diversity in giant clams. Here I discuss each chapter of my thesis in turn, highlighting the significance and limitations of each in the light of existing knowledge before making suggestions of future investigations that could build on my work.

The scope of published population genetic data for Indo-Pacific marine fauna and future research opportunities in the region

In the second chapter of this thesis I showed that, in molecular genetic studies of Indo-Pacific marine species, the majority (>82%) report genetic diversity for just a single species. There are several reasons for this single-species approach: the expense and logistical difficulties of sampling from numerous remote locations, setting up a functional lab and optimising DNA analysis for

multiple species are not just financial burdens but require technical expertise. Commonly, a multi-species research program will be divided up into units suitable for ownership by individual student investigators. The resulting single-species focus, combined with uncohesive sampling among groups working in the Indo-Pacific, has led to a situation where combination of data sets across groups is a pressing need. This chapter highlighted several species of which there has been significant sampling in complementary regions that could be linked to assess genetic patterns across the entire species range. For example, *Scylla serrata*, the mud crab, is a valuable fisheries target that has been studied using mtDNA sequence data intensively in Australia and east Africa but with no studies in the centre of its range. Linking these two data sets together with additional sampling in the Coral Triangle would allow evaluation of population connectivity in the region, and potentially more accurate stock delineation. I identified several such species where existing sampling by different research groups could be bolstered by additional samples, or where combining existing data could be beneficial. One recent example shows the benefits of such an approach: *Linckia laevigata*, the blue sea star, was the target of a study combining data from 38 locations (Crandall et al. 2014), both published data and new collections, to perform the most detailed examination of population structure within a marine invertebrate to date (but see my work on *Tridacna* spp. in Chapter 4).

The 108 studies considered here yielded over 1400 data points of genetic diversity within 116 species across the Indo-Pacific. These data were summary statistics (haplotype diversity, nucleotide diversity and heterozygosity), calculated within species for several different genetic markers; most commonly mtDNA sequence data. This argues for the continued use of sequence data by molecular genetic studies in the region to facilitate data combination. Critically, frequency based genetic data such as allozymes and microsatellites are difficult to combine across research groups, which limits their utility for data synthesis. This continued use of sequence data could easily be supplemental to more intensive genomic studies that can provide insight into contemporary processes acting on marine systems and reveal ongoing selection. Utility of data for spatial analyses relies heavily on the inclusion of accurate geographical information, studies not providing these metadata were excluded from my study resulting in the loss of data from 20 studies. These data might be thought of as ‘dead ends’ in that they can not contribute quantitatively to future spatial analyses.

Finally, Chapter 2 highlighted locations where sets of species had been co-sampled, marking out discontinuities in sampling among regions. Given that making broad-scale inference about concordant patterns across species requires that these species be sampled fairly continuously across the study region, these disconnections represent a significant problem. Specifically, the Great Barrier Reef in Australia was generally sampled for a set of species that differed from those sampled in western Australia or the Coral Triangle. Consequently, patterns of gene flow across these national boundaries and the north of Australia are often untested, leaving knowledge of shared diversity across these boundaries unknown. This chapter did identify several co-sampled locations that span the region that, were they prioritised in future studies, could facilitate combination of data sets at a broad spatial extent (the Seychelles, Christmas and Cocos-Keeling Islands in the Indian Ocean and

Moorea, the Marquesas and Palmyra in the Pacific). These are locations that could be prioritised by those planning new studies with a view to maximising the possibilities for multi-species studies.

Some of the ideas in this chapter grew from discussions at a Catalysis Meeting on the Molecular Ecology of the Indo-Pacific in 2012 (www.nescent.org/science/awards_summary.php?id=290). This has since developed into a working group tasked with compiling and making accessible all the current genetic data on marine life in the Indo-Pacific for a collaborative online community (DIPnet: Diversity in the Indo-Pacific network: http://indopacificnetwork.wikispaces.com/). My collation of these data formed the foundation for the online data base, but the intention is to populate this with raw, georeferenced data. Raw data have several advantages over summary statistics. Firstly, they allow quality control where poor quality DNA sequences, that might greatly alter summary statistics, can be excluded. Secondly, provided each sequence is georeferenced, they allow the redefinition of populations. This prevents incorrectly defined populations falsely increasing summary statistics, such as nucleotide diversity. It also allows the spatial resolution of the genetic data to be altered by splitting or combining populations. Finally, access to raw data allows the re-analysis of existing data with newer techniques that might reveal patterns not visible with the techniques available at the time of publication. The state of currently available raw data does not, however, allow these benefits to be reaped. Poorly georeferenced, incomplete data sets render these data less useful for the purposes of mapping genetic diversity than are summary statistics, which are currently available for a large number of species across a broad spatial scale. For this reason I chose to use summary statistics of genetic diversity in Chapter 3, as discussed below.

Multi-species approach to testing the species-genetic diversity correlation in Indo-Pacific reef animals.

Testing the concordance between patterns of species diversity and within-species genetic diversity has application to fields as diverse as conservation biology and evolutionary ecology. Knowledge of where biodiversity accumulates allows efficient partitioning of conservation resources (Ball 2009). Uncovering some of the correlates of neutral within-species genetic variation informs study of the processes underlying biological diversity. By combining summary statistics of genetic diversity for 75 species of Indo-Pacific reef animals I was able to address the question of whether diversity within and among species is correlated in the Indo-Pacific. This is the first time that the Species-Genetic Diversity Correlation (SGDC (Vellend 2003)), which is predicted by neutral theories such as the Island Theory of Biogeography (MacArthur and Wilson 1967) and the Island Model of Population Genetics (Wright 1940), has been assessed at such a broad spatial extent in the marine environment. Given the fluidity of this environment and the potential for long-range dispersal of many marine species, one might expect that strong ecological patterns would be obscured by high levels of gene flow. However, shallow-water habitats of the Indo-Pacific represent a network of discrete habitat patches characterised by a gradient of species richness that makes this a good system within which to assess the SGDC. Such structure has previously been associated with a higher probability of observing positive correlations between hierarchical levels of diversity

(Vellend et al. 2014). Previous consideration of this question in the marine environment has been scarce and has been characterised by great variability among species in the strength of the correlation (Robinson et al. 2010; Messmer et al. 2012; Noyer and Becerro 2012). Previous work had also been limited in its geographic scope; the broadest spatial extent covered 6000 km of Pacific Ocean but included only three locations (Messmer et al. 2012). In this chapter I sought to examine the generality of the SGDC in the marine environment by testing across as broad a range of species and spatial extent as possible. I maximised the number of species and geographic coverage, including 277 locations across the Indo-Pacific.

My finding of a positive correlation between species richness and genetic diversity had modest support based on lower DIC scores (Table 1, page 90). Individual species slopes were highly variable. This is concordant with previous coral reef work using a fraction of the number of species (Messmer et al 2012). The key difference here is that I was able to average across multiple species in a robust way, accounting for among-species variation using mixed beta regression. Overall, the slope of the SGDC was shallow (Fig. 3, page 91), and significant for only one of three marker subsets. There was a weak positive relationship between these two levels of diversity, but predicting genetic diversity with species richness, or *vice versa* (Papadopoulou et al. 2011), is still out of reach.

The correlation between species and genetic diversity is expected based on the similar processes of migration, differentiation and extinction acting at both levels of diversity. These factors will differ according to the species in question, but when we average across multiple species or control for these life histroy differences (discussed below) we should see concordant patterns. Modelling genetic diversity as a function of factors affecting these processes potentially offers an avenue to predicting genetic diversity. In particular, my future analyses will include the area of suitable habitat available near sampling locations as a predictor of genetic diversity. The expectation is that larger areas will support larger populations, or present larger targets for immigration, and should have higher genetic diversity. These effects are supported by evidence for marine species diversity (Tittensor et al. 2010), but remain untested with genetic diversity. A measure of the isolation of the habitat patch is also highly desirable as a future predictor for genetic diversity. This has been neatly demonstrated for islands in the context of terrestrial plants (Weigelt and Kreft 2013) but the technique needs modification to be suitable for representing the isolation of reef patches for marine species. Network analyses of connectivity, while computationally intensive, hold great promise for the calculation of isolation (Treml et al. 2008; Urban et al. 2009). Quantifying the effects of area and isolation on genetic diversity within species would allow me to begin to see the relative influence of neutral processes in generating and maintaining biodiversity.

Genetic diversity may also be expected to follow the predictions of the intermediate disturbance hypothesis (Connell 1961; 1978): minimum diversity is expected in areas experiencing either very low or very high levels of disturbance. Proxies for disturbance on reefs can be used to predict genetic diversity, ranging in spatial and temporal scale from the frequency and duration of storms (Puotinen 2007; Fabricius et al. 2008) through human population density (Mora et al. 2011) to

long-term climatic stability (Sbrocco and Barber 2013). The relative influence of contemporary (storm intensity and human density) or historical factors (climatic stability) will likely scale with the temporal resolution of the genetic marker used, as seen in Chapter 4 of this thesis and Crandall et al. (2014). Previous work on global patterns of marine species richness has highlighted the importance of environmental factors such as sea surface temperature and productivity (Tittensor et al. 2010). The mixed *beta* regression analysis method I present here can easily be used to test these questions. It allows assessment of the relative importance of multiple predictor variables in the context of among-species variation.

In the near future, raw genetic data with precise georeferenced metadata will become available to enable a more robust assessment of the generality of the SGDC across the Indo-Pacific. Ideally, future analyses of this correlation will use data generated with the same genetic marker for multiple species sampled evenly across the region. If genetic marker is held constant, greater consideration of species life history characteristics in explaining the variation between individual slopes becomes possible. This will allow investigation of the effects of characteristics thought to influence dispersal such as reproductive mode (Riginos et al. 2011) or schooling behaviour (Luiz et al. 2013) in fishes. Knowledge of the species characters underlying the strength of correlation with predictor variables will allow assessment of the generality of these predictions. Limitations of my approach may have stemmed from the restrictions imposed by a scale mismatch between genetic and species diversity data or from my use of published summary statistics of genetic diversity. Either of these limitations could be solved by the use of raw DNA sequence data, provided it is complete and accompanied by informative metadata.

Strong signature of the Torres Strait landbridge on population structure in three codistributed species of *Tridacna* giant clams.

In Chapter 4 I demonstrated the utility of geographically extensive raw, georeferenced sequence data for investigating patterns of genetic diversity across broad spatial extents in closely related species. I focus on three co-distributed species of giant clams in the genus *Tridacna*, a species group identified in Chapter 2 as well sampled in the Indo-Pacific but with data existing as multiple, disconnected data sets. By filling a sampling gap in the west Pacific I was able to unite dense sampling in the Coral Triangle with samples from the central Pacific, Taiwan and the Red Sea to form the most extensive sample of sympatric congeners in the Indo-Pacific. This allowed assessment of a barrier previously untested for this group, the Torres Strait, which was shown to have a strong influence on gene flow in this group. I showed that populations of two of these species, *T. maxima* and T. *crocea*, show strong genetic structure at mtDNA COI between the West Pacific and the Coral Triangle. While further data are needed to determine the levels of gene flow among these regions, these findings have important consequences for management of these species: making the assumption that these strongly differentiated regions are exchanging demographically important levels of migrants would be a risky strategy. These results call for the targeted, multi-locus analysis of west Pacific, central Pacific, Coral Triangle and Indian Ocean populations of giant clams to determine stock structure.

Previously unpublished data from Indonesia (from Dr Timery DeBoer) allowed further characterisation of the geographic range of a cryptic species previously known only from western Australia, the Solomon Islands and Taiwan (Huelsken et al. 2013) and the first population genetic study of this species. The discovery of a cryptic species within this group joins growing evidence of underestimated biodiversity in the region (Vogler et al. 2008; Duda et al. 2009) and for tropical species more generally (Knowlton 1993). Characterising the geographic range of this species serves an important role in contributing to its assessment for conservation status. The morphological characters of this species overlap considerably with both *Tridacna maxima* and *T. squamosa*, indeed, trained collectors have been unable to tell them apart in the field (Penny and Willan 2014). A recent attempt to characterise the range of this species used recognition of mantle patterns typical of Taiwanese individuals (four individauls) to extrapolate species identity from georeferenced photographs of wild clams (Borsa et al. 2014). That these mantle patterns bear no resemblance to those of this species from western Australia (pers. obs.), which clusters with Taiwan in phylogenetic analyses, underscores the importance of matching genetic data with morphology. Such use of mantle morphology, in the absence of definitive characters from multiple samples confirmed by genetic data, can lead to misleading conclusions. The consequences of overestimating the range of such a species could be its under-protection under IUCN or CITES categories. Ideally, the morphology of this species will be defined using multiple individuals examined morphologically and sequenced to confirm their identity.

The data set I assemble for these three codistributed species of *Tridacna* numbered 2031 sequences from four different research groups. This data set would not have been possible without direct collaborations with each research group as the data could not be assembled from publicly accessible databases. In several cases the COI sequences available on NCBI Genbank represented only unique haplotypes and geographic information was often missing or merely specified the country of collection. While direct collaboration can be rewarding (as has been my experience) it is often not easy to make speedy progress when multiple researchers are involved in a project. Specifically, unless clear expectations are set out early, issues may arise concerning authorship, editing of text can take significantly longer when author number is high, and differences of opinion can potentially cause collaborations to break down.

Synthesising so many DNA sequences presents its own challenges. File formats for sequence data can vary substantially across different analysis programs (Lischer and Excoffier 2012), making combination of data time-consuming. While each lab involved in Chapter 4 had sequenced approximately the same section of the COI gene, the overlapping portion across all was substantially shorter than each individual alignment. These limitations aside, one important benefit of this approach is that I was able to check DNA sequences for errors that might significantly alter the measurement of genetic diversity. Ambiguous nucleotide bases were excluded from analyses and early analyses did indeed prompt re-evaluation of some sections of the sequence alignments. Initial visualisation of haplotype networks revealed several ‘odd’ individuals that appeared to be

highly differentiated from others; this was subsequently found to be due to a section of low quality sequence in a single individual. When working with published indices, one has no knowledge of, or control over, these potential issues.

Conclusions

The tropical Indo-Pacific region is vast both in geographic extent and biological complexity. Molecular genetics has made significant contributions to documenting this diversity, although challenges to the use of these plentiful data remain. Data for multiple, co-distributed species allow inference of shared processes driving the distribution of within-species diversity. Variation among species can be substantial and requires further work to elucidate the main causative factors. This thesis has provided a road map for those considering using published genetic data for studies attempting to combine data to perform analyses across a broad geographic and taxonomic range. I have illustrated the limitations and possibilities of these data for mapping genetic diversity and identify sampling gaps to be filled. I have demonstrated and tested a method for analysis of genetic diversity summary statistics which has potential for investigating factors associated with high genetic diversity while allowing explicit consideration of variation among species. Finally, I demonstrated the potential gains of targeted sampling to fill gaps in existing knowledge of factors driving diversity within a specific group of marine invertebrates. My findings underscore the complexity in patterns of genetic diversity within species. Even between closely related species such as the Tridacnids we see marked differences in patterns of genetic diversity. We are still unable to accurately predict patterns of diversity within species, but we have access to a growing repository of genetic data with which to pursue this goal. In our current situation of rapid biodiversity loss, shifting climate norms and growing demands on natural resources we have a responsibility to make the best use of new and existing data with the goal of understanding this fundamental building block of biodiversity.

LITERATURE CITED

Allen, A. P. and J. F. Gillooly. 2006. Assessing latitudinal gradients in speciation rates and biodiversity at the global scale. Ecology Letters 9:947-954.

Ball, I. R., Possingham, H.P., Watts, M. 2009. Marxan and relatives: Software for spatial conservation prioritisation. Pages 185-195 in A. Moilanen, Wilson, K.A., Possingham, H.P., editor. Spatial conservation prioritisation: Quantitative methods and computational tools. Oxford University Press, Oxford, UK., Oxford.

Bellwood, D. R., W. Renema, and B. R. Rosen. 2012. Biodiversity hotspots, evolution and coral reef biogeography: a review.in D. Gower, Johnson, K., Richardson, R., Rosen, B., Rüber, L., Williams, S., editor. Biotic Evolution and Environmental Change in Southeast Asia Cambridge University Press., Cambridge, UK.

Borsa, P., C. Fauvelot, J. Tiavouane, D. Grulois, C. Wabnitz, M. A. Naguit, and S. Andréfouét. 2014. Distribution of Noah's giant clam, *Tridacna noae*. Marine Biodiversity:1-6.

Carpenter, K. E. and V. G. Springer. 2005. The center of the center of marine shore fish biodiversity: the Philippine Islands. Environmental Biology of Fishes 72:467-480.

Connell, J. H. 1961. Effects of competition, predation by *Thais lapillus*, and other factors on natural populations of the barnacle *Balanus balanoides*. Ecological Monographs 31:61-104.

Connell, J. H. 1978. Diversity in tropical rain forests and coral reefs. Science 199:1302-1310.

Crandall, E. D., E. A. Treml, L. Liggins, L. Gleeson, N. Yasuda, P. H. Barber, G. Wörheide, and C. Riginos. 2014. Return of the ghosts of dispersal past: historical spread and contemporary gene flow in the blue sea star *Linckia laevigata*. Bulletin of Marine Science 90:399-425.

Duda, T. F., A. J. Kohn, and A. M. Matheny. 2009. Cryptic species differentiated in *Conus ebraeus*, a widespread tropical marine gastropod. The Biological Bulletin 217:292-305.

Fabricius, K. E., G. De'ath, M. L. Puotinen, T. Done, T. F. Cooper, and S. C. Burgess. 2008. Disturbance gradients on inshore and offshore coral reefs caused by a severe tropical cyclone. Limnology & Oceanography 53:690–704.

Hoeksema, B. 2007. Delineation of the Indo-Malayan centre of maximum marine biodiversity: The Coral Triangle. Biogeography, time, and place: distributions, barriers, and islands 29:117-178.

Huelsken, T., J. Keyse, L. Liggins, S. Penny, E. A. Treml, and C. Riginos. 2013. A Novel Widespread Cryptic Species and Phylogeographic Patterns within Several Giant Clam Species (Cardiidae: *Tridacna*) from the Indo-Pacific Ocean. PLOS one 8:e80858.

Hughes, A. R., B. D. Inouye, M. T. J. Johnson, N. Underwood, and M. Vellend. 2008. Ecological consequences of genetic diversity. Ecology Letters 11:609-623.

Knowlton, N. 1993. Sibling species in the sea. Annual Review of Ecology & Systematics 24:189-216.

Lischer, H. and L. Excoffier. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298-299.

Luiz, O. J., A. P. Allen, D. R. Robertson, S. R. Floeter, M. Kulbicki, L. Vigliola, R. Becheler, and J. S. Madin. 2013. Adult and larval traits as determinants of geographic range size among tropical reef fishes. Proceedings of the National Academy of Sciences 110:16498-16502.

MacArthur, R. H. and E. O. Wilson. 1967. The theory of island biogeography. Princeton Univ Pr.

Magurran, A. E. 2004. Measuring biological diversity. Blackwell Science Ltd, London.

Messmer, V., G. P. Jones, P. L. Munday, and S. Planes. 2012. Concordance between genetic and species diversity in coral reef fishes across the Pacific Ocean biodiversity gradient. Evolution 66:3902-3917.

Mora, C., O. Aburto-Oropeza, A. Ayala Bocos, P. M. Ayotte, S. Banks, A. G. Bauman, M. Beger, S. Bessudo, D. J. Booth, E. Brokovich, A. Brooks, P. Chabanet, J. E. Cinner, J. Cortés, J. J. Cruz-Motta, A. Cupul Magaña, E. E. DeMartini, G. J. Edgar, D. A. Feary, S. C. A. Ferse, A. M. Friedlander, K. J. Gaston, C. Gough, N. A. J. Graham, A. Green, H. Guzman, M. Hardt, M. Kulbicki, Y. Letourneur, A. López Pérez, M. Loreau, Y. Loya, C. Martinez, I. Mascareñas-Osorio, T. Morove, M.-O. Nadon, Y. Nakamura, G. Paredes, N. V. C. Polunin, M. S. Pratchett, H. Reyes Bonilla, F. Rivera, E. Sala, S. A. Sandin, G. Soler, R. Stuart-Smith, E. Tessier, D. P. Tittensor, M. Tupper, P. Usseglio, L. Vigliola, L. Wantiez, I. Williams, S. K. Wilson, and F. A. Zapata. 2011. Global human footprint on the linkage between biodiversity and ecosystem functioning in reef fishes. Plos Biol 9:1-9.

Mora, C., P. M. Chittaro, P. F. Sale, J. P. Kritzer, and S. A. Ludsin. 2003. Patterns and processes in reef fish diversity. Nature 421:933-936.

Noyer, C. and M. A. Becerro. 2012. Relationship between genetic, chemical, and bacterial diversity in the Atlanto-Mediterranean bath sponge *Spongia lamella*. Hydrobiologia 687:85-99.

Papadopoulou, A., I. Anastasiou, F. Spagopoulou, M. Stalimerou, S. Terzopoulou, A. Legakis, and A. P. Vogler. 2011. Testing the species-genetic diversity correlation in the Aegean archipelago: towards a haplotype-based macroecology? The American Naturalist 178:241-255.

Penny, S. S. and R. C. Willan. 2014. Description of a new species of giant clam (Bivalvia: Tridacnidae) from Ningaloo Reef, Western Australia. Molluscan Research 34:201-211.

Puotinen, M. L. 2007. Modelling the risk of cyclone wave damage to coral reefs using GIS: a case study of the Great Barrier Reef, 1969–2003. International Journal of Geographical Information Science 21:97-120.

Riginos, C., K. E. Douglas, Y. Jin, D. F. Shanahan, and E. A. Treml. 2011. Effects of geography and life history traits on genetic differentiation in benthic marine fishes. Ecography 34:566-575.
Robinson, J. D., E. Diaz-Ferguson, M. F. Poelchau, S. Pennings, T. D. Bishop, and J. Wares. 2010. Multiscale diversity in the marshes of the Georgia coastal ecosystems LTER. Estuaries and Coasts

33:865-877.

Sbrocco, E. J. and P. H. Barber. 2013. MARSPEC: ocean climate layers for marine spatial ecology. Ecology 94:979-979.

Tittensor, D. P., C. Mora, W. Jetz, H. K. Lotze, D. Ricard, E. V. Berghe, and B. Worm. 2010. Global patterns and predictors of marine biodiversity across taxa. Nature 466:1098-1101.

Treml, E. A., P. N. Halpin, D. L. Urban, and L. F. Pratson. 2008. Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. Landscape Ecology 23:19-36.

Urban, D. L., E. S. Minor, E. A. Treml, and R. S. Schick. 2009. Graph models of habitat mosaics. Ecology Letters 12:260-273.

Vellend, M. 2003. Island biogeography of genes and species. American Naturalist 162:358-365.

Vellend, M., G. Lajoie, A. Bourret, C. Murria, S. W. Kembel, and D. Garant. 2014. Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. Molecular Ecology 23:2890-2901.

Vogler, C., J. Benzie, H. Lessios, P. Barber, and G. Wörheide. 2008. A threat to coral reefs multiplied? Four species of crown-of-thorns starfish. Biology Letters 4:696-699.

Weigelt, P. and H. Kreft. 2013. Quantifying island isolation. Insights from global patterns of insular plant species richness. Ecography 36:417-429.

Wright, S. G. 1940. Breeding structure of populations in relation to speciation. The American Naturalist 74:232-248.

APPENDICES

Huelsken T, Keyse J, Liggins L, Penny S, Treml EA, Riginos C. 2013. A Novel Widespread Cryptic Species and Phylogeographic Patterns within Several Giant Clam Species (Cardiidae: *Tridacna*) from the Indo-Pacific Ocean. PloS one 8(11):e80858.

Pope LC, Liggins L, Keyse J, Carvalho SB, Riginos C (2015) Not the time or the place: the missing spatio-temporal link in publicly available genetic data. Molecular Ecology. DOI 10.1111/mec.13254

PLOS ONE

A Novel Widespread Cryptic Species and Phylogeographic Patterns within Several Giant Clam Species (Cardiidae: *Tridacna*) from the Indo-Pacific Ocean

Thomas Huelsken[1], Jude Keyse[1], Libby Liggins[1], Shane Penny[2], Eric A. Treml[1,3], Cynthia Riginos[1]*

1 The University of Queensland, School of Biological Sciences, St Lucia, Australia, **2** Charles Darwin University, Research Institute for Environment and Livelihoods, Casuarina, Australia, **3** University of Melbourne, Department of Zoology, Melbourne, Australia

Abstract

Giant clams (genus *Tridacna*) are iconic coral reef animals of the Indian and Pacific Oceans, easily recognizable by their massive shells and vibrantly colored mantle tissue. Most *Tridacna* species are listed by CITES and the IUCN Redlist, as their populations have been extensively harvested and depleted in many regions. Here, we survey *Tridacna crocea* and *Tridacna maxima* from the eastern Indian and western Pacific Oceans for mitochondrial (*COI* and *16S*) and nuclear (*ITS*) sequence variation and consolidate these data with previous published results using phylogenetic analyses. We find deep intraspecific differentiation within both *T. crocea* and *T. maxima*. In *T. crocea* we describe a previously undocumented phylogeographic division to the east of Cenderawasih Bay (northwest New Guinea), whereas for *T. maxima* the previously described, distinctive lineage of Cenderawasih Bay can be seen to also typify western Pacific populations. Furthermore, we find an undescribed, monophyletic group that is evolutionarily distinct from named *Tridacna* species at both mitochondrial and nuclear loci. This cryptic taxon is geographically widespread with a range extent that minimally includes much of the central Indo-Pacific region. Our results reinforce the emerging paradigm that cryptic species are common among marine invertebrates, even for conspicuous and culturally significant taxa. Additionally, our results add to identified locations of genetic differentiation across the central Indo-Pacific and highlight how phylogeographic patterns may differ even between closely related and co-distributed species.

Citation: Huelsken T, Keyse J, Liggins L, Penny S, Treml EA, et al. (2013) A Novel Widespread Cryptic Species and Phylogeographic Patterns within Several Giant Clam Species (Cardiidae: *Tridacna*) from the Indo-Pacific Ocean. PLoS ONE 8(11): e80858. doi:10.1371/journal.pone.0080858

Editor: Mikhail V. Matz, University of Texas, United States of America

Received July 30, 2013; **Accepted** October 14, 2013; **Published** November 20, 2013

Copyright: © 2013 Huelsken et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Funding for this work was provided by the Australian Research Council (www.arc.gov.au, DP0878306 to CR), the German Research Foundation (www.dfg.de/en, DFG, Hu 1806/1-1, Hu 1806/2-1 to TH), the World Wildlife Fund (worldwildlife.org/initiatives/fuller-science-for-nature-fund, Kathryn Fuller Post-doctoral Research Fellowship to EAT), the Malacological Society of Australasia (www.malsocaus.org, to TH), and the Joyce Vickery Fund (linneansocietynsw.org.au/grants.html, to JK). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: c.riginos@uq.edu.au

These authors contributed equally to this work.

Introduction

Giant clams of the genus *Tridacna* are among the most conspicuous marine invertebrates on coral reefs due to their large size and brilliantly colored mantle that contains photosynthesizing symbionts. Giant clams have traditionally provided raw material for tools, containers, and ornaments [1], and many populations are harvested for meat, shells, and the ornamental aquarium trade [2,3]. Despite local management efforts, including mariculture [3], wild stocks of giant clams are depleted and some species are locally extinct in many areas of Southeast Asia and the South Pacific [3–5]. Consequently, most *Tridacna* species are listed by CITES (Appendix II)[6] and the IUCN Redlist [7].

There are currently eight [8] described species within the genus *Tridacna* (*T. crocea* Lamarck, 1819, *T. derasa* (Röding 1798), *T. gigas* (Linnaeus 1758), *T. maxima* (Röding 1798), *T. mbalavuana* Ladd, 1934, *T. rosewateri* Sirenko and Scarlato 1991, *T. squamosa* Lamarck 1819, and *T. squamosina* Sturany 1899), differentiated by morphology and habitat preference [9–12]. *Tridacna squamosina*, *T. rosewateri*, and *T. mbalavuana* have restricted distributions (Red Sea, Mauritius, and Fiji to Tonga, respectively), whereas *T. derasa*, *T. gigas*, *T. crocea*, *T. squamosa* and *T. maxima* are widely distributed in the Indian and Pacific Oceans, with the latter two extending their distribution into the Red Sea [8,9]. Molecular phylogenetic investigations support monophyly of the described species [13–15], albeit with some disagreement among species relationships. An unpublished Master's thesis [16] also reports a morphologically distinct clam from Taiwan and uses mtDNA loci to show that this clam is highly divergent from sympatric *T. maxima*, potentially indicative of an additional unnamed species.

The juncture between the Indian and Pacific Oceans (Fig. 1), where several species of *Tridacna* are sympatric [8], is a well-known epicenter of tropical marine biodiversity [17,18]. Genetic surveys in this region have revealed cryptic species, even among conspicuous and well-studied marine invertebrates [19,20]. Many species show substantial intraspecific genetic division between the ocean basins (reviewed by [21]), with the Sunda Shelf, Molucca and Flores Seas, Makassar Strait, and Bird's Head region of northwest New Guinea emerging as locations of genetic discontinuities [21,22]. These locations span the archipelago commonly referred to as Wallacea, which falls between the Sunda (southeast Asia) and Sahul (Australia and New Guinea) continental shelves

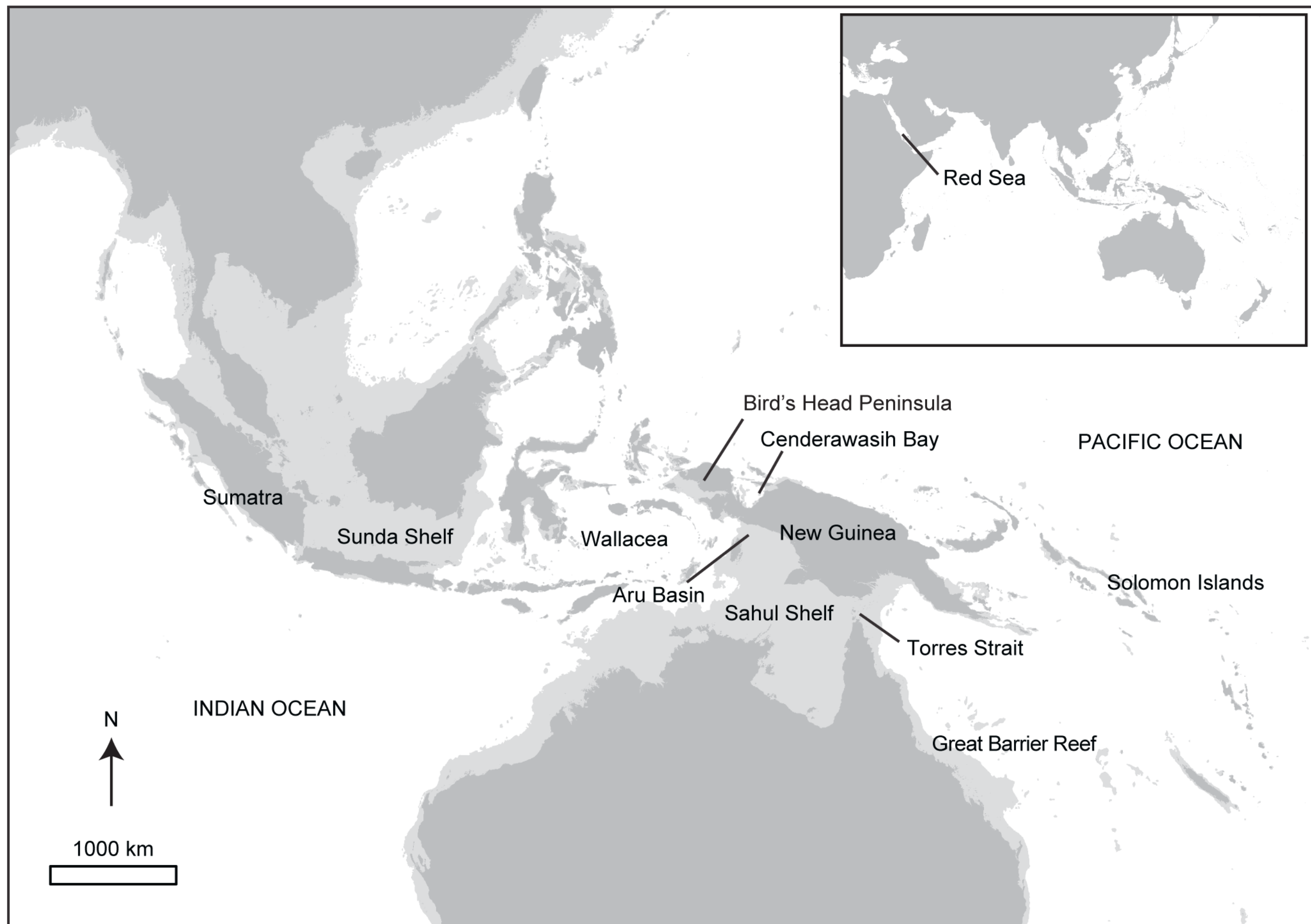


Figure 1. Study region. The light grey outline represents the lowest Pleistocene sea level (120 m depth contour).
doi:10.1371/journal.pone.0080858.g001

and was the only point of permanent oceanic connection between the Indian and Pacific Oceans throughout the Pleistocene [23].

Phylogeographic and population genetic surveys have intensely sampled *T. maxima* and *T. crocea* throughout Wallacea using mitochondrial (mtDNA) markers [24–26], allozymes [27,28], and microsatellites [29]. Both *T. crocea* and *T. maxima* have been shown to contain distinct mtDNA clades associated with Sumatra (Sunda), Wallacea, and northwest New Guinea (Sahul, particularly in Cenderawasih Bay) [24–26]. These lineages are sympatric in some populations, for instance *T. maxima* from northern Java has both Sumatran and Wallacean mitotypes, and similarly *T. crocea* populations from Halmahera eastward through Cenderawasih Bay contain both Wallacean and northwest New Guinean lineages [26,29]. Microsatellite genotyping of *T. crocea* corroborates the distinctiveness of Sumatran and Cenderawasih populations, with evidence for mixing in Wallacea of local genotypes with Cenderawasih-like genotypes [29]. Thus, substantial genetic differentiation typifies at least two *Tridacna* species in this region.

In the Pacific Ocean, *T. derasa, T. gigas, T. maxima* and *T. crocea* have been genetically surveyed, primarily with allozyme markers [27,28,30–35], but also with mtDNA [36]. These studies show genetic divisions between western and central Pacific populations but with some indication that eastern Australian populations show greater affinities with Philippine populations than they do with other western Pacific populations [30,33]. Great Barrier Reef populations (eastern Sahul) form a cluster distinct from, but closely related to, Philippine populations for *T. maxima* and *T. derasa* but with low sampling in the Philippines (two and one populations, respectively) and no sampling in Wallacea or Sunda regions. Thus, it is unknown whether substantial genetic divergence reflects the geographic distance separating the Philippines and eastern Sahul or is indicative of distinct regional groupings.

Here, we examine DNA sequence diversity of *T. crocea* and *T. maxima* whose sampled distributions include the eastern Indian Ocean, Wallacea, and western Pacific Oceans. Data from new samples, predominantly from the western Pacific, are merged with data from previous studies, especially from Wallacea (e.g. [24,25,26]), to present a unified summary of phylogeographic patterns and a point of contrast to earlier broadscale studies based on allozymes [30,32,33,35]. We use phylogenetic analyses to assess evolutionary relationships among species and also gauge regional geographical divisions within species.

Materials and Methods

Sampling and permits

Small mantle biopsies were non-lethally collected from animals with morphology characteristic of *Tridacna maxima* and *T. crocea* at 0–20 m depth from the Solomon Islands, and in Australia from Ningaloo Reef, Heron Island, Lizard Island, the Torres Strait and Lihou Reef. All sampling and tissue transport was in accordance with local and international regulations. Permit details are as

follows: Lihou Reef, Australia: Department of Sustainability, Environment, Water, Population & Communities (Access to Biological Resources in a Commonwealth Area for Non-Commercial Purposes permit number: AU-COM2008042); Lizard Island and Heron Island, Australia: Great Barrier Reef Marine Park Authority and Queensland Parks and Wildlife (Marine Parks Permits: G08/28114.1, G09/31678.1, G10/33597.1, G11/34640.1); Ningaloo Reef, Australia: Western Australia Department of Environment and Conservation (License to take Fauna for Scientific Purposes: SF007126, SF006619, SF008861; Authority to Enter Calm Land/or Waters: CE002227, CE002627, Department of Fisheries, Western Australia Exemption 2046); Queensland: Queensland Government Department of Primary Industries (General Fisheries Permits: 118636, 150981); Torres Strait Islands, Australia: Commonwealth of Australia Torres Strait Fisheries Act 1984 and Australian Fisheries Management Authority (Permit for Scientific Purposes: 8562); Solomon Islands: Solomon Islands Government Ministry of Education and Human Resource Development and Ministry of Fisheries and Marine Resources (research permit: to S Albert, expiry 31/10/2011); Solomon Islands Government Ministry of Environment, Conservation and Meteorology (Convention on International Trade in Endangered Species of Wild Fauna and Flora export permit: EX2010/102); Australian Government Department of the Environment, Water, Heritage and the Arts (Convention on International Trade in Endangered Species of Wild Fauna and Flora import permit: 2010-AU-616020); Australian Quarantine Inspection Service (Permit to Import Quarantine Material: IP10017966).

DNA sequences

DNA was extracted using a modification of the Qiagen DNeasy protocol [37]. Primers that targeted mitochondrial cytochrome oxidase 1 (*COI*) [24,26,38] and ribosomal *16S* [39] were used to amplify 390 and 417 basepair segments of the respective gene regions. A subset of samples were amplified for the partial nuclear *18S* and *ITS1* region (referred to as *ITS* in text) to provide independent estimates of phylogenetic relationships using primers from [13,40]. PCR products were purified following a standard Exo-Sap protocol (New England Biolabs) and were sequenced by Macrogen (Korea). Trace files were edited in CodonCode Aligner (ver. 4.0.3). In addition, the NCBI repository of nucleotide sequences was searched for all published *Tridacna COI* and *16S* sequences (August 2012) representing both intraspecific [24–26,41] and interspecific [9,15,16] surveys. These sequences were manually aligned [42] against our new sequences and against outgroups (*Hippopus hippopus, Hippopus porcellanus, Cerastoderma glaucum, Fragum sueziense*, and *Corculum cardissa*) and trimmed to a common length. For *ITS* there were several insertions/deletions that could not be reconciled, so these areas of low overlap were masked and not used for phylogenetic analyses.

Phylogenetic analyses

Previous mtDNA surveys have used either *16S* [9,15,26] or *COI* [24–26,41] gene regions. To unify these sources of data and address interspecific relationships, we initially took representative sequences across studies and linked them by our samples for which both gene regions had been sequenced in a concatenated search. For samples with only a single gene region (that is, sequences acquired from NCBI), information from the missing gene region was treated as missing data. Up to four individuals per species were retained representing the diversity of their species clade and prioritizing individuals with both *16S* and *COI* sequenced. Using StarBEAST v. 1.6.2 [43] each mtDNA gene region was treated as a separate partition. A general time reversible model with gamma distributed and invariant sites (GTR+G+I) was applied to each gene, with additional partitioning by codon position (1+2, 3) for *COI*. A relaxed molecular clock with an uncorrelated lognormal mutation rate was used for each gene. The *COI* and *16S* gene trees were linked, as mtDNA is a single linked locus (i.e. concatenated gene regions). Priors were set for nodes defining species as a log normal date (mean $=0$, SD $=1$) with an offset representing the most recent estimate of the earliest fossil (*T. crocea*: 1.8, *T. maxima*: 5.3, and *T. squamosa*: 1.8 million years). The root of the Tridacninae was set as normal with mean date of 14 and SD of 2.5 million years. All fossil dates were based on [15,44]. Speciation was modeled both as birth-death and Yule processes in independent runs of 250 million steps, with a burn-in of 25%, and yielded similar results.

Additional genealogical searches were performed using MrBayes ver. 3.1.2 [45] and RAxML (Randomized Axelerated Maximum Likelihood, Blackbox interface) [46]. Using the concatenated file of the same mtDNA sequences as above, searches were partitioned such that *16S* formed one partition, and *COI* formed a second partition with third codon positions partitioned separately from first and second (1+2, 3) for *COI*. In MrBayes, a GTR+G+I (nst $=6$, invgamma) model for all three partitions was used, with a search length of 10 million steps, sampling every 10,000 steps, and a burn-in of 25% (2.5 mill steps). Similarly, the GTR+G+I models were applied to these partitions in RAxML in a maximum likelihood search with 100 bootstrap replicates.

Locus-specific genealogies were also inferred for *COI*, *16S*, and *ITS* using both MrBayes and RAxML. Total data sets for each locus were assembled from all available sequences and then simplified by removing any identical haplotypes. Searches were performed under the same conditions previously described for *16S* (no partitions) and for *COI* (1+2, 3) with four separate searches of 10 million steps and the final 25% percent of trees retained (effectively a burn-in of 7.5 million steps). Search conditions for the partial nuclear *ITS* sequences were as above with indels treated as missing data and no partitioning.

The software Figtree (Rambaut: http://tree.bio.ed.ac.uk/software/figtree/) was used to assist with tree visualization and graphics preparation.

Phylogeographic patterns

Intraspecific phylogeographic patterns were assessed by examining all available *COI* and *16S* sequences for *T. crocea*, *T. maxima*, and the distinct clade (*Tridacna sp.*) identified in the previous analyses. For each species-locus combination, a heuristic maximum parsimony search was conducted in PAUP* [47]. Because frequencies of published haplotypes are not consistently available, it was not possible to conduct standard population genetic analyses such as measures of diversity and differentiation. For intraspecific parsimony searches, the maximum number of trees was set to 1000 in PAUP*[47].

Results

DNA sequences

New DNA sequence data was generated for individuals from five locations (including 55 *COI*, 65 *16S*, and 50 *ITS* sequences: Genbank Acc. Nos. JX974838-JX975007). Combining these new sequence data with previously published data yielded aggregations of 405 *COI*, 132 *16S*, and 50 *ITS* sequences for *Tridacna* species, with 335 unique haplotypes for *COI* and 54 unique haplotypes for *16S*. In the new data generated for this study nearly all included

individuals were sequenced for both *COI* and *16S* allowing us to link results from these two loci and provide a common context for the aggregated sequences from previous studies. Similarly, *ITS* sequences were obtained from an overlapping subset of individuals sequenced for *COI* and *16S*. Nexus files have been deposited in Treebase (http://purl.org/phylo/treebase/phylows/study/TB2:S13501).

Phylogenetic analyses

Phylogenetic analyses resulted in well-resolved topologies defining several clades within *Tridacna*. Tree topologies for the concatenated and single gene datasets were similar (Figs. 2–4), providing evidence for a robust and consistent phylogenetic signal. The concatenated analyses of mitochondrial *COI* and *16S* loci (Fig. 2) strongly support monophyly of *T. squamosa, T. crocea*, and a previously undescribed clade (but reported in [16]) formed well-supported terminal taxa, with more modest support for the monophyly of *T. maxima*. This undescribed clade (which we refer to as *Tridacna sp.*) was also well supported in single gene analyses of *COI* and *16S* (Fig. 3) and *ITS* (Fig. 4). *T. sp.* sequences were evolutionarily distinct from other species; the average pairwise *COI* sequence divergence between *T. sp* and *T. crocea* was 14.4% and was 12.6% between *T. sp.* and *T. squamosa*, as compared to 9.5% between *T. crocea* and *T. squamosa* (uncorrected pairwise distances).

Gene trees for *COI* and *16S* show concordant relationships among species (Fig. 3), confirming that independent research groups have sampled similar genotypes. The notable exception to the consistency across studies was the *16S T. derasa* sequence from [15] which did not cluster consistently with our *16S T. derasa* sequence (specimen ET358) even though our *COI* sequence from this same individual clustered with other *T. derasa* sequences including GQ166591 from [48]. For this reason, the *T. derasa* sequence from [15] was retained in the *16S* tree, but excluded from the joint *COI* and *16S* searches. All mtDNA-based genealogies supported *T. squamosa* and *T. crocea* as sister species (Figs. 2 and 3) whereas *ITS* based analyses gave modest support for *T. sp.* and *T. crocea* as sister species (Fig. 4). Within the mtDNA-based analyses, *T. derasa*, *T. gigas*, and *T. mbalavuana* appear consistently as basal lineages within *Tridacna* (Figs. 2 and 3). (No *ITS* sequences were available for these taxa.)

Phylogeographic patterns

Within *T. crocea* and *T. maxima*, there was broadscale phylogeographic concordance of mtDNA gene trees (as shown in Fig. 5). *T. crocea* and *T. maxima* haplotypes from the Solomon

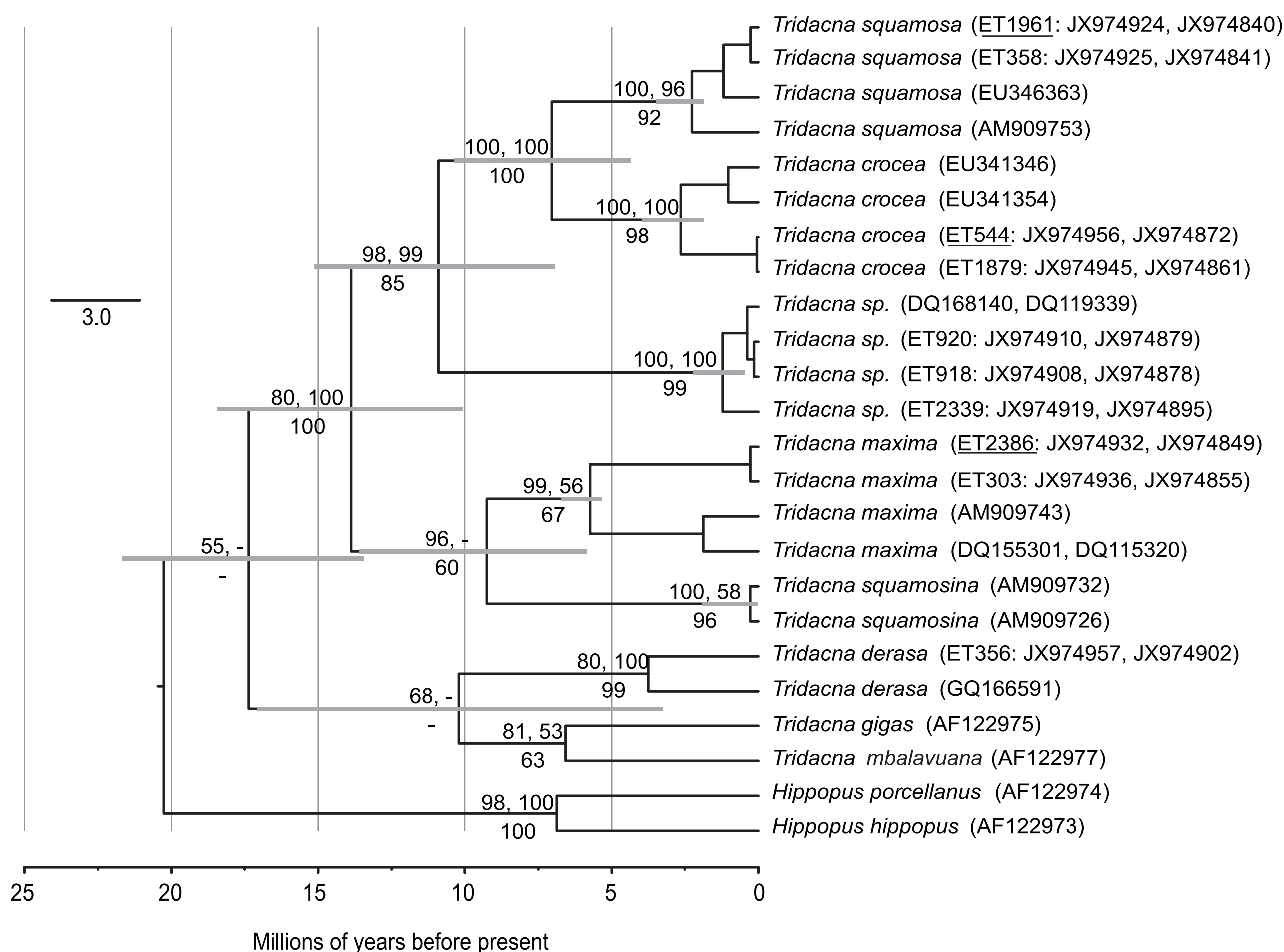


Figure 2. Species relationships within *Tridacna* based on concatenated mitochondrial DNA (*COI* and *16S*) sequences. The topology shown is a time calibrated maximum clade credibility tree inferred with StarBEAST under a birth-death model. Bayesian posterior probabilities from StarBEAST and MrBayes are above branches and RAxML bootstrap support percentages are below branches. Individuals with two accession numbers include both *COI* and *16S* sequences. Individuals that are underlined also appear in Fig. 4.
doi:10.1371/journal.pone.0080858.g002

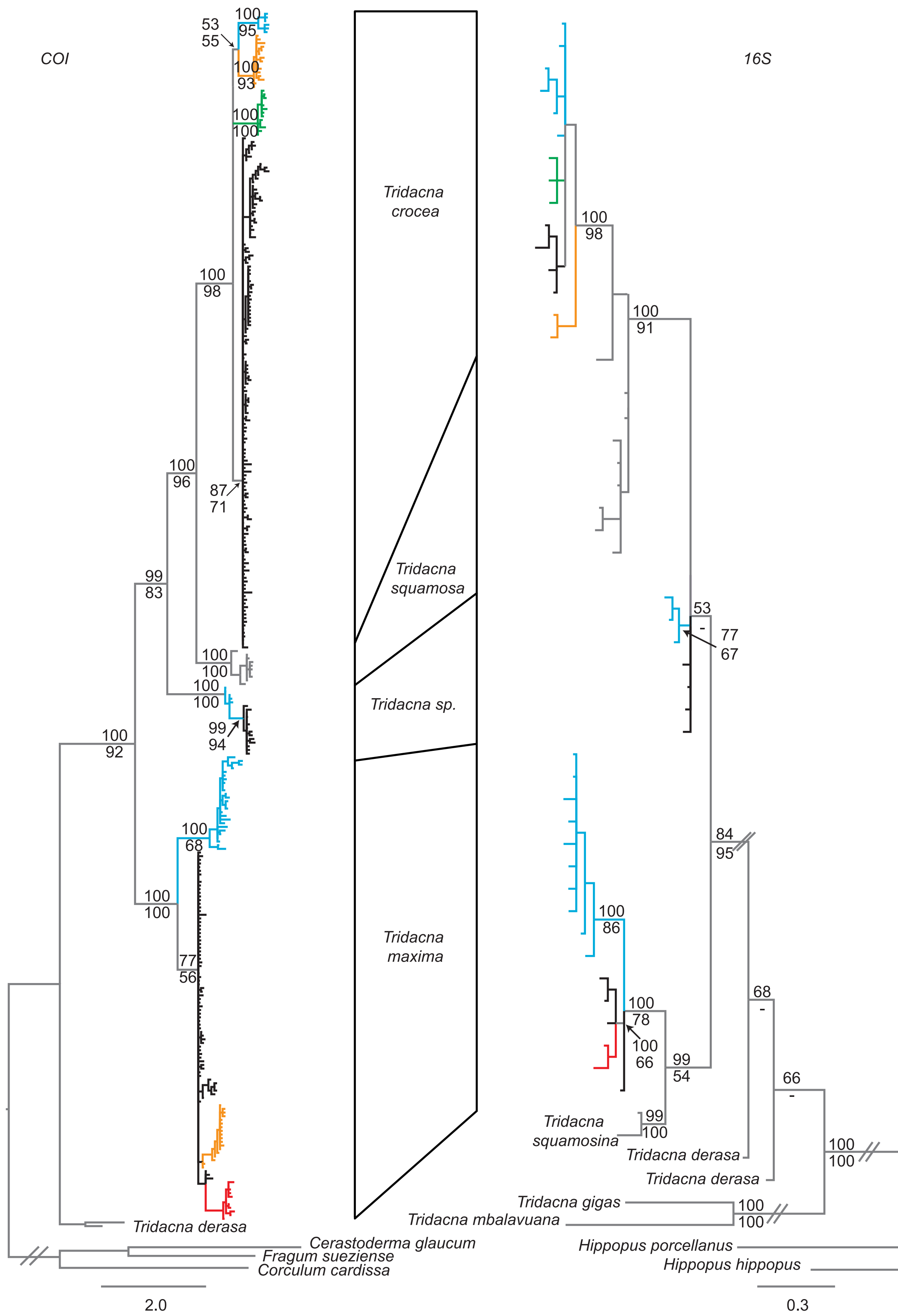

COI
16S
Tridacna crocea
Tridacna squamosa
Tridacna sp.
Tridacna maxima
Tridacna derasa
Cerastoderma glaucum
Fragum sueziense
Corculum cardissa
Tridacna squamosina
Tridacna gigas
Tridacna mbalavuana
Hippopus porcellanus
Hippopus hippopus
2.0
0.3

Figure 3. Bayesian phylogenetic trees for mitochondrial *CO1* and *16S*. MrBayesian consensus trees constructed for each gene region using all available data. Although different species and regions have differential representation, the two gene trees are concordant, as is expected for linked loci. Thus, overall patterns are consistent among research groups. Branch colors correspond to distinct lineages whose geographic distributions are described in Fig. 5.
doi:10.1371/journal.pone.0080858.g003

Islands, the Torres Strait and Lizard Island (and additionally western New Guinea/Cenderwasih Bay, Lihou Reef and Heron Island for *T. maxima*) formed a distinct monophyletic 'Pacific' group (colored blue in Fig. 5). Sequences from the Sunda Shelf formed a second monophyletic group (colored orange in Fig. 5) as described in the original publications [24–26], although the location or the genetic break differed slightly for each species. Finally, sequences from Indonesia, Singapore, western New Guinea/Cenderwasih Bay and Taiwan formed a third group (black in Fig. 5). Most sequences published in Genbank are not georeferenced. We were, however, able to deduce the distinct clades typifying major regions from previously published surveys by recreating previously published analyses; *T. crocea* (yellow haplotypes of [24], grey clade of [26]) are shown in green and orange respectively, and *T. maxima* (yellow haplotypes of [25]) are shown in blue and orange respectively in Fig. 5.

For *T. maxima*, the northwest New Guinea clade formed a cluster with the Pacific clade, although no haplotypes were shared between the two locations. For *T. crocea*, however, haplotypes from northwest New Guinea and the western Pacific were members of two distinct monophyletic groups: the Pacific (blue) and the Wallacea (black) groups (Fig. 5). The *T. crocea* and *T. maxima 16S* sequences from [15], described as having been obtained from individuals sourced from aquarium stores, both fell within Pacific haplotype groups, suggesting that these purchased specimens had a Pacific origin.

Despite the reduced sampling for *T. sp.*, a 'Pacific' lineage was similarly positioned in the Solomon Islands, and a distinct lineage, comprising samples from western Australia and Taiwan, geographically overlapped with the Wallacea (black) lineage portrayed in *T. crocea* and *T. maxima*. Similar phylogeographic patterns were evident for *COI* and *16S* for each species despite only partially overlapping sets of individuals forming the basis for each tree.

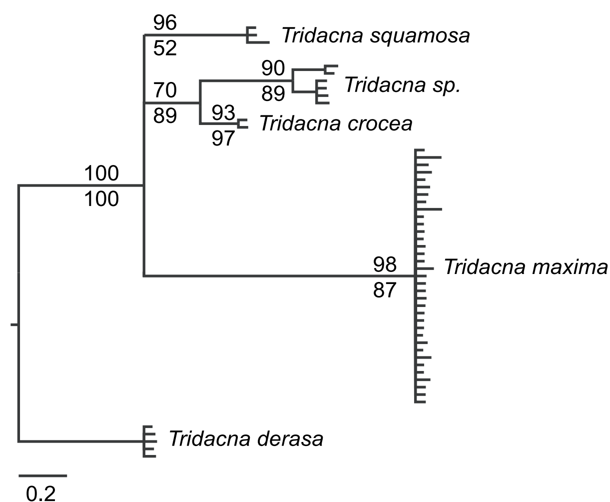


Figure 4. Species relationships within *Tridacna* based an *ITS* MrBayes consensus tree. Unalignable regions have been excluded. Bayesian posterior probabilities are above branches and RAxML bootstrap support percentages are below branches.
doi:10.1371/journal.pone.0080858.g004

Discussion

Despite their distinct shell morphology and longstanding cultural and commercial significance, our data reveal cryptic diversity within giant clams. Here, we find a previously undescribed clade of *Tridacna* (*Tridacna sp.*). This clade is supported by both mtDNA and nuclear gene regions (Figs. 2–4), which identify it as a unique, evolutionarily significant unit [49] with reference to previously described species. Our molecular phylogenetic analyses place *T. sp.* as a sister clade to *T. squamosa* and/or *T. crocea*, but in no instance was a close relationship between *T. sp.* and *T. maxima* suggested in our gene trees. Thus, molecular data do not support *T. sp.* being a variety of *T. maxima* as was suggested by Tang [16]. Clams with *T. sp.* mitotypes were found both at Ningaloo Reef in western Australia and in the Solomon Islands. Although only *T. sp.* and *T. squamosa* were identified among our clam samples from Ningaloo, it is likely that *T. maxima* also occur at Ningaloo (Penny unpub., [50]), and we found *T. sp.* sympatric with *T. maxima* and *T. crocea* in the Solomons.

The *T. sp.* clade includes the single haplotype (*COI* and a *16S*) described from Taiwan [16]. Tang *et al.* (2005) suggested that there are morphological differences between *T. sp.* and *T. maxima*, including mantle pattern, shell lip shape, posterior adductor weight and the position of the incurrent aperture. Qualitative examination of an individual from Ningaloo Reef with *T. sp.* mtDNA shows shell characters typical of *T. maxima*: asymmetry of the valve with posterior elongation and dense rows of scales on folds (Fig. 6). *T. maxima* is well known for its morphological variability [51] and thus it is possible that previous morphological examinations of *T. sp* may have been identified it as *T. maxima*. (Additional morphological samples are not presently available as most collecting permits only allow non-lethal sampling of giant clams.) Our findings, therefore, lend support to Tang's conclusion that *T. sp.* is an undescribed species but we show that, rather than being a narrow-range endemic (such as *Tridacna rosewateri* from Mauritius [10]), *T. sp.* is widely distributed. Although it is not possible at present to delineate the distribution of *T. sp.*, it seems probable that *T. sp.* occurs at locations in between Australia, Taiwan and the Solomon Islands. *T. sp.* individuals from the western Pacific were reciprocally monophyletic from the individuals from Ningaloo (Indian Ocean) and the single sequence from Taiwan (Fig. 5).

MtDNA genealogies place *T. sp* as sister species to *T. crocea* and *T. squamosa*, with strong support for monophyly of this group of three species (Figs. 2 and 3). *Tridacna maxima* and *T. squamosina* formed a second clade, but with less support across phylogenetic analyses (Fig. 2) probably because only *16S* sequences were available for *T. squamosina*. Monophyly of *T. crocea* and *T. squamosa* was reported in previous mtDNA based phylogenetic analyses [9,15], but not in allozyme analyses [14] where *T. squamosa* was sister to *T. crocea* and *T. maxima*. Monophyly of the *Chametrachea* subgenus (including *T. squamosina*, *T. crocea*, *T. maxima*, *T. sp.* and *T. squamosa*) [15,44] was supported in individual gene analyses and the concatenated StarBEAST searches (Fig. 2). Monophyly of the *Tridacna* subgenus (including *T. derasa*, *T. mbalavuana*, and *T. gigas*) was not well supported in any of our mtDNA analyses, with these taxa appearing basal to the *Chametrachea*, but missing and non-overlapping data may have contributed to the low resolution.

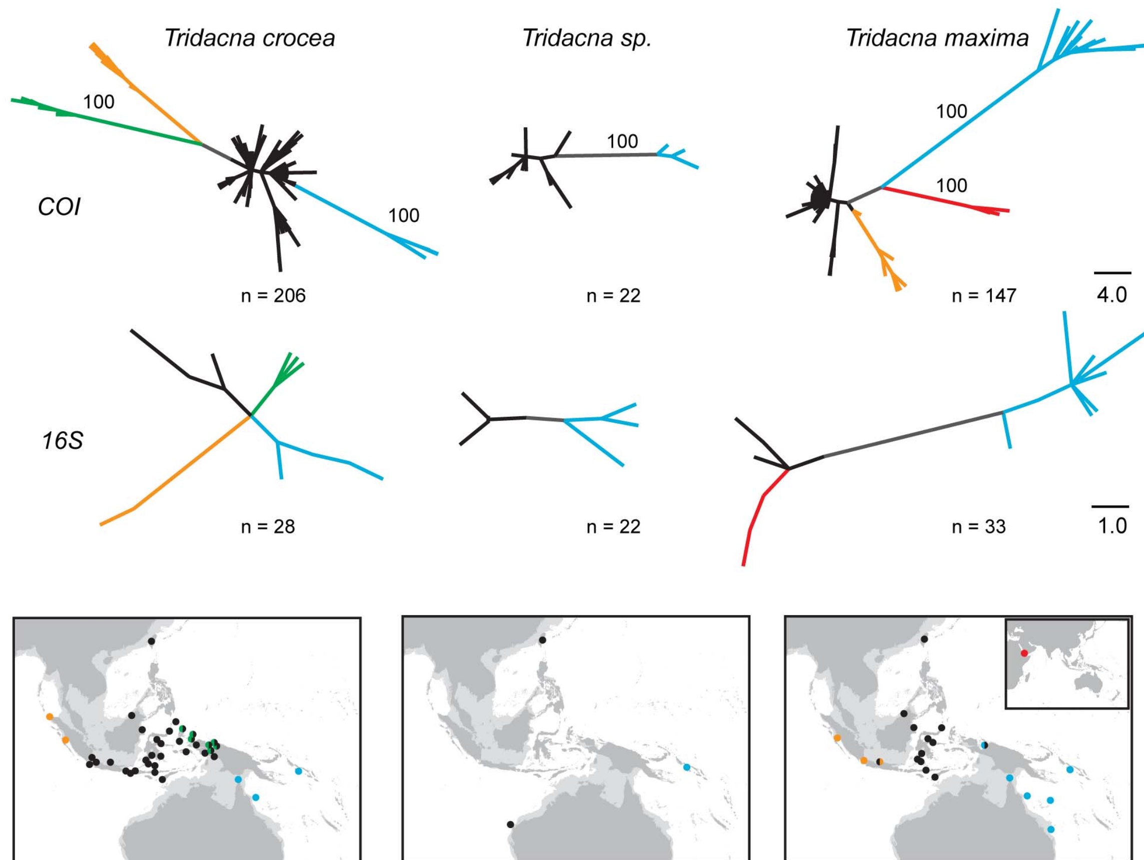


Figure 5. Unrooted parsimony trees and sampling locations for *Tridacna crocea*, *Tridacna sp.*, and *Tridacna maxima*. Major lineages on networks are colored and the geographic extent of each lineage is indicated on the map. Relative frequencies of each haplotype are not depicted; each haplotype is shown in equal size (see text). Dots on maps indicate sampling locations and locations with two distinct sympatric lineages are shown as bisected circles (not indicative of relative frequencies). Support for monophyly of major clades among *COI* trees is based on 100 percent consistency of each branch among all equally parsimonious trees (a randomly chosen tree is depicted). Among *16S* trees, the single most parsimonious tree for *T. sp.* and *T. maxima* are shown, and for *T. crocea* both green and blue lineages were present in all six equally parsimonious trees. Colors indicate geographic locations of haplotypes and internal branches are in gray.
doi:10.1371/journal.pone.0080858.g005

Previous phylogeographic studies of *T. crocea* [24,26,29] and *T. maxima* [25] from Indonesia show geographic restriction of several clades. The mtDNA gene trees within these papers delineate clusters comprising haplotypes from western Sumatra (Sunda), Wallacea, and northwest New Guinea (Sahul) [24–26,29] with some mixing between clades particularly in the Bird's Head Peninsula of northwest New Guinea [26]. Our samples showed an additional and deeper evolutionary break for *T. crocea* to the east of Cenderawasih Bay, whereby individuals from the Solomon Islands, Torres Strait, and Great Barrier Reef form a monophyletic group and do not share any mtDNA haplotypes with northwest New Guinea or locations in Wallacea (Fig. 5). Therefore, it appears that the distinct clade of *T. crocea* haplotypes from northwest New Guinea (with some spillover westward into Wallacea [26]) is regionally endemic and does not extend into the west Pacific. These patterns are not due to differences in DNA sequencing interpretation between research groups, as samples (from [24,26,41]) are mutually consistent and a single *T. crocea* (from [15]) falls within the larger Pacific *T. crocea* clade. Based on present sampling, we can place this newly discovered genetic discontinuity between Cenderawasih Bay and the Solomon Islands in the north and between the Aru Basin and Torres Strait in the south. For *T. maxima*, in contrast, the distinct haplotypes from northwest New Guinea fall in the same clade as west Pacific haplotypes. Thus the northwest New Guinea clade of *T. maxima* can now be viewed as a westward extension of Pacific variants, albeit with no shared haplotypes between locations.

With only two species to compare, we can only speculate as to why the mtDNA patterns differ between species, although greater overall population genetic structure in *T. crocea* compared to *T. maxima* is consistent with previously co-sampled regions (for instance, [24] in comparison to [25]). Because of the diffuse sampling for *T. crocea*, we cannot pinpoint a specific location of geographic differentiation east of Cenderawasih Bay, yet at a macroscale this observation is consistent with mtDNA patterns in a butterflyfish [52], a reef fish [53], and a sea star (Crandall pers. comm.) and may be associated with a long stretch (>700 km) of coastline east of Cenderawasih Bay with sparse reef habitat [54].

In *T. maxima*, we found that Solomon Islands haplotypes cluster with haplotypes from the Great Barrier Reef; this affinity contrasts with allozyme results that show substantial divergence between Solomon Islands and Great Barrier Reef populations [33]. The nature of these differing patterns cannot be explored further as allozyme results are not directly comparable across research groups.

The broadscale geographic and multispecies phylogenetic results of this study, consolidated with those of previous investigations, reveal new aspects of regional patterns and

Figure 6. An individual with *Tridacna sp.* mtDNA demonstrating valve morphology consistent with *Tridacna maxima*. A) *Tridacna maxima* from Hibernia Reef, WA, Australia. Accession No# P.52722 (Museum Art Gallery Northern Territory (MAGNT)), original identification based on morphology, B) *Tridacna sp.* from Five Finger Reef, South of Coral Bay, Ningaloo Marine Park, WA, Australia. Accession No#. P.51911 (Museum Art Gallery Northern Territory (MAGNT)), C) *Tridacna maxima*, from north western WA, Australia, unregistered (Museum Art Gallery Northern Territory (MAGNT)). Photo credit: Shane Penny.
doi:10.1371/journal.pone.0080858.g006

highlight key uncertainties in the current knowledge of *Tridacna*. A common result among population genetic studies of *Tridacna* species to date is that there is substantial population structure. Such genetic differentiation may be due in part to the relatively short planktonic larval duration of approximately 9 days [12] that is likely to restrict dispersal distances. The discovery of an undescribed species adds to other recent species discoveries in *Tridacna* [9–11], but the broad distribution of *T. sp.* illustrates that cryptic species can remain undetected even in such conspicuous groups as giant clams.

Both the discovery of a new species and the observation of substantial geographic differentiation are relevant to monitoring of local stocks and human transport of clams. First, the presence of a cryptic sympatric species would result in overestimates of species

abundance where clam populations are censused. Second, human-aided movements could cause species to be introduced to regions outside their natural range and, similarly, are likely to introduce foreign genetic material into local populations. *Tridacna maxima*, *T. squamosa*, *T. derasa*, *T. mbalavuana* and *T. gigas* were frequently translocated during the 1980's and 1990's (some human assisted movements continuing into this century) by governmental, commercial and conservation organizations to combat local depletion and facilitate the live culture trade [55]. Third, depleted populations are unlikely to receive immigrants from geographically distant locations via planktonic dispersal and, therefore, recovery may be slow or negligible even when local harvesting has ceased. Results from giant clams underscore two important themes emerging from genetic investigations of marine organisms: cryptic species are common [19,20,56,57], and many species are genetically heterogeneous across their geographic range [58].

Supporting Information

Document S1 Genbank accession numbers for all included sequences.
(XLS)

Acknowledgments

Sampling in the Coral Sea was supported by the Marine Division of the Australian Government Department of Sustainability, Environment, Water, Population & Communities. We are grateful to the staff of the Australian Museum's Lizard Island Research Station and the Heron Island Research Station for their facilities and support. Sampling in the Torres Strait Islands was assisted by the staff and students of Tagai State College, Thursday Island Primary and the Torres Strait Regional Authority. Sampling in the Solomon Islands was made possible via the Pacific Strategy Assistance Program within the Australian Government Department of Climate Change and Energy Efficiency and with the assistance of the Roviana Conservation Foundation. We especially thank JD Aguirre, S Albert, A Denzin, N Gemmell, M Jimuru, F MacGregor, V McGrath (Senior Community Liaison Officer, Land and Sea Management Unit, Torres Strait Regional Authority), A Mirams, R Pearce, Stephen, Lavud and Takenda for their logistical support and field assistance. JS Lucas, LG Cook, A Toon, L Pope and JM Pandolfi provided helpful comments and suggestions, as did several anonymous reviewers.

Author Contributions

Conceived and designed the experiments: TH JK LL SP EAT CR. Performed the experiments: TH JK LL SP EAT CR. Analyzed the data: TH JK SP CR. Contributed reagents/materials/analysis tools: TH JK SP EAT CR. Wrote the paper: CR JK TH LL.

References

1. Heslinga G (1996) Clams to cash: How to make and sell giant clam shell products. Hawaii, USA: Center for Tropical and Subtropical Aquaculture. Publication Number 125: 47.
2. Wabnitz C, Taylor M, Green E, Razak T (2003) From ocean to aquarium: The global trade in marine ornamental species. Cambridge, UK: UNEP World Conservation Monitoring Centre. 64 p.
3. Tisdell C, Menz KM (1992) Giant clam farming and sustainable development: An overview. In: Tisdell C, editor. Giant clams in the sustainable development of the South Pacific: Socioeconomic issues in mariculture and conservation. Canberra, Australia: ACIAR pp. 3–14.
4. Tomascik T, Mah AJ, Nontji A, Moosa MK (1997) The ecology of the Indonesian seas: Part one. Hong Kong: Periplus Editions. 656 p.
5. Wells SM (1997) Giant clams: Status, trade and mariculture, and the role of CITES in management. Gland, Switzerland and Cambridge, UK: IUCN-The World Conservation Union. 77 p.
6. UNEP-WCMC (2013) Convention on international trade in endangered species of wild fauna and flora. Available: www.cites.org. Accessed 2013 October 20.
7. IUCN (2013) IUCN red list of threatened species. Available: www.IUCNredlist.org. Accessed 2013 October 20.
8. bin Othman AS, Goh GHS, Todd PA (2010) The distribution and status of giant clams (family Tridacnidae) - a short review. Raffles Bull Zool 58: 103–111.
9. Richter C, Roa-Quiaoit H, Jantzen C, Al-Zibdah M, Kochzius M (2008) Collapse of a new living species of giant clam in the Red Sea. Curr Biol 18: 1349–1354.
10. Sirenko BI, Scarlato OA (1991) *Tridacna rosewateri* sp. n., a new species of giant clam from the Indian Ocean. La Conchiglia 22: 4–9.
11. Lucas JS, Ledua E, Braley RD (1991) *Tridacna tevoroa* Lucas, Ledua and Braley: A recently-described species of giant clam (Bivalvia; Tridacnidae) from Fiji and Tonga. Nautilus 105: 92–103.
12. Lucas JS (1988) Giant clams: Description, distribution and life history. In: Copland JW, LucasJS, editor. Giant clams in Asia and the Pacific. pp. 21–32.
13. Maruyama T, Ishikura M, Yamazaki S, Kanai S (1998) Molecular phylogeny of zooxanthellate bivalves. Biol Bull 195: 70–77.
14. Benzie JAH, Williams ST (1998) Phylogenetic relationships among giant clam species (Mollusca: Tridacnidae) determined by protein electrophoresis. Mar Biol 132: 123–133.
15. Schneider JA, O'Foighil D (1999) Phylogeny of giant clams (Cardiidae: Tridacninae) based on partial mitochondrial 16S rDNA gene sequences. Mol Phyl Evol 13: 59–66.
16. Tang YC (2005) The systematic status of *Tridacna maxima* (Bivalvia: Tridacnidae) based on morphological and molecular evidence. Master's thesis. Keelung, Taiwan, Republic of China: National Taiwan Ocean University. 80 p.
17. Hoeksema B (2007) Delineation of the Indo-Malayan centre of maximum marine biodiversity: The Coral Triangle. In: Renema W, editor. Biogeography, time, and place: Distributions, barriers, and islands. pp. 117–178.
18. Ekman S (1953) Zoogeography of the sea. London: Sidgwick and Jackson. 417 p.
19. Vogler C, Benzie J, Lessios H, Barber P, Wörheide G (2008) A threat to coral reefs multiplied? Four species of crown-of-thorns starfish. Biol Lett 4: 696–699.
20. Duda TF, Kohn AJ, Matheny AM (2009) Cryptic species differentiated in *Conus ebraeus*, a widespread tropical marine gastropod. Biol. Bull. 217:292–305
21. Carpenter KE, Barber PH, Crandall ED, Ablan-Lagman MCA, Ambariyanto, et al. (2010) Comparative phylogeography of the Coral Triangle and implications for marine management. J Mar Biol 2011: 1–14.
22. Barber PH, Erdmann MV, Palumbi SR (2006) Comparative phylogeography of three codistributed stomatopods: Origins and timing of regional lineage diversification in the Coral Triangle. Evolution 60: 1825–1839.
23. Voris HK (2000) Maps of Pleistocene sea levels in southeast Asia: Shorelines, river systems and time durations. J Biogeogr 27: 1153–1167.
24. Kochzius M, Nuryanto A (2008) Strong genetic population structure in the boring giant clam, *Tridacna crocea*, across the Indo-Malay archipelago: Implications related to evolutionary processes and connectivity. Mol Ecol 17: 3775–3787.
25. Nuryanto A, Kochzius M (2009) Highly restricted gene flow and deep evolutionary lineages in the giant clam *Tridacna maxima*. Coral Reefs 28: 607–619.
26. DeBoer TS, Subia MD, Ambariyanto, Erdmann MV, Kovitvongsa K, et al. (2008) Phylogeography and limited genetic connectivity in the endangered boring giant clam across the Coral Triangle. Cons Biol 22: 1255–1266.
27. Ravago-Gotanco RG, Magsino RM, Juinio-Meñez MA (2007) Influence of the North Equatorial Current on the population genetic structure of *Tridacna crocea* (Mollusca: Tridacnidae) along the eastern Philippine seaboard. Mar Ecol Prog Ser 336: 161–168.
28. Juinio-Meñez MA, Magsino RM, Ravago-Gotanco RG, Yu ET (2003) Genetic structure of *Linckia laevigata* and *Tridacna crocea* populations in the Palawan shelf and shoal reefs. Mar Biol 142: 717–726.
29. DeBoer TS, Naguit MR, Erdmann MV, Ablan-Lagman MCA, Ambariyanto, et al. (2014) Concordant phylogenetic patterns inferred from mitochondrial and microsatellite DNA in the giant clam *Tridacna crocea*. Bull Mar Sci in press.
30. Macaranas JM, Ablan CA, Pante MJR, Benzie JAH, Williams ST (1992) Genetic structure of giant clam (*Tridacna derasa*) populations from reefs in the Indo-Pacific. Mar Biol 113: 231–238.
31. Laurant V, Planes S, Salvat B (2002) High variability of genetic pattern in giant clam (*Tridacna maxima*) populations within French Polynesia. Biol J Linn Soc 77: 221–231.
32. Benzie JAH, Williams ST (1995) Gene flow among giant clam (*Tridacna gigas*) populations in the Pacific does not parallel ocean circulation. Mar Biol 123: 781–787.
33. Benzie JAH, Williams ST (1997) Genetic structure of giant clam (*Tridacna maxima*) populations in the west Pacific is not consistent with dispersal by present-day ocean currents. Evolution 51: 768–783.
34. Benzie JAH, Williams ST (1992) No genetic differentiation of giant clam (*Tridacna gigas*) populations in the Great Barrier Reef, Australia. Mar Biol 113: 373–377.
35. Benzie JAH, Williams ST (1992) Genetic structure of giant clam (*Tridacna maxima*) populations from reefs in the western Coral Sea. Coral Reefs 11: 135–141.
36. Gardner JPA, Boesche C, Meyer JM, Wood AR (2012) Analyses of DNA obtained from shells and brine-preserved meat of the giant clam *Tridacna maxima* from the central Pacific Ocean. Mar Ecol Prog Ser 453: 297–301.
37. Huelsken T, Schreiber S, Hollmann M (2011) COI amplification success from mucus-rich marine gastropods (Gastropoda: Naticidae) depends on DNA

extraction method and preserving agent. Mitteilungen der Deutschen Malakozoologischen Gesellschaft 85: 17–26.
38. 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 Biotech 3: 294–299.
39. Palumbi S (1996) Nucleic acids ii: The polymerase chain reaction. In: Hillis DM Moritz C, Mable BK, editors. Molecular systematics. Sunderland, Massachusetts: Sinauer & Associates Inc. pp. 205–247.
40. Yu ET, Juinio-Menez MA, Monje VD (2000) Sequence variation in the ribosomal DNA internal transcribed spacer of *Tridacna crocea*. Mar Biotechnol 2: 511–516.
41. Neo ML, Todd PA (2012) Population density and genetic structure of the giant clams *Tridacna crocea* and *T. squamosa* on Singapore's reefs. Aquatic Biology 14: 265–275.
42. Rambaut AE (1996) Se-Al: Sequence alignment editor v2.0a11. Available: http://tree.bio.ed.ac.uk/software/. Accessed 2013 October 20.
43. Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data Mol Biol Evol 27: 570–580.
44. Harzhauser M, Mandic O, Piller WE, Reuter M, Kroh A (2008) Tracing back the origin of the Indo-Pacific mollusc fauna: Basal Tridacninae from the Oligocene and Miocene of the Sultanate of Oman. Palaeontology 51: 199–213.
45. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.
46. Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web-servers. Syst Biol 57: 758–771.
47. Swofford DL (1998) PAUP*, phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland, Massachusetts: Sinauer Associates.
48. Plazzi F, Ceregato A, Taviani M, Passamonti M (2011) A molecular phylogeny of bivalve mollusks: Ancient radiations and divergences as revealed by mitochondrial genes. PLoSOne 6: e27147.
49. Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. Ann Rev Ecol Syst 18: 269–292.
50. Black R, Johnson MS, Prince J, Brearley A, Bond T (2011) Evidence of large, local variations in recruitment and mortality in the small giant clam, *Tridacna maxima*, at Ningaloo Marine Park, Western Australia. Mar FW Res 62: 1318–1326.
51. Rosewater J (1965) The family Tridacnidae in the Indo-Pacific. Mollusc Res 1: 347–396.
52. McMillan WO, Weigt LA, Palumbi SR (1999) Color pattern evolution, assortative mating, and genetic differentiation in brightly colored butterflyfishes (Chaetodontidae). Evolution 53: 247–260.
53. Sims CA, Riginos C, Blomberg SP, Huelsken T, Drew J, et al. (2014) Cleaning up the biogeography of *Labroides dimidiatus* using phylogenetics and morphometrics. Coral Reefs: in press.
54. Spalding MD, Fox HE, Allen GE, Davidson N, Ferdaña ZA, et al. (2007) Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. Bioscience 57: 573–583.
55. Kinch J, Teitelbaum A (2009) Proceedings of the regional workshop on the management of sustainable fisheries for giant clams (Tridacnidae) and CITES capacity building. Nadi, Fiji: Secretariat of the Pacific Community. 54 p.
56. Bickford D, Lohman DJ, Sodhi NS, Ng PK, Meier R, et al. (2006) Cryptic species as a window on diversity and conservation. Tr Ecol Evol 22: 148–155.
57. Knowlton N (1993) Sibling species in the sea. Ann Rev Ecol Syst 24: 189–216.
58. Hellberg ME (2009) Gene flow and isolation among populations of marine animals. Ann Rev Ecol Syst 40: 291–310.

Supporting Information
Document S1. Genbank accession numbers for all included sequences.

COI						
Study Source	**Acc No**	**Locality**	**Name used in study**	**Putative species**		
Tang unpublished	DQ269479	Taiwan		crocea		
Nuryanto et al. 2007	EU003606	W. Sulawesi	Spermonde01	crocea		
Nuryanto et al. 2007	EU003607	W. Sulawesi	Spermonde02	crocea		
Nuryanto et al. 2007	EU003608	W. Java	PulauSeribu01	crocea		
Nuryanto et al. 2007	EU003609	W. Papua	Biak01	crocea		
DeBoer et al. 2008	EU341350	Indonesia	TSD_20	crocea		
DeBoer et al. 2008	EU341351	Indonesia	TSD_21	crocea		
DeBoer et al. 2008	EU341352	Indonesia	TSD_22	crocea		
DeBoer et al. 2008	EU341353	Indonesia	TSD_23	crocea		
DeBoer et al. 2008	EU341354	Indonesia	TSD_24	crocea		
DeBoer et al. 2008	EU341355	Indonesia	TSD_25	crocea		
DeBoer et al. 2008	EU341356	Indonesia	TSD_26	crocea		
DeBoer et al. 2008	EU341357	Indonesia	TSD_27	crocea		
DeBoer et al. 2008	EU341358	Indonesia	TSD_28	crocea		
DeBoer et al. 2008	EU341359	Indonesia	TSD_29	crocea		
DeBoer et al. 2008	EU341360	Indonesia	TSD_30	crocea		
DeBoer et al. 2008	EU341361	Indonesia	TSD_31	crocea		
DeBoer et al. 2008	EU341362	Indonesia	TSD_32	crocea		
DeBoer et al. 2008	EU341363	Indonesia	TSD_33	crocea		
DeBoer et al. 2008	EU341364	Indonesia	TSD_34	crocea		
DeBoer et al. 2008	EU341365	Indonesia	TSD_35	crocea		
DeBoer et al. 2008	EU341366	Indonesia	TSD_36	crocea		
DeBoer et al. 2008	EU341367	Indonesia	TSD_37	crocea		
DeBoer et al. 2008	EU341368	Indonesia	TSD_38	crocea		
DeBoer et al. 2008	EU341369	Indonesia	TSD_39	crocea		
DeBoer et al. 2008	EU341370	Indonesia	TSD_40	crocea		
DeBoer et al. 2008	EU341371	Indonesia	TSD_41	crocea		
DeBoer et al. 2008	EU341372	Indonesia	TSD_42	crocea		
DeBoer et al. 2008	EU341373	Indonesia	TSD_43	crocea		
DeBoer et al. 2008	EU341374	Indonesia	TSD_44	crocea		
DeBoer et al. 2008	EU341375	Indonesia	TSD_45	crocea		
DeBoer et al. 2008	EU341376	Indonesia	TSD_46	crocea		
DeBoer et al. 2008	EU341377	Indonesia	TSD_47	crocea		
DeBoer et al. 2008	EU341378	Indonesia	TSD_48	crocea		
DeBoer et al. 2008	EU341379	Indonesia	TSD_49	crocea		
Kochzius & Nuryanto 2008	FM253431	Indonesia	Hap_1	crocea		
Kochzius & Nuryanto 2008	FM253432	Biak	Hap_2_yell	crocea		
Kochzius & Nuryanto 2008	FM253433	Biak	Hap_3_yell	crocea		
Kochzius & Nuryanto 2008	FM253434	Indonesia	Hap_4	crocea		
Kochzius & Nuryanto 2008	FM253435	Indonesia	Hap_5	crocea		
Kochzius & Nuryanto 2008	FM253436	Indonesia	Hap_6	crocea		
Kochzius & Nuryanto 2008	FM253437	Biak	Hap_7_yell	crocea		
Kochzius & Nuryanto 2008	FM253438	Biak	Hap_8_yell	crocea		
Kochzius & Nuryanto 2008	FM253439	Indonesia	Hap_9	crocea		
Kochzius & Nuryanto 2008	FM253440	Indonesia	Hap_10	crocea		
Kochzius & Nuryanto 2008	FM253441	Biak	Hap_11_yell	crocea		
Kochzius & Nuryanto 2008	FM253442	Indonesia	Hap_12	crocea		
Kochzius & Nuryanto 2008	FM253443	Indonesia	Hap_13	crocea		
Kochzius & Nuryanto 2008	FM253444	Indonesia	Hap_14	crocea		
Kochzius & Nuryanto 2008	FM253445	Indonesia	Hap_15	crocea		
Kochzius & Nuryanto 2008	FM253446	Indonesia	Hap_16	crocea		
Kochzius & Nuryanto 2008	FM253447	Indonesia	Hap_17	crocea		
Kochzius & Nuryanto 2008	FM253448	Indonesia	Hap_18	crocea		
Kochzius & Nuryanto 2008	FM253449	Indonesia	Hap_19	crocea		
Kochzius & Nuryanto 2008	FM253450	Indonesia	Hap_20	crocea		
Kochzius & Nuryanto 2008	FM253451	Indonesia	Hap_21	crocea		
Kochzius & Nuryanto 2008	FM253452	Indonesia	Hap_22	crocea		
Kochzius & Nuryanto 2008	FM253453	Indonesia	Hap_23	crocea		
Kochzius & Nuryanto 2008	FM253454	Indonesia	Hap_24	crocea		
Kochzius & Nuryanto 2008	FM253455	Indonesia	Hap_25	crocea		
Kochzius & Nuryanto 2008	FM253456	Indonesia	Hap_26	crocea		
Kochzius & Nuryanto 2008	FM253457	Indonesia	Hap_27	crocea		
Kochzius & Nuryanto 2008	FM253458	Indonesia	Hap_28	crocea		
Kochzius & Nuryanto 2008	FM253459	Indonesia	Hap_29	crocea		
Kochzius & Nuryanto 2008	FM253460	Indonesia	Hap_30	crocea		
Kochzius & Nuryanto 2008	FM253461	Indonesia	Hap_31	crocea		
Kochzius & Nuryanto 2008	FM253462	Indonesia	Hap_32	crocea		

Kochzius & Nuryanto 2008	FM253463	Indonesia	Hap_33	crocea		
Kochzius & Nuryanto 2008	FM253464	Indonesia	Hap_34	crocea		
Kochzius & Nuryanto 2008	FM253465	Indonesia	Hap_35	crocea		
Kochzius & Nuryanto 2008	FM253466	Indonesia	Hap_36	crocea		
Kochzius & Nuryanto 2008	FM253467	Indonesia	Hap_37	crocea		
Kochzius & Nuryanto 2008	FM253468	Indonesia	Hap_38	crocea		
Kochzius & Nuryanto 2008	FM253469	Indonesia	Hap_39	crocea		
Kochzius & Nuryanto 2008	FM253470	Indonesia	Hap_40	crocea		
Kochzius & Nuryanto 2008	FM253471	Indonesia	Hap_41	crocea		
Kochzius & Nuryanto 2008	FM253472	Indonesia	Hap_42	crocea		
Kochzius & Nuryanto 2008	FM253473	Indonesia	Hap_43	crocea		
Kochzius & Nuryanto 2008	FM253474	Indonesia	Hap_44	crocea		
Kochzius & Nuryanto 2008	FM253475	Indonesia	Hap_45	crocea		
Kochzius & Nuryanto 2008	FM253476	Indonesia	Hap_46	crocea		
Kochzius & Nuryanto 2008	FM253477	Indonesia	Hap_47	crocea		
Kochzius & Nuryanto 2008	FM253478	Indonesia	Hap_48	crocea		
Kochzius & Nuryanto 2008	FM253479	Indonesia	Hap_49	crocea		
Kochzius & Nuryanto 2008	FM253480	Indonesia	Hap_50	crocea		
Kochzius & Nuryanto 2008	FM253481	Indonesia	Hap_51	crocea		
Kochzius & Nuryanto 2008	FM253482	Indonesia	Hap_52	crocea		
Kochzius & Nuryanto 2008	FM253483	Indonesia	Hap_53	crocea		
Kochzius & Nuryanto 2008	FM253484	Indonesia	Hap_54	crocea		
Kochzius & Nuryanto 2008	FM253485	Indonesia	Hap_55	crocea		
Kochzius & Nuryanto 2008	FM253486	Indonesia	Hap_56	crocea		
Kochzius & Nuryanto 2008	FM253487	Indonesia	Hap_57	crocea		
Kochzius & Nuryanto 2008	FM253488	Indonesia	Hap_58	crocea		
Kochzius & Nuryanto 2008	FM253489	Indonesia	Hap_59	crocea		
Kochzius & Nuryanto 2008	FM253490	Indonesia	Hap_60	crocea		
Kochzius & Nuryanto 2008	FM253491	Indonesia	Hap_61	crocea		
Kochzius & Nuryanto 2008	FM253492	Indonesia	Hap_62	crocea		
Kochzius & Nuryanto 2008	FM253493	Indonesia	Hap_63	crocea		
Kochzius & Nuryanto 2008	FM253494	Indonesia	Hap_64	crocea		
Kochzius & Nuryanto 2008	FM253495	Indonesia	Hap_65	crocea		
Kochzius & Nuryanto 2008	FM253496	Indonesia	Hap_66	crocea		
Kochzius & Nuryanto 2008	FM253497	Indonesia	Hap_67	crocea		
Kochzius & Nuryanto 2008	FM253498	Indonesia	Hap_68	crocea		
Kochzius & Nuryanto 2008	FM253499	Indonesia	Hap_69	crocea		
Kochzius & Nuryanto 2008	FM253500	Indonesia	Hap_70	crocea		
Kochzius & Nuryanto 2008	FM253501	Indonesia	Hap_71	crocea		
Kochzius & Nuryanto 2008	FM253502	Indonesia	Hap_72	crocea		
Kochzius & Nuryanto 2008	FM253503	Indonesia	Hap_73	crocea		
Kochzius & Nuryanto 2008	FM253504	Indonesia	Hap_74	crocea		
Kochzius & Nuryanto 2008	FM253505	Indonesia	Hap_75	crocea		
Kochzius & Nuryanto 2008	FM253506	Indonesia	Hap_76	crocea		
Kochzius & Nuryanto 2008	FM253507	Indonesia	Hap_77	crocea		
Kochzius & Nuryanto 2008	FM253508	Indonesia	Hap_78	crocea		
Kochzius & Nuryanto 2008	FM253509	Indonesia	Hap_79	crocea		
Kochzius & Nuryanto 2008	FM253510	Indonesia	Hap_80	crocea		
Kochzius & Nuryanto 2008	FM253511	Indonesia	Hap_81	crocea		
Kochzius & Nuryanto 2008	FM253512	Indonesia	Hap_82	crocea		
Kochzius & Nuryanto 2008	FM253513	Indonesia	Hap_83	crocea		
Kochzius & Nuryanto 2008	FM253514	Indonesia	Hap_84	crocea		
Kochzius & Nuryanto 2008	FM253515	Indonesia	Hap_85	crocea		
Kochzius & Nuryanto 2008	FM253516	Indonesia	Hap_86	crocea		
Kochzius & Nuryanto 2008	FM253517	Indonesia	Hap_87	crocea		
Kochzius & Nuryanto 2008	FM253518	Indonesia	Hap_88	crocea		
Kochzius & Nuryanto 2008	FM253519	Indonesia	Hap_89	crocea		
Kochzius & Nuryanto 2008	FM253520	Indonesia	Hap_90	crocea		
Kochzius & Nuryanto 2008	FM253521	Indonesia	Hap_91	crocea		
Kochzius & Nuryanto 2008	FM253522	Indonesia	Hap_92	crocea		
Kochzius & Nuryanto 2008	FM253523	Indonesia	Hap_93	crocea		
Kochzius & Nuryanto 2008	FM253524	Indonesia	Hap_94	crocea		
Kochzius & Nuryanto 2008	FM253525	Sumatra	Hap_95_Red	crocea		
Kochzius & Nuryanto 2008	FM253526	Sumatra	Hap_96_Red	crocea		
Kochzius & Nuryanto 2008	FM253527	Sumatra	Hap_97_Red	crocea		
Kochzius & Nuryanto 2008	FM253528	Sumatra	Hap_98_Red	crocea		
Kochzius & Nuryanto 2008	FM253529	Sumatra	Hap_99_Red	crocea		
Kochzius & Nuryanto 2008	FM253530	Indonesia	Hap_100	crocea		
Kochzius & Nuryanto 2008	FM253531	Indonesia	Hap_101	crocea		
Kochzius & Nuryanto 2008	FM253532	Indonesia	Hap_102	crocea		

Kochzius & Nuryanto 2008	FM253533	Indonesia	Hap_103	crocea		
Kochzius & Nuryanto 2008	FM253534	Indonesia	Hap_104	crocea		
Kochzius & Nuryanto 2008	FM253535	Indonesia	Hap_105	crocea		
Kochzius & Nuryanto 2008	FM253536	Indonesia	Hap_106	crocea		
Kochzius & Nuryanto 2008	FM253537	Indonesia	Hap_107	crocea		
Kochzius & Nuryanto 2008	FM253538	Indonesia	Hap_108	crocea		
Kochzius & Nuryanto 2008	FM253539	Indonesia	Hap_109	crocea		
Kochzius & Nuryanto 2008	FM253540	Indonesia	Hap_110	crocea		
Kochzius & Nuryanto 2008	FM253541	Indonesia	Hap_111	crocea		
Kochzius & Nuryanto 2008	FM253542	Indonesia	Hap_112	crocea		
Kochzius & Nuryanto 2008	FM253543	Indonesia	Hap_113	crocea		
Kochzius & Nuryanto 2008	FM253544	Indonesia	Hap_114	crocea		
Kochzius & Nuryanto 2008	FM253545	Indonesia	Hap_115	crocea		
Kochzius & Nuryanto 2008	FM253546	Indonesia	Hap_116	crocea		
Kochzius & Nuryanto 2008	FM253547	Indonesia	Hap_117	crocea		
Kochzius & Nuryanto 2008	FM253548	Indonesia	Hap_118	crocea		
Kochzius & Nuryanto 2008	FM253549	Indonesia	Hap_119	crocea		
Kochzius & Nuryanto 2008	FM253550	Indonesia	Hap_120	crocea		
Kochzius & Nuryanto 2008	FM253551	Indonesia	Hap_121	crocea		
Kochzius & Nuryanto 2008	FM253552	Indonesia	Hap_122	crocea		
Kochzius & Nuryanto 2008	FM253553	Indonesia	Hap_123	crocea		
Kochzius & Nuryanto 2008	FM253554	Indonesia	Hap_124	crocea		
Kochzius & Nuryanto 2008	FM253555	Indonesia	Hap_125	crocea		
Kochzius & Nuryanto 2008	FM253556	Indonesia	Hap_126	crocea		
Kochzius & Nuryanto 2008	FM253557	Indonesia	Hap_127	crocea		
Kochzius & Nuryanto 2008	FM253558	Indonesia	Hap_128	crocea		
Kochzius & Nuryanto 2008	FM253559	Indonesia	Hap_129	crocea		
Kochzius & Nuryanto 2008	FM253560	Indonesia	Hap_130	crocea		
Kochzius & Nuryanto 2008	FM253561	Indonesia	Hap_131	crocea		
Kochzius & Nuryanto 2008	FM253562	Indonesia	Hap_132	crocea		
Neo & Todd 2012	JN392040	Singapore	NML_32	crocea		
Neo & Todd 2012	JN392041	Singapore	NML_56	crocea		
Neo & Todd 2012	JN392042	Singapore	NML_66	crocea		
Neo & Todd 2012	JN392043	Singapore	NML_41	crocea		
Neo & Todd 2012	JN392044	Singapore	NML_30	crocea		
Neo & Todd 2012	JN392045	Singapore	NML_43	crocea		
Neo & Todd 2012	JN392046	Singapore	NML_29	crocea		
Neo & Todd 2012	JN392047	Singapore	NML_31	crocea		
Neo & Todd 2012	JN392048	Singapore	NML_39	crocea		
Neo & Todd 2012	JN392049	Singapore	NML_57	crocea		
Neo & Todd 2012	JN392050	Singapore	NML_59	crocea		
Neo & Todd 2012	JN392051	Singapore	NML_64	crocea		
Neo & Todd 2012	JN392052	Singapore	NML_58	crocea		
Neo & Todd 2012	JN392053	Singapore	NML_67	crocea		
Neo & Todd 2012	JN392054	Singapore	NML_35	crocea		
Neo & Todd 2012	JN392055	Singapore	NML_42	crocea		
Neo & Todd 2012	JN392056	Singapore	NML_65	crocea		
Neo & Todd 2012	JN392057	Singapore	NML_34	crocea		
Neo & Todd 2012	JN392058	Singapore	NML_36	crocea		
Neo & Todd 2012	JN392059	Singapore	NML_60	crocea		
Neo & Todd 2012	JN392060	Singapore	NML_62	crocea		
Neo & Todd 2012	JN392061	Singapore	NML_40	crocea		
Neo & Todd 2012	JN392062	Singapore	NML_33	crocea		
Neo & Todd 2012	JN392063	Singapore	NML_63	crocea		
Neo & Todd 2012	JN392064	Singapore	NML_38	crocea		
Neo & Todd 2012	JN392065	Singapore	NML_37	crocea		
Neo & Todd 2012	JN392066	Singapore	NML_61	crocea		
Plazz & Passamont 2010	GQ166591	?		derasa		
Nuryanto et al. 2008	EU003616	E. Sulawesi (Togian)	Indonesia01	gigas		
Tang & Chen, unplublished	DQ155301	Taiwan	Tridacna maxima	maxima		
Nuryanto et al. 2007	EU003610	W. Sumatra	Padang01	maxima		
Nuryanto et al. 2007	EU003611	W. Sulawesi	Spermonde01	maxima		
Nuryanto et al. 2007	EU003612	E. Sulawesi	Togian01	maxima		
Nuryanto et al. 2007	EU003613	W. Papua	Biak01	maxima		
Nuryanto et al. 2007	EU003614	W. Java	PulauSeribu01	maxima		
DeBoer et al. unpublished	EU346364	Indonesia	TSD53	maxima		
DeBoer et al. unpublished	EU346365	Indonesia	TSD54	maxima		
DeBoer et al. unpublished	EU346366	Indonesia	TSD55	maxima		
DeBoer et al. unpublished	EU346367	Indonesia	TSD56	maxima		
DeBoer et al. unpublished	EU346368	Indonesia	TSD57	maxima		

Nuryanto & Kochzius 2009	FM244476	Red Sea	Hap_1_Red	maxima		
Nuryanto & Kochzius 2009	FM244477	Red Sea	Hap_2_Red	maxima		
Nuryanto & Kochzius 2009	FM244478	Red Sea	Hap_3_Red	maxima		
Nuryanto & Kochzius 2009	FM244479	Red Sea	Hap_4_Red	maxima		
Nuryanto & Kochzius 2009	FM244480	Red Sea	Hap_5_Red	maxima		
Nuryanto & Kochzius 2009	FM244481	Red Sea	Hap_6_Red	maxima		
Nuryanto & Kochzius 2009	FM244482	Red Sea	Hap_7_Red	maxima		
Nuryanto & Kochzius 2009	FM244483	Red Sea	Hap_8_Red	maxima		
Nuryanto & Kochzius 2009	FM244484	Red Sea	Hap_9_Red	maxima		
Nuryanto & Kochzius 2009	FM244485	Red Sea	Hap_10_Red	maxima		
Nuryanto & Kochzius 2009	FM244513	Java & Sumatra	Hap_11_orange	maxima		
Nuryanto & Kochzius 2009	FM244514	Java & Sumatra	Hap_12_orange	maxima		
Nuryanto & Kochzius 2009	FM244515	Java & Sumatra	Hap_13_orange	maxima		
Nuryanto & Kochzius 2009	FM244516	Java & Sumatra	Hap_14_orange	maxima		
Nuryanto & Kochzius 2009	FM244517	Java & Sumatra	Hap_15_orange	maxima		
Nuryanto & Kochzius 2009	FM244518	Java & Sumatra	Hap_16_orange	maxima		
Nuryanto & Kochzius 2009	FM244519	Java & Sumatra	Hap_17_orange	maxima		
Nuryanto & Kochzius 2009	FM244520	Java & Sumatra	Hap_18_orange	maxima		
Nuryanto & Kochzius 2009	FM244521	Java & Sumatra	Hap_19_orange	maxima		
Nuryanto & Kochzius 2009	FM244522	Java & Sumatra	Hap_20_orange	maxima		
Nuryanto & Kochzius 2009	FM244523	Java & Sumatra	Hap_21_orange	maxima		
Nuryanto & Kochzius 2009	FM244524	Java & Sumatra	Hap_22_orange	maxima		
Nuryanto & Kochzius 2009	FM244525	Java & Sumatra	Hap_23_orange	maxima		
Nuryanto & Kochzius 2009	FM244526	Java & Sumatra	Hap_24_orange	maxima		
Nuryanto & Kochzius 2009	FM244527	Java & Sumatra	Hap_25_orange	maxima		
Nuryanto & Kochzius 2009	FM244528	Java & Sumatra	Hap_26_orange	maxima		
Nuryanto & Kochzius 2009	FM244529	Java & Sumatra	Hap_27_orange	maxima		
Nuryanto & Kochzius 2009	FM244530	Java & Sumatra	Hap_28	maxima		
Nuryanto & Kochzius 2009	FM244531	Java & Sumatra	Hap_29_orange	maxima		
Nuryanto & Kochzius 2009	FM244532	Java & Sumatra	Hap_30_orange	maxima		
Nuryanto & Kochzius 2009	FM244533	Java & Sumatra	Hap_31_orange	maxima		
Nuryanto & Kochzius 2009	FM244534	Java & Sumatra	Hap_32_orange	maxima		
Nuryanto & Kochzius 2009	FM244535	Central Indo	Hap_33	maxima		
Nuryanto & Kochzius 2009	FM244536	Central Indo	Hap_34	maxima		
Nuryanto & Kochzius 2009	FM244537	Central Indo	Hap_35	maxima		
Nuryanto & Kochzius 2009	FM244538	Central Indo	Hap_36	maxima		
Nuryanto & Kochzius 2009	FM244539	Central Indo	Hap_37	maxima		
Nuryanto & Kochzius 2009	FM244540	Central Indo	Hap_38	maxima		
Nuryanto & Kochzius 2009	FM244541	Central Indo	Hap_39	maxima		
Nuryanto & Kochzius 2009	FM244542	Central Indo	Hap_40	maxima		
Nuryanto & Kochzius 2009	FM244543	Central Indo	Hap_41	maxima		
Nuryanto & Kochzius 2009	FM244544	Central Indo	Hap_42	maxima		
Nuryanto & Kochzius 2009	FM244545	Central Indo	Hap_43	maxima		
Nuryanto & Kochzius 2009	FM244546	Central Indo	Hap_44	maxima		
Nuryanto & Kochzius 2009	FM244547	Central Indo	Hap_45	maxima		
Nuryanto & Kochzius 2009	FM244548	Central Indo	Hap_46	maxima		
Nuryanto & Kochzius 2009	FM244549	Central Indo	Hap_47	maxima		
Nuryanto & Kochzius 2009	FM244550	Central Indo	Hap_48	maxima		
Nuryanto & Kochzius 2009	FM244551	Central Indo	Hap_49	maxima		
Nuryanto & Kochzius 2009	FM244552	Central Indo	Hap_50	maxima		
Nuryanto & Kochzius 2009	FM244553	Central Indo	Hap_51	maxima		
Nuryanto & Kochzius 2009	FM244554	Central Indo	Hap_52	maxima		
Nuryanto & Kochzius 2009	FM244555	Central Indo	Hap_53	maxima		
Nuryanto & Kochzius 2009	FM244556	Central Indo	Hap_54	maxima		
Nuryanto & Kochzius 2009	FM244557	Central Indo	Hap_55	maxima		
Nuryanto & Kochzius 2009	FM244558	Central Indo	Hap_56	maxima		
Nuryanto & Kochzius 2009	FM244559	Central Indo	Hap_57	maxima		
Nuryanto & Kochzius 2009	FM244560	Central Indo	Hap_58	maxima		
Nuryanto & Kochzius 2009	FM244561	Central Indo	Hap_59	maxima		
Nuryanto & Kochzius 2009	FM244562	Central Indo	Hap_60	maxima		
Nuryanto & Kochzius 2009	FM244563	Central Indo	Hap_61	maxima		
Nuryanto & Kochzius 2009	FM244564	Central Indo	Hap_62	maxima		
Nuryanto & Kochzius 2009	FM244565	Central Indo	Hap_63	maxima		
Nuryanto & Kochzius 2009	FM244566	Central Indo	Hap_64	maxima		
Nuryanto & Kochzius 2009	FM244567	Central Indo	Hap_65	maxima		
Nuryanto & Kochzius 2009	FM244568	Central Indo	Hap_66	maxima		
Nuryanto & Kochzius 2009	FM244569	Central Indo	Hap_67	maxima		
Nuryanto & Kochzius 2009	FM244570	Central Indo	Hap_68	maxima		
Nuryanto & Kochzius 2009	FM244571	Central Indo	Hap_69	maxima		
Nuryanto & Kochzius 2009	FM244572	Central Indo	Hap_70	maxima		

Nuryanto & Kochzius 2009	FM244573	Central Indo	Hap_71	maxima		
Nuryanto & Kochzius 2009	FM244574	Central Indo	Hap_72	maxima		
Nuryanto & Kochzius 2009	FM244575	Central Indo	Hap_73	maxima		
Nuryanto & Kochzius 2009	FM244576	Central Indo	Hap_74	maxima		
Nuryanto & Kochzius 2009	FM244577	Central Indo	Hap_75	maxima		
Nuryanto & Kochzius 2009	FM244578	Central Indo	Hap_76	maxima		
Nuryanto & Kochzius 2009	FM244579	Central Indo	Hap_77	maxima		
Nuryanto & Kochzius 2009	FM244580	Central Indo	Hap_78	maxima		
Nuryanto & Kochzius 2009	FM244581	Central Indo	Hap_79	maxima		
Nuryanto & Kochzius 2009	FM244582	Central Indo	Hap_80	maxima		
Nuryanto & Kochzius 2009	FM244583	Central Indo	Hap_81	maxima		
Nuryanto & Kochzius 2009	FM244584	Central Indo	Hap_82	maxima		
Nuryanto & Kochzius 2009	FM244585	Central Indo	Hap_83	maxima		
Nuryanto & Kochzius 2009	FM244586	Central Indo	Hap_84	maxima		
Nuryanto & Kochzius 2009	FM244587	Central Indo	Hap_85	maxima		
Nuryanto & Kochzius 2009	FM244588	Central Indo	Hap_86	maxima		
Nuryanto & Kochzius 2009	FM244589	Central Indo	Hap_87	maxima		
Nuryanto & Kochzius 2009	FM244590	Central Indo	Hap_88	maxima		
Nuryanto & Kochzius 2009	FM244591	Central Indo	Hap_89	maxima		
Nuryanto & Kochzius 2009	FM244592	Central Indo	Hap_90	maxima		
Nuryanto & Kochzius 2009	FM244593	Central Indo	Hap_91	maxima		
Nuryanto & Kochzius 2009	FM244594	Central Indo	Hap_92	maxima		
Nuryanto & Kochzius 2009	FM244595	Central Indo	Hap_93	maxima		
Nuryanto & Kochzius 2009	FM244596	Central Indo	Hap_94	maxima		
Nuryanto & Kochzius 2009	FM244597	Central Indo	Hap_95	maxima		
Nuryanto & Kochzius 2009	FM244598	Central Indo	Hap_96	maxima		
Nuryanto & Kochzius 2009	FM244599	Central Indo	Hap_97	maxima		
Nuryanto & Kochzius 2009	FM244600	Central Indo	Hap_98	maxima		
Nuryanto & Kochzius 2009	FM244601	Central Indo	Hap_99	maxima		
Nuryanto & Kochzius 2009	FM244602	Central Indo	Hap_100	maxima		
Nuryanto & Kochzius 2009	FM244603	Central Indo	Hap_101	maxima		
Nuryanto & Kochzius 2009	FM244604	Central Indo	Hap_102	maxima		
Nuryanto & Kochzius 2009	FM244605	Central Indo	Hap_103	maxima		
Nuryanto & Kochzius 2009	FM244606	Central Indo	Hap_104	maxima		
Nuryanto & Kochzius 2009	FM244607	Indonesia-Biak	Hap_105_yell	maxima		
Nuryanto & Kochzius 2009	FM244608	Indonesia-Biak	Hap_106_yell	maxima		
Nuryanto & Kochzius 2009	FM244609	Indonesia-Biak	Hap_107_yell	maxima		
Nuryanto & Kochzius 2009	FM244610	Indonesia-Biak	Hap_108_yell	maxima		
Nuryanto & Kochzius 2009	FM244611	Indonesia-Biak	Hap_109_yell	maxima		
Nuryanto & Kochzius 2009	FM244612	Indonesia-Biak	Hap_110_yell	maxima		
Nuryanto & Kochzius 2009	FM244613	Indonesia-Biak	Hap_111_yell	maxima		
Nuryanto & Kochzius 2009	FM244614	Indonesia-Biak	Hap_112_yell	maxima		
Nuryanto & Kochzius 2009	FM244615	Indonesia-Biak	Hap_113_yell	maxima		
Nuryanto & Kochzius 2009	FM244616	Indonesia	Hap_114	maxima		
Nuryanto & Kochzius 2009	FM244617	Indonesia-Biak	Hap_115_yell	maxima		
Nuryanto & Kochzius 2009	FM244618	Indonesia-Biak	Hap_116_yell	maxima		
Nuryanto & Kochzius 2009	FM244619	Indonesia	Hap_117	maxima		
Tang & Chen, unpublished	DQ168140	Taiwan	Tridacna sp.	sp.		
Nuryanto et al. 2007	EU003615	Red Sea	RedSea01	squamosa		
DeBoer et al. unpublished	EU346361	Indonesia	TSD50	squamosa		
DeBoer et al. unpublished	EU346362	Indonesia	TSD51	squamosa		
DeBoer et al. unpublished	EU346363	Indonesia	TSD52	squamosa		
Neo & Todd 2012	JN392020	Singapore	NML_53	squamosa		
Neo & Todd 2012	JN392021	Singapore	NML_54	squamosa		
Neo & Todd 2012	JN392022	Singapore	NML_44	squamosa		
Neo & Todd 2012	JN392023	Singapore	NML_55	squamosa		
Neo & Todd 2012	JN392024	Singapore	NML_50	squamosa		
Neo & Todd 2012	JN392025	Singapore	NML_49	squamosa		
Neo & Todd 2012	JN392026	Singapore	NML_48	squamosa		
Neo & Todd 2012	JN392027	Singapore	NML_51	squamosa		
Neo & Todd 2012	JN392028	Singapore	NML_52	squamosa		
Neo & Todd 2012	JN392029	Singapore	NML_46	squamosa		
Neo & Todd 2012	JN392030	Singapore	NML_25	squamosa		
Neo & Todd 2012	JN392031	Singapore	NML_45	squamosa		
Neo & Todd 2012	JN392032	Singapore	NML_28	squamosa		
Neo & Todd 2012	JN392033	Singapore	NML_47	squamosa		
Neo & Todd 2012	JN392034	Singapore	NML_21	squamosa		
Neo & Todd 2012	JN392035	Singapore	NML_26	squamosa		
Neo & Todd 2012	JN392036	Singapore	NML_22	squamosa		
Neo & Todd 2012	JN392037	Singapore	NML_24	squamosa		

Neo & Todd 2012	JN392038	Singapore	NML_23	squamosa		
Neo & Todd 2012	JN392039	Singapore	NML_27	squamosa		
Ladhar-Chaabouni et al., 2010.	FJ179505			Cerastoderma glaucum		
Kirkendale & Middelfart, unpublis	FJ745302			Fragum sueziense		
Kirkendale & Middelfart, unpublis	FJ745346			Corculum cardiss		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET781_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET782_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET783_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET790_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET916_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET918_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET919_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET920_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET921_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET927_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET929_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET930_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET933_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET935_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Ningaloo, Western Au	ET937_NIN_spp	sp.		
Huelsken et al. (present study)	pending	Solomon Islands	ET2272_SOL_spp	sp.		
Huelsken et al. (present study)	pending	Solomon Islands	ET2339_SOL_spp	sp.		
Huelsken et al. (present study)	pending	Solomon Islands	ET2422_SOL_spp	sp.		
Huelsken et al. (present study)	pending	Solomon Islands	ET2425_SOL_spp	sp.		
Huelsken et al. (present study)	pending	Solomon Islands	ET2452_SOL_spp	sp.		
Huelsken et al. (present study)	pending	Solomon Islands	ET2454_SOL_spp	sp.		
Huelsken et al. (present study)	pending	Torres Strait, Austral	ET1961_TOR_squa	squamosa		
Huelsken et al. (present study)	pending	Lihou Reef, Coral Sea	ET358_LIH_squa	squamosa		
Huelsken et al. (present study)	pending	Torres Strait, Austral	ET1995_TOR_max	maxima		
Huelsken et al. (present study)	pending	Torres Strait, Austral	ET2006_TOR_max	maxima		
Huelsken et al. (present study)	pending	Torres Strait, Austral	ET2011_TOR_max	maxima		
Huelsken et al. (present study)	pending	Torres Strait, Austral	ET2015_TOR_max	maxima		
Huelsken et al. (present study)	pending	Solomon Islands	ET2145_SOL_max	maxima		
Huelsken et al. (present study)	pending	Solomon Islands	ET2348_SOL_max	maxima		
Huelsken et al. (present study)	pending	Solomon Islands	ET2386_SOL_max	maxima		
Huelsken et al. (present study)	pending	Solomon Islands	ET2395_SOL_max	maxima		
Huelsken et al. (present study)	pending	Heron Island, Austra	ET26_HER_max	maxima		
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET2793_LIZ_max	maxima		
Huelsken et al. (present study)	pending	Lihou Reef, Coral Sea	ET303_LIH_max	maxima		
Huelsken et al. (present study)	pending	Lihou Reef, Coral Sea	ET305_LIH_max	maxima		
Huelsken et al. (present study)	pending	Lihou Reef, Coral Sea	ET357_LIH_max	maxima		
Huelsken et al. (present study)	pending	Lihou Reef, Coral Sea	ET375_LIH_max	maxima		
Huelsken et al. (present study)	pending	Heron Island, Austra	ET43_HER_max	maxima		
Huelsken et al. (present study)	pending	Heron Island, Austra	ET49_HER_max	maxima		
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET650_LIZ_max	maxima		
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET654_LIZ_max	maxima		
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET655_LIZ_max	maxima		
Huelsken et al. (present study)	pending	Torres Strait, Austral	ET1879_TOR_croc	crocea		
Huelsken et al. (present study)	pending	Torres Strait, Austral	ET1880_TOR_croc	crocea		
Huelsken et al. (present study)	pending	Torres Strait, Austral	ET1884_TOR_croc	crocea		
Huelsken et al. (present study)	pending	Torres Strait, Austral	ET1885_TOR_croc	crocea		
Huelsken et al. (present study)	pending	Solomon Islands	ET2180_SOL_croc	crocea		
Huelsken et al. (present study)	pending	Solomon Islands	ET2198_SOL_croc	crocea		
Huelsken et al. (present study)	pending	Solomon Islands	ET2333_SOL_croc	crocea		
Huelsken et al. (present study)	pending	Solomon Islands	ET2334_SOL_croc	crocea		
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET527_LIZ_croc	crocea		
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET529_LIZ_croc	crocea		
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET542_LIZ_croc	crocea		
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET544_LIZ_croc	crocea		
Huelsken et al. (present study)	pending	Lihou Reef, Coral Sea	ET356_LIH_dera	derasa		

16S				
Study Source	**Acc No**	**Locality**	**Name used in study**	**Putative species**
Study Source	Acc No	Locality	Study name	Putative species
DeBoer et al. 2008	EU341331	Indonesia	TSD_1	crocea
DeBoer et al. 2008	EU341332	Indonesia	TSD_2	crocea
DeBoer et al. 2008	EU341333	Indonesia	TSD_3	crocea
DeBoer et al. 2008	EU341334	Indonesia	TSD_4	maxima
DeBoer et al. 2008	EU341335	Indonesia	TSD_5	maxima
DeBoer et al. 2008	EU341336	Indonesia	TSD_6	crocea
DeBoer et al. 2008	EU341337	Indonesia	TSD_7	crocea
DeBoer et al. 2008	EU341338	Indonesia	TSD_8	squamosa
DeBoer et al. 2008	EU341339	Indonesia	TSD_9	crocea
DeBoer et al. 2008	EU341340	Indonesia	TSD_10	crocea
DeBoer et al. 2008	EU341341	Indonesia	TSD_11	crocea
DeBoer et al. 2008	EU341342	Indonesia	TSD_12	crocea
DeBoer et al. 2008	EU341343	Indonesia	TSD_13	maxima
DeBoer et al. 2008	EU341344	Indonesia	TSD_14	crocea
DeBoer et al. 2008	EU341345	Indonesia	TSD_15	squamosa
DeBoer et al. 2008	EU341346	Indonesia	TSD_16	crocea
DeBoer et al. 2008	EU341347	Indonesia	TSD_17	crocea
DeBoer et al. 2008	EU341348	Indonesia	TSD_18	crocea
DeBoer et al. 2008	EU341349	Indonesia	TSD_19	crocea
Schneider & O'Foighil 1999	AF122975	aquarium store		gigas
Schneider & O'Foighil 1999	AF122976	aquarium store		derasa
Schneider & O'Foighil 1999	AF122977	aquarium store		tevoroa
Schneider & O'Foighil 1999	AF122978	aquarium store		squamosa
Schneider & O'Foighil 1999	AF122979	aquarium store		maxima
Schneider & O'Foighil 1999	AF122980	aquarium store		crocea
Richter et al. 2008	AM909726	Jordan (RedSea)	Star1	costata
Richter et al. 2008	AM909727	Jordan (RedSea)	Star2	costata
Richter et al. 2008	AM909728	Jordan (RedSea)	Star3	costata
Richter et al. 2008	AM909729	Jordan (RedSea)	Star4	costata
Richter et al. 2008	AM909730	Jordan (RedSea)	Star5	costata
Richter et al. 2008	AM909731	Jordan (RedSea)	Star6	costata
Richter et al. 2008	AM909732	Jordan (RedSea)	Star10	costata
Richter et al. 2008	AM909733	Egypt (RedSea)	Tnov015	costata
Richter et al. 2008	AM909734	Egypt (RedSea)	Tnov01	costata
Richter et al. 2008	AM909735	Egypt (RedSea)	Tnov026	costata
Richter et al. 2008	AM909736	Egypt (RedSea)	Tnov02	costata
Richter et al. 2008	AM909737	Egypt (RedSea)	Tnov037	costata
Richter et al. 2008	AM909738	Egypt (RedSea)	Tnov03	costata
Richter et al. 2008	AM909739	Egypt (RedSea)	Tnov048	costata
Richter et al. 2008	AM909740	Egypt (RedSea)	Tnov04	costata
Richter et al. 2008	AM909741	Egypt (RedSea)	Tnov059	costata
Richter et al. 2008	AM909742	Jordan (RedSea)	Tmax03	maxima
Richter et al. 2008	AM909743	Jordan (RedSea)	Tmax04	maxima
Richter et al. 2008	AM909744	Jordan (RedSea)	Tmax06	maxima
Richter et al. 2008	AM909745	Jordan (RedSea)	Tmax01CR	maxima
Richter et al. 2008	AM909746	Jordan (RedSea)	Tmax02CR	maxima
Richter et al. 2008	AM909747	Jordan (RedSea)	Tmax03CR	maxima
Richter et al. 2008	AM909748	Egypt (RedSea)	Tmax01Hu	maxima
Richter et al. 2008	AM909749	Indonesia	TmaInd12	maxima
Richter et al. 2008	AM909750	Indonesia	TmaInd13	maxima
Richter et al. 2008	AM909751	Egypt (RedSea)	Tmax02Eg	maxima
Richter et al. 2008	AM909752	Egypt (RedSea)	Tmax3Eg	maxima
Richter et al. 2008	AM909753	Jordan (RedSea)	Tsqua17	squamosa
Richter et al. 2008	AM909754	Jordan (RedSea)	Tsqua27	squamosa
Richter et al. 2008	AM909755	Jordan (RedSea)	Tsqua42	squamosa

Richter et al. 2008	AM909756	Jordan (RedSea)	Tsqua49	squamosa
Richter et al. 2008	AM909757	Jordan (RedSea)	Tsqua56	squamosa
Richter et al. 2008	AM909758	Jordan (RedSea)	Tsqua65	squamosa
Richter et al. 2008	AM909759	Jordan (RedSea)	Tsqua67	squamosa
Richter et al. 2008	AM909760	Jordan (RedSea)	Tsqua81	squamosa
Richter et al. 2008	AM909761	Jordan (RedSea)	Tsqua04	squamosa
Richter et al. 2008	AM909762	Jordan (RedSea)	Tsq61_3	squamosa
Richter et al. 2008	AM909763	Indonesia	Tcr232_6	crocea
Richter et al. 2008	AM909764	Indonesia	Tcr232_7	crocea
Tang & Chen unpub	DQ115320	Taiwan		maxima
Tang & Chen unpub	DQ119339	Taiwan	YCT-2005	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET784	squamosa
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET786	squamosa
Huelsken et al. (present study)	pending	Torres Strait, Austra	ET1961	squamosa
Huelsken et al. (present study)	pending	Lihou Reef, Coral Se	ET358	squamosa
Huelsken et al. (present study)	pending	Lihou Reef, Coral Se	ET309	squamosa
Huelsken et al. (present study)	pending	Torres Strait, Austra	ET1995	maxima
Huelsken et al. (present study)	pending	Torres Strait, Austra	ET2006	maxima
Huelsken et al. (present study)	pending	Torres Strait, Austra	ET2011	maxima
Huelsken et al. (present study)	pending	Torres Strait, Austra	ET2015	maxima
Huelsken et al. (present study)	pending	Solomon Islands	ET2145	maxima
Huelsken et al. (present study)	pending	Solomon Islands	ET2348	maxima
Huelsken et al. (present study)	pending	Solomon Islands	ET2386	maxima
Huelsken et al. (present study)	pending	Solomon Islands	ET2395	maxima
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET654	maxima
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET655	maxima
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET2793	maxima
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET2789	maxima
Huelsken et al. (present study)	pending	Lihou Reef, Coral Se	ET303	maxima
Huelsken et al. (present study)	pending	Lihou Reef, Coral Se	ET305	maxima
Huelsken et al. (present study)	pending	Lihou Reef, Coral Se	ET357	maxima
Huelsken et al. (present study)	pending	Heron Island, Austra	ET43	maxima
Huelsken et al. (present study)	pending	Heron Island, Austra	ET49	maxima
Huelsken et al. (present study)	pending	Heron Island, Austra	ET21	maxima
Huelsken et al. (present study)	pending	Torres Strait, Austra	ET1879	crocea
Huelsken et al. (present study)	pending	Torres Strait, Austra	ET1880	crocea
Huelsken et al. (present study)	pending	Torres Strait, Austra	ET1884	crocea
Huelsken et al. (present study)	pending	Torres Strait, Austra	ET1885	crocea
Huelsken et al. (present study)	pending	Solomon Islands	ET2180	crocea
Huelsken et al. (present study)	pending	Solomon Islands	ET2198	crocea
Huelsken et al. (present study)	pending	Solomon Islands	ET2333	crocea
Huelsken et al. (present study)	pending	Solomon Islands	ET2334	crocea
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET527	crocea
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET529	crocea
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET542	crocea
Huelsken et al. (present study)	pending	Lizard Island, Austra	ET544	crocea
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET781	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET782	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET783	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET790	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET916	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET918	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET920	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET921	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET927	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET935	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET937	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET917	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET923	sp

Huelsken et al. (present study)	pending	Ningaloo, Western A	ET924	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET925	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET926	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET928	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET931	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET932	sp
Huelsken et al. (present study)	pending	Ningaloo, Western A	ET933	sp
Huelsken et al. (present study)	pending	Solomon Islands	ET2175	sp
Huelsken et al. (present study)	pending	Solomon Islands	ET2450	sp
Huelsken et al. (present study)	pending	Solomon Islands	ET2339	sp
Huelsken et al. (present study)	pending	Solomon Islands	ET2422	sp
Huelsken et al. (present study)	pending	Solomon Islands	ET2425	sp
Huelsken et al. (present study)	pending	Solomon Islands	ET2454	sp
Huelsken et al. (present study)	pending	Lihou Reef, Coral Se	ET306	derasa
Huelsken et al. (present study)	pending	Lihou Reef, Coral Se	ET307	derasa
Huelsken et al. (present study)	pending	Lihou Reef, Coral Se	ET308	derasa
Huelsken et al. (present study)	pending	Lihou Reef, Coral Se	ET356	derasa

ITS				
Study Source		Acc No	Name used in study	Putative species
Huelsken et al. (present study)		pending	ET309	squamosa
Huelsken et al. (present study)		pending	ET358	squamosa
Huelsken et al. (present study)		pending	ET1961	squamosa
Huelsken et al. (present study)		pending	ET556	crocea
Huelsken et al. (present study)		pending	ET543	crocea
Huelsken et al. (present study)		pending	ET765	sp.
Huelsken et al. (present study)		pending	ET918	sp.
Huelsken et al. (present study)		pending	ET766	sp.
Huelsken et al. (present study)		pending	ET781	sp.
Huelsken et al. (present study)		pending	ET783	sp.
Huelsken et al. (present study)		pending	ET1896	maxima
Huelsken et al. (present study)		pending	ET1921	maxima
Huelsken et al. (present study)		pending	ET1923	maxima
Huelsken et al. (present study)		pending	ET1993	maxima
Huelsken et al. (present study)		pending	ET1995	maxima
Huelsken et al. (present study)		pending	ET2011	maxima
Huelsken et al. (present study)		pending	ET2016	maxima
Huelsken et al. (present study)		pending	ET21	maxima
Huelsken et al. (present study)		pending	ET22	maxima
Huelsken et al. (present study)		pending	ET657	maxima
Huelsken et al. (present study)		pending	ET658	maxima
Huelsken et al. (present study)		pending	ET659	maxima
Huelsken et al. (present study)		pending	ET660	maxima
Huelsken et al. (present study)		pending	ET661	maxima
Huelsken et al. (present study)		pending	ET2272P	maxima
Huelsken et al. (present study)		pending	ET2386	maxima
Huelsken et al. (present study)		pending	ET2424	maxima
Huelsken et al. (present study)		pending	ET2793	maxima
Huelsken et al. (present study)		pending	ET300	maxima
Huelsken et al. (present study)		pending	ET301	maxima
Huelsken et al. (present study)		pending	ET303	maxima
Huelsken et al. (present study)		pending	ET310	maxima
Huelsken et al. (present study)		pending	ET374	maxima
Huelsken et al. (present study)		pending	ET375	maxima
Huelsken et al. (present study)		pending	ET377	maxima
Huelsken et al. (present study)		pending	ET648	maxima
Huelsken et al. (present study)		pending	ET650	maxima
Huelsken et al. (present study)		pending	ET652	maxima
Huelsken et al. (present study)		pending	ET653	maxima
Huelsken et al. (present study)		pending	ET47	maxima
Huelsken et al. (present study)		pending	ET44	maxima
Huelsken et al. (present study)		pending	ET46	maxima
Huelsken et al. (present study)		pending	ET1946	maxima
Huelsken et al. (present study)		pending	ET2014	maxima
Huelsken et al. (present study)		pending	ET2017	maxima
Huelsken et al. (present study)		pending	ET2018	maxima
Huelsken et al. (present study)		pending	ET306	derasa
Huelsken et al. (present study)		pending	ET656	derasa
Huelsken et al. (present study)		pending	ET308	derasa
Huelsken et al. (present study)		pending	ET356	derasa

Huelsken et al. (present study)	pending	ET307	derasa

MOLECULAR ECOLOGY

Molecular Ecology (2015)

NEWS AND VIEWS

OPINION

Not the time or the place: the missing spatio-temporal link in publicly available genetic data

LISA C. POPE,* LIBBY LIGGINS,† ‡ JUDE KEYSE,* SILVIA B CARVALHO§ and CYNTHIA RIGINOS*

*School of Biological Sciences, The University of Queensland, Brisbane, Qld 4072, Australia, †Allan Wilson Centre for Molecular Ecology and Evolution, New Zealand Institute for Advanced Study, Institute of Natural and Mathematical Sciences, Massey University, Auckland 0745, New Zealand, ‡Auckland War Memorial Museum, Tāmaki Paenga Hira, Auckland 1142, New Zealand, §CIBIO/InBIO – Centro de Investigação em Biodiversidade e Recursos Genéticos da Universidade do Porto, R. Padre Armando Quintas, 4485-661 Vairão, Portugal

Abstract

Genetic data are being generated at unprecedented rates. Policies of many journals, institutions and funding bodies aim to ensure that these data are publicly archived so that published results are reproducible. Additionally, publicly archived data can be 'repurposed' to address new questions in the future. In 2011, along with other leading journals in ecology and evolution, *Molecular Ecology* implemented mandatory public data archiving (the Joint Data Archiving Policy). To evaluate the effect of this policy, we assessed the genetic, spatial and temporal data archived for 419 data sets from 289 articles in *Molecular Ecology* from 2009 to 2013. We then determined whether archived data could be used to reproduce analyses as presented in the manuscript. We found that the journal's mandatory archiving policy has had a substantial positive impact, increasing genetic data archiving from 49 (pre-2011) to 98% (2011–present). However, 31% of publicly archived genetic data sets could not be recreated based on information supplied in either the manuscript or public archives, with incomplete data or inconsistent codes linking genetic data and metadata as the primary reasons. While the majority of articles did provide some geographic information, 40% did not provide this information as geographic coordinates. Furthermore, a large proportion of articles did not contain any information regarding date of sampling (40%). Although the inclusion of spatio-temporal data does require an increase in effort, we argue that the enduring value of publicly accessible genetic data to the molecular ecology field is greatly compromised when such metadata are not archived alongside genetic data.

Correspondence: Lisa C. Pope, Fax: +61 7 334 67646; E-mail: l.pope@uq.edu.au

Keywords: biological ontology, data accessibility, metadata, reproducibility, reuse, standards

Received 4 January 2015; revision received 7 May 2015; accepted 22 May 2015

Introduction

Molecular ecology is a rapidly growing field, and genetic data are being generated at an exponential rate (Kodama *et al.* 2012; Parr *et al.* 2012). Reliable archiving and public access to such data are essential to allow the reproducibility of published research to be assessed, a central tenet of science. Furthermore, data archives with public access can support the application of new statistical approaches, syntheses across studies, and allow the 'repurposing' of data, that is enabling researchers to address questions that differ from those for which the data were originally collected (Sidlauskas *et al.* 2010; Stoltzfus *et al.* 2012).

Molecular ecology encompasses a broad range of topics, illustrated by the sections of this journal. Questions aligning with different topics can often be addressed using the same genetic markers (see Fig. S1, Supporting information), providing extensive opportunities for genetic data to be repurposed in this field. Examples of data repurposing include the construction of megaphylogenies (e.g. the open tree of life project – http://blog.opentreeoflife.org/), delineating genetic 'hot spots' (e.g. Vandergast *et al.* 2008; Wood *et al.* 2012), testing the generality of the central-margin hypothesis (Eckert *et al.* 2008) and predicting the spread of invasive species (Gaither *et al.* 2013), to name but a few. The future value of spatio-temporal genetic data to investigating questions such as the impact of climate change, the ongoing biodiversity crisis and disease spread is incalculable.

In 2011, *Molecular Ecology* entered into the Joint Data Archiving Policy (http://datadryad.org/pages/jdap) motivated by low voluntary rates of public data archiving among contributors (Rausher *et al.* 2010); the JDAP calls for published studies to be reproducible and to facilitate data reuse (see Box 1). Several 'best-practice guides' and recommendations for the provisioning of genetic data and metadata have been contributed (e.g. Leebens-Mack *et al.* 2006; Whitlock 2011; White *et al.* 2013; Cranston *et al.* 2014). Multilocus genotypes and/or DNA sequences identifiable to the level of individual (i.e. individual-based genetic data) are preferable for assessing both the reproducibility of

© 2015 John Wiley & Sons Ltd

Box 1. The Joint Data Archiving Policy (JDAP) describes a requirement that data supporting publications be publicly available. This policy was adopted in a joint and coordinated fashion by many leading journals in the field of evolution in 2011 (Rausher *et al.* 2010), and JDAP has since been adopted by additional journals across various disciplines.

Molecular ecology policy on data archiving

Data are important products of the scientific enterprise, and they should be preserved and usable for decades in the future. As such, *Molecular Ecology* requires authors to archive the data supporting their results and conclusions along with sufficient details so that a third party can interpret them correctly. Studies with exemplary data and code archiving are more valuable for future research and, all else being equal, will be given higher priority for publication. Data should be archived in an appropriate public archive, such as GenBank, Gene Expression Omnibus, TreeBASE, Dryad, the Knowledge Network for Biocomplexity, and your own institutional or funder repository, or as Supporting Information on the *Molecular Ecology* website. The utility of archived data is greatly enhanced when the scripts and input files used in the analyses are also made available. Given that scripts may be a mix of proprietary and freely available code, their deposition is not compulsory, but we nonetheless strongly encourage authors to make these scripts available whenever possible. As discussed by Whitlock *et al.* (2010), accurate interpretation of data will likely 'require a short additional text document, with details specifying the meaning of each column in the data set. The preparation of such shareable data sets will be easiest if these files are prepared as part of the data analysis phase of the preparation of the paper, rather than after acceptance of a manuscript'. For additional guidelines on data deposition best practice, please visit http://datadryad.org/depositing. Data must be publicly available at the time of publication. Embargos may be granted in exceptional instances at the discretion of the Managing Editors. Exemptions to this policy may also be granted, especially for sensitive information such as human subject data or the location of endangered species.

published research and reuse (discussed in Whitlock 2011). It is more difficult, however, to specify which metadata may be relevant for the reproducibility of a study and/or future repurposing of genetic data. Nonetheless, spatio-temporal information, such as the location and time of genetic sampling, is of central importance to most ecological and evolutionary studies, and their inclusion is certain to expand the scope for future data reuse or repurpose.

Despite the universal nature of geographic and temporal information, there is often no requirement for these metadata to be associated with genetic data by existing public databases, journals or institutions. For example, since 2005, DNA sequences submitted to the National Centre for Biotechnology Information (NCBI) have been encouraged, but not required, to contain geographic information. Voluntary inclusion of this information appears limited, with fewer than 7% of sequences for 'barcoding genes' submitted to NCBI since mid-2011 containing geographic coordinates (Marques *et al.* 2013). However, some data archives do require such metadata to be deposited and linked to genetic data (e.g. Metagenomics Analysis Server, http://press.igsb.anl.gov/mg-rast/metadata-in-mg-rast/, requires latitude and longitude, but not time; the NCBI Bioproject requires both spatial and temporal information, Dugan *et al.* 2014). How biological databases should cross-communicate and how genetic and biodiversity ontologies can support such information exchange has been an active topic of discussion and implementation. For example, Gene Ontology (GO) standards promote the exchange of information among the GO Consortium, including FlyBase, WormBase, J Craig Venter Institute and Mouse Genome Informatics (http://geneontology.org/page/go-consortium-contributors-list). Similarly, metadata standards such as the Darwin Core underpin biodiversity databases such as the Global Biodiversity Information Facility (GBIF) and the Ocean Biogeographic Information System (OBIS). The need for more inclusive standards that encompass all aspects of biodiversity, including genetic biodiversity, is recognized, and such standards are under development (Yilmaz *et al.* 2011; Walls *et al.* 2014).

At present, there are no universal genetic or biodiversity databases to suit the variety of studies published in *Molecular Ecology*, and NCBI and DataDryad are currently the most used archive facilities for this journal (based on studies reviewed herein). NCBI has provided an immense resource to molecular ecologists via their restricted entry format and active data quality-checking facilities, enabling uniform and programmatic data retrieval. On the other hand, DataDryad provides a platform that allows the user to provide any data in any format. While this flexibility is extremely useful, it becomes difficult to assess exactly which data have been archived without an informative, accompanying text or key. Furthermore, although geographic and temporal information is sometimes contained within a publication, differences between reference codes used in metadata and genetic data files can render this information useless. While it is possible to contact the original authors to obtain data, this approach was found to have low success in other areas of genetic research (Magee *et al.* 2014).

Increasingly, the importance of public data archiving is being recognized by funding bodies (e.g. National Institutes of Health 2003, National Science Foundation, Natural

© 2015 John Wiley & Sons Ltd

Sciences and Engineering Research Council of Canada, National Environment Research Council of the UK, The Austrian Science Fund, the Deutsche Forschungsgemeinschaft in Germany, and Australian Research Council 2013), universities and journals (Moore *et al.* 2010; Fairbairn 2011; Baker 2013; Lin & Strasser 2014). A dramatic increase in public archiving of genetic data has resulted from these institutional policies (Vines *et al.* 2013). However, the importance of archiving associated metadata is less recognized. Consequently, while great progress has been made towards the public availability of genetic data, the lack of emphasis on provision of associated information, such as geographic location and time of sampling, may impede our ability to fully reproduce such studies or use their genetic data in new ways (Anonymous, 2008).

Given that *Molecular Ecology* was one of the first journals in evolution and ecology to adopt a mandatory public data access policy, here we gauge the impact of this policy on public archiving of genetic data. Then, given the importance of geographic and temporal information to many ecological and evolutionary studies, we evaluate the extent to which spatio-temporal data associated with genetic data are being made publicly available by the molecular ecology community. To do this, we examined manuscripts from 20 issues of *Molecular Ecology* from 2009 to 2013. For these articles, we determined the following: Have genetic data been made publicly available? Could the analyses presented be reproduced based on the total information made publicly available? Has geographic or temporal information been provided and at what scale? What is the scope for repurposing the associated data for future studies?

Methods

We examined all original articles containing newly generated genetic data from 20 issues of *Molecular Ecology* in the 2009 to 2013 period (two issues from July, and two from December each year), a total of 289 articles. Many articles utilized multiple markers, which were often archived differently. From 289 articles, we obtained 419 genetic data sets (one data set for each different genetic marker used in each manuscript) for which we assessed public archiving rates as well as the ability to 'reproduce' analyses based on the provisioning of genetic data and spatial and temporal sampling information.

We defined articles as having 'publicly available' genetic data if *any* genetic data were lodged in a public repository (e.g. NCBI, DataDryad) or provided as supplementary material on the journal website; thus, our criterion for publically available data was very lenient. We searched the text of each article for reference to public data archives and, for articles published after 2011, utilized the 'Data Accessibility' section. We recorded the following: genetic marker type, type of genetic data archived (e.g. individual, population level), and the location of the genetic data if it had been archived.

To address the 'reproducibility' of an article, we assessed whether the genetic data could be recreated and whether sufficient metadata had been provided such that *all* the analyses presented in that article could be reproduced. As different public archiving practices are often used for different genetic marker types (e.g. sequences versus microsatellites), and rates of public archiving have varied for different genetic marker types over time, genetic data were considered at the level of data set. For articles that included links or references to publically available genetic data files, we started by examining these files and applied a simple set of criteria to gauge whether it would be possible to *recreate* the original genetic data set(s): (i) reference codes used to identify individuals in the publically archived genetic data had to match those used in the manuscript or a linking file, and (ii) individual genetic data could be determined. For multilocus nuclear markers, if only summary allele frequencies were provided, the data set was not deemed recreatable as full genotypes with linkage relationships could not be inferred. For DNA sequence-based studies where only unique sequences were archived, data sets were only considered recreatable if haplotype frequencies and sample sizes were provided, allowing individual genetic data to be recreated. Haplotype information provided non-numerically, such as pie graphs, was not considered sufficient for reconstructing genotypes. For phylogenetic studies, a single sequence per species was sufficient to meet the recreatable genetic data set criterion. These criteria were selected based on common problems we had personally experienced when attempting to reanalyse data sets. We were conservative when designating a data set as not 'recreatable', and if there was any doubt, we assumed that the archived data could be used to recreate a data set.

The ability to recreate the relevant genetic data set(s) (using data set criteria described previously) was deemed essential for an article to be considered 'reproducible'. In some cases, article 'reproducibility' only required that the genetic data set(s) were recreatable, for instance when spatial and temporal information was irrelevant to the study objectives. In other cases, spatial and/or temporal information of an appropriate scale (i.e. metadata) was also required for the analyses to be reproducible. If these metadata were not provided at a sufficiently accurate scale to allow the presented analyses to be performed, the article was not classed as reproducible. Because we did not recreate the actual presented analyses, our assessment of complete article-level reproducibility is certain to be upwardly biased.

We examined the number of articles for which spatio-temporal metadata were provided, and assessed the precision of these data. We excluded a small number of articles for which it could be argued that geography was not relevant (e.g. laboratory/methodological/within-population studies); thus, 252 articles were examined for spatio-temporal metadata. As with genetic data, all publicly available materials, including the text of the article, supplementary text and publically archived data, were searched to determine whether geographic or temporal information was provided. When geographic information was present, we categorized its level of precision:

© 2015 John Wiley & Sons Ltd

1. Where geographic information was provided as text only, we defined this as two categories: 'locality' and 'region'. Text was classified as 'region' if the area specified was 'large', for example ocean, country, state, region, or province; otherwise, it was classed as 'locality' (town, district, etc.).
2. Where coordinates were provided (latitude, longitude or UTM), we defined their precision using three categories: ~ > 100 km (degrees only); ~ 1 – 100 km (degrees up to two decimal places, or minutes); and ~ < 1 km (remainder).
3. Where geographic information was provided using an undefined coordinate system, if only a map was provided with no text, this was categorized as 'other'.

Where a record was provided of the time of sampling, we categorized the precision as: year range, year, and more accurate than a year (< year). In the majority of articles, the same spatio-temporal information applied to all data sets within the article. In the small number of cases where spatio-temporal information differed between data sets (9 of 252), the more precise spatio-temporal data were used.

Finally, to examine the potential to 'repurpose' archived data, we combined information on publicly archived genetic data sets that provided linking codes and individual genetic data ('recreatable' genetic data sets), with information on spatio-temporal metadata, for articles published post-JDAP. We plotted the number of recreatable data sets for which geographic and temporal information was provided at various levels of precision, providing an indication of the extent to which genetic data sets and accompanying spatio-temporal data are available for 'repurposing'.

Results

Many articles contained multiple data sets and used more than one public database to store their genetic data. Nearly half of the articles examined stored some data in NCBI (147/289; 47%). The majority of DNA sequence data sets were stored in NCBI (133/156; 85%), and 45% of these included some kind of geographic information in NCBI itself (60 of 156 sequence data sets), although this was generally 'country'. The second most commonly used public database was DataDryad (112/289; 39%).

Public archiving of data increased greatly over the 5-year period examined, starting at 49% and ending at 98% (based on 289 articles, Fig. 1a). This gain was primarily due to increased public archiving of non-DNA sequence data such as microsatellite and SNP genotypes (Fig. S2, Supporting information). While public archiving rates improved over time, increasing the overall proportion of genetic data sets that could be 'recreated', other issues remained steady: in addition to data sets with no public archiving (72/419; 12%), 19% failed to provide individual level genetic data (79/419) and 10% did not provide a link between codes used in the manuscript and those used in the archived data (43/419), with some studies failing to provide both codes and individual genotypes (Fig. 1b). In summary, 31% of genetic data sets that have nominally been publically archived could not be recreated (106/347). Articles evaluated as presenting completely reproducible analyses also increased over time, although again the proportion was not as great as might be expected given public data archiving rates (18–72%; Fig. 1a). Restricting consideration to only those articles that included public data archiving (242 articles), only 41% (100) presented fully reproducible analyses.

In contrast to the gains in genetic data archiving, the provisioning of geographic and temporal data changed little from 2009 to 2013 (Fig. 1c, d). All articles for which geography was deemed relevant provided geographic information of some kind. However, over a third of articles provided geographic information as text only (90/252, 36%), with 18% describing geography in the text at a regional-level only (ocean, country, state, region or province; 45/252). Only 60% of articles provided geographic coordinates (151/252). There has been an increase in the level of precision of geographic coordinates when provided (<1 km increased from 29 to 46%); however, the overall rate of latitude and longitude reporting has remained steady (Fig. 1c). Similarly, reporting of time of sampling remained fairly constant (Fig. 1d). Around 40% of articles did not provide any temporal information (100/252), and many provided only a range of years (50/252, 20%). Thus, only 40% of articles (102/252) reported year of sampling (or greater precision).

For genetic data sets from 2011 onwards that were able to be recreated (178 from a total of 228 geographically relevant data sets, Fig. 2a), Fig. 2(b) illustrates the varying levels of precision of archived spatio-temporal metadata. The proportion of data sets available for repurposing will vary depending on the spatio-temporal needs of the new study. If temporal information is not required and if authors are willing to use locality text information, in addition to geographic coordinates, a large proportion of recreatable data sets could be reused (83%; 148/178). However, if latitude and longitude are required, fewer data sets are repurposable (64%; 115/178), and if latitude and longitude along with year of sampling or better are desired, a much smaller pool of data sets are available for repurposing (21%; 35/178).

Discussion

Policies mandating public data archiving have clearly increased archiving of genetic data, as shown in Fig. 1 (see also Vines *et al.* 2013). These developments in *Molecular Ecology* align with a sociological shift towards data sharing in ecology and evolution (e.g. Jones *et al.* 2006; Poisot *et al.* 2013; and discussed in Constable *et al.* 2010). In general, researchers in the fields of molecular ecology, phylogenetics and genomics have accepted this practice as fundamental to the requirement of reproducibility in science. Several institutions exist with the intention of making molecular genetic data publically accessible (e.g. NCBI, EBI, datadryad.org, http://www.free-the-data.org/ etc.); however,

© 2015 John Wiley & Sons Ltd

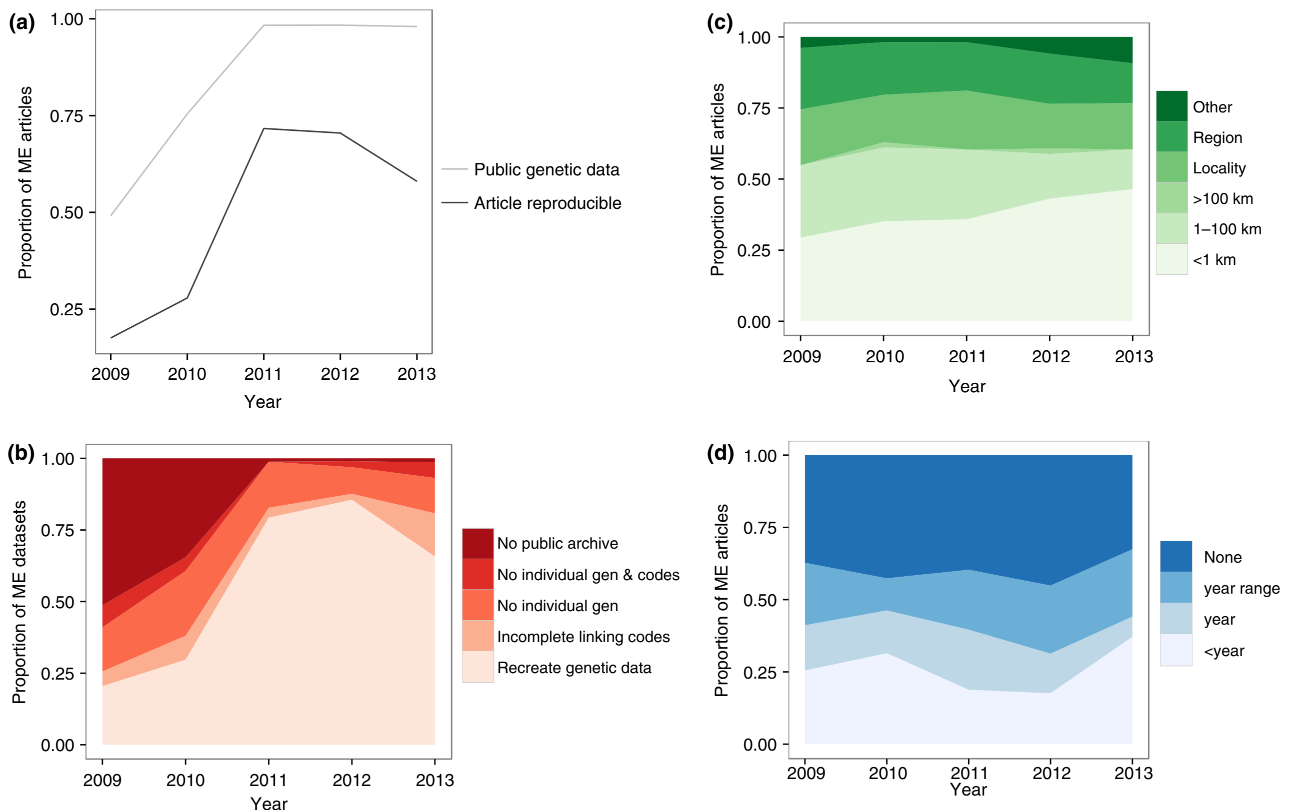


Fig. 1 Public availability of genetic data and associated spatio-temporal information. Results are based on all articles in both July and December issues for the year indicated (289 articles). (a) Rates of public data archiving and 'reproducibility' over time. 'Public genetic data' refers to the deposition of any kind of genetic data in a public database or publication. 'Article reproducible' refers to articles for which sufficient information was provided such that all analyses presented in the original article could be reproduced (see text for a full definition). (b) Rates of genetic data set archiving over time (n = 419). Genetic data sets were assessed as 'recreatable', and if not, the reason for this was classed as either: no public archive of genetic data, incomplete linking codes, no individual genetic data, or both no linking codes and no individual genetic data. c) Precision of geographic information in articles over time. For articles where geography was relevant (i.e. not captive bred/experimental, a total of 252 articles), we determined with what precision geographic information was made available: 'nothing/other', text only regional, text only locality, or coordinates > 100 km, 1–100 km, and <1 km precision. d) Precision of time of sampling information over time. Articles where time of sampling was relevant (252 articles) were graded on the amount of information provided: nothing, a range of years, year, and greater precision than a year.

data utility and linkages to other biodiversity databases are limited by submission formats and ontologies (Yilmaz *et al.* 2011; Walls *et al.* 2014). We argue that in order for molecular ecological data to be truly accessible to the public, at a minimum, individual genotypes should be recoverable and linked to geographic and temporal information. Our study indicates that voluntary rates of supplying this information could be substantively improved (post-JDAP introduction in 2011, 21% of genetic data sets could not be recreated, 45% of data sets provided no temporal information, and 40% no geographic coordinates). Both the JDAP and the *Molecular Ecology* policy on data archiving (see Box 1) emphasize that all data supporting the publications be available (not just genetic data files), which is consistent with our opinion. Thus, we suspect that shortcomings in full implementation stem from misinterpretation of these data archiving policies, difficulties in cross-referencing without clear standards or appropriately structured databases, unintentional oversights of busy people, and poor (self-) regulation of the field. Undoubtedly, we are also personally guilty of inadequate data archiving.

There are many reasons why spatio-temporal metadata may not be associated with genetic data. The location of samples may not be deemed relevant, such as for captively reared or artificially selected organisms. In other cases, the original time and place of genetic sampling will be unknown. This might occur where samples have been 'inherited' from previous projects, deposited in museums with locality unknown, or collected in such a manner that a precise locality cannot be determined (e.g. markets). Occasionally, locations of endangered species or sites of archaeological importance might be withheld from public release (see Rausher *et al.* 2010).

In other cases, the place and time of genetic sampling are known, but these metadata are not publically archived. Data submission can be a lengthy process, and fast ways to

© 2015 John Wiley & Sons Ltd

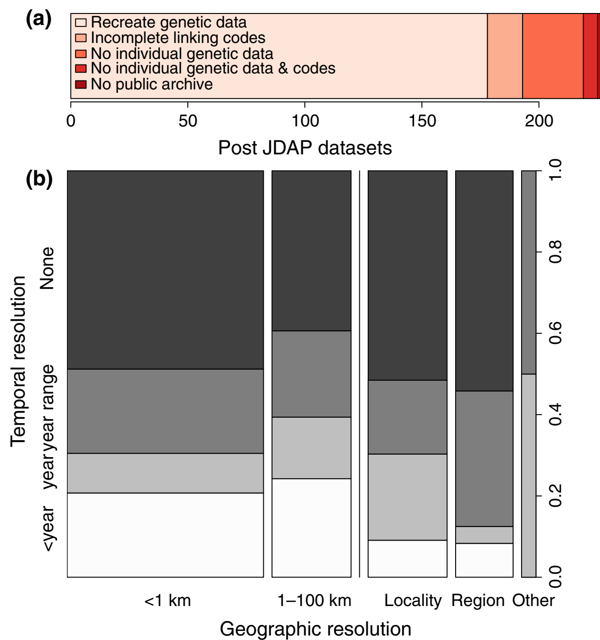


Fig. 2 Public archive rates, spatio-temporal data availability and opportunities for 'repurposing' of data sets (from 228 articles for which geography was relevant to the article objective published post-JDAP): (a) This bar represents the number of data sets with: no publically archived metadata (3/228), insufficient information to recreate individual genetic data (26/228), did provide individual genetic data, but without codes to link all the provided data together to recreate genetic data (16/228), both no linking codes and no individual genetic data (6/228). (b) The remaining 'recreatable' data sets (n = 178) are presented in terms of the precision of the available spatio-temporal data to indicate the data available for repurposing based on a researcher's requirements. Categories used are described in Fig. 1.

include metadata are often not obvious. Many popular population genetic data formats have no method for appending metadata (e.g. Arlequin, GENEPOP, Nexus and STRUCTURE formats). Some formats do, however, allow the inclusion of spatial information (GENALEX, Peakall & Smouse 2006; geneGIS, Dick *et al.* 2014; TESS, Chen *et al.* 2007). Unfortunately, for many of the loci employed in molecular ecology studies (especially microsatellites and AFLPs), there are no standard data repositories; thus, extra care is required in preparing archived files for these data. Additionally, because manuscript acceptance is typically decoupled from public data submission, often changes in the reference codes linking genetic data to metadata creep in during the revision process.

To improve the standards of public genetic data and spatio-temporal metadata in our field will require the effort of all parties: authors, reviewers, journals, institutions, public data repositories and the *Molecular Ecology* community as a whole. Based on the results from this study, we recommend that best practice for genetic data archiving for most *Molecular Ecology* studies (consistent with the JDAP) should include the following: (i) genetic data files that present individual genotypes, (ii) unified reference codes identifying individuals across any archived data sets from a single publication, (iii) year (ideally date) of sampling, and (iv) sampling locations provided as geographic coordinates.

In particular, we stress the need for higher community standards regarding geographic reporting with the expectation that spatial information be provided as georeferenced coordinates (presently missing in 40% of examined articles). The best practice for spatial data should include both a text description of the locality and geographic coordinates (including a description of the system used), as several location names are shared worldwide (e.g. Bird Rock), and/or only locally known (e.g. Bob's corner; discussed in a Anonymous 2008). In some locations, the use and precision of GPS is limited due to signal weakness and/or disturbance (i.e. underwater, under forest canopy, little satellite coverage). However, in such situations, geographic coordinates can be complemented with an estimate of spatial uncertainty. Tools to facilitate the estimation of uncertainty are available (e.g. GeoLocate, http://www.museum.tulane.edu/geolocate/) and are already incorporated in record keeping protocols for other forms of biological data (e.g. VertNet, Constable *et al.* 2010).

Preferably, genetic data should be deposited in appropriate repositories, rather than as supplemental files, which have been shown to decay over time (Evangelou *et al.* 2005; Anderson *et al.* 2006). Structured repositories with controlled ontologies can be efficiently queried and searched by end–users, and there are growing efforts to link genetic and other biodiversity databases via shared ontologies (see Introduction), preserving long-term value to the field. Where possible, spatial and temporal information should accompany database submissions; for example, these data can be included in NCBI records. In the short term, however, many *Molecular Ecology* data will not find an obvious home in a structured repository, and thus, flexible methods of data archiving (such as DataDryad) are extremely valuable. We recommend that authors prepare files in line with the recommendations listed above (utilizing commonly used genotype based files, consistent codes, date and location of sampling) along with an overarching readme file (see Whitlock *et al.* 2010) and review these files at the time of final submission. A quick check that these minimal elements are available could be undertaken by handling editors (based on our reviews of studies here, we found that well-prepared files can be summarily checked in a few minutes).

The last 5 years have shown a massive increase in the public archiving of genetic data. Despite these positive developments, many of the studies published in *Molecular Ecology* today are not reproducible, a central tenet of public archiving. *Molecular Ecology* represents one of the leaders in the call for essential data archiving, so this situation is likely worse for journals without clear and enforced data access policies. Additionally, many studies do not include geographic coordinates, or even year of sampling, restricting the future reuse of these genetic data. We advocate 'a

© 2015 John Wiley & Sons Ltd

higher expectation for the quality and quantity of descriptive data' (Field 2008). How this is best achieved is open to debate. Whether higher rates and quality of spatio-temporal data can be achieved through raised awareness and standards, without explicit mandates, remains to be seen. We do know that careful archiving of genetic data with associated spatio-temporal data *now* will result in a much more valuable legacy for future research. To fully realize the future potential of this data legacy, there should now be a greater push to link spatio-temporal metadata to genetic data and to develop standards and repositories that facilitate data deposition, curation and searchability.

Acknowledgements

LCP was funded by a University of Queensland Women's Postdoctoral Research Fellowship. LL was funded by a New Zealand Allan Wilson Centre for Molecular Ecology and Evolution Postdoctoral Research Fellowship. SBC was funded by a postdoctoral grant from Fundacão para a Ciência e Tecnologia (FCT) (SFRH/BPD/74423/2010), and through the project PTDC/BIA-BIC/118624/2010-FCOMP-01-0124-FEDER-019676, supported by FEDER funds through the Operational Programme for Competitiveness Factors – COMPETE and by National Funds through FCT. The authors thank J. Sheehan with assistance in retrieving information from manuscripts.

References

Anderson NR, Tarczy-Hornoch P, Bumgarner RE (2006) On the persistence of supplementary resources in biomedical publications. *BMC Bioinformatics*, **7**, 260.

Anonymous (2008) A place for everything. *Nature*, **453**, 2.

Australian Research Council (2013) *ARC Open Access Policy - Version 2013.1*. Australian Research Council AG, Canberra, ACT, Australia.

Baker CS (2013) Journal of heredity adopts joint data archiving policy. *Journal of Heredity*, **104**, 1.

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, **7**, 747–756.

Constable H, Guralnick R, Wieczorek J *et al.* (2010) VertNet: a new model for biodiversity data sharing. *PLoS Biology*, **8**, e1000309.

Cranston K, Harmon LJ, O'Leary MA, Lisle C (2014) Best practices for data sharing in phylogenetic research. *PLOS Currents Tree of Life*, **6**. doi: 10.1371/currents.tol.bf01eff4a6b60ca4825c69293dc59645.

Dick DM, Walbridge S, Wright DJ *et al.* (2014) geneGIS: Geoanalytical Tools and Arc Marine Customization for Individual-Based Genetic Records. *Transactions in GIS*, **18**, 324–350.

Dugan VG, Emrich SJ, Giraldo-Calderon GI *et al.* (2014) Standardized metadata for human pathogen/vector genomic sequences. *PLoS ONE*, **9**, e99979.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology*, **17**, 1170–1188.

Evangelou E, Trikalinos TA, Ioannidis JP (2005) Unavailability of online supplementary scientific information from articles published in major journals. *FASEB Journal*, **19**, 1943–1944.

Fairbairn DJ (2011) The advent of mandatory data archiving. *Evolution*, **65**, 1–2.

Field D (2008) Working together to put molecules on the map. *Nature*, **453**, 978.

Gaither MR, Bowen BW, Toonen RJ (2013) Population structure in the native range predicts the spread of introduced marine species. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20130409.

Jones MB, Schildhauer MP, Reichman OJ, Bowers S (2006) The new bioinformatics: integrating ecological data from the gene to the biosphere. *Annual Review of Ecology and Systematics*, **37**, 519–544.

Kodama Y, Shumway M, Leinonen R, International Nucleotide Sequence Database (2012) The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Research* **40**, D54–D56.

Leebens-Mack J, Vision T, Brenner E *et al.* (2006) Taking the first steps towards a standard for reporting on phylogenies: minimum information about a phylogenetic analysis (MIAPA). *OMICS: A Journal of Integrative Biology*, **10**, 231–237.

Lin J, Strasser C (2014) Recommendations for the role of publishers in access to data. *PLoS Biology*, **12**, e1001975.

Magee AF, May MR, Moore BR (2014) The dawn of open access to phylogenetic data. *PLoS ONE*, **9**, e110268.

Marques AC, Maronna MM, Collins AG (2013) Putting GenBank data on the map. *Science*, **341**, 1341.

Moore AJ, McPeek MA, Rausher MD, Rieseberg L, Whitlock MC (2010) The need for archiving data in evolutionary biology. *Journal of Evolutionary Biology*, **23**, 659–660.

National Institutes of Health (2003) *NOT-OD-03-032: Final NIH Statement on Sharing Research Data*. Department of Health and Human Services National Institutes of Health Office of Extramural Research, Rockville, Maryland.

Parr CS, Guralnick R, Cellinese N, Page RD (2012) Evolutionary informatics: unifying knowledge about the diversity of life. *Trends in Ecology and Evolution*, **27**, 94–103.

Peakall ROD, Smouse PE (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Poisot T, Mounce R, Gravel D (2013) Moving toward a sustainable ecological science: don't let data go to waste!. *Ideas in Ecology and Evolution*, **6**, 11–19.

Rausher MD, McPeek MA, Moore AJ, Rieseberg L, Whitlock MC (2010) Data archiving. *Evolution*, **64**, 603–604.

Sidlauskas B, Ganapathy G, Hazkani-Covo E *et al.* (2010) Linking big: the continuing promise of evolutionary synthesis. *Evolution*, **64**, 871–880.

Stoltzfus A, O'Meara B, Whiteacre J *et al.* (2012) Sharing and re-use of phylogenetic trees (and associated data) to facilitate synthesis. *BMC Research Notes*, **5**, 574.

Vandergast AG, Bohonak AJ, Hathaway SA, Boys J, Fisher RN (2008) Are hotspots of evolutionary potential adequately protected in southern California? *Biological Conservation*, **141**, 1648–1664.

Vines TH, Andrew RL, Bock DG *et al.* (2013) Mandated data archiving greatly improves access to research data. *FASEB Journal*, **27**, 1304–1308.

Walls RL, Deck J, Guralnick R *et al.* (2014) Semantics in support of biodiversity knowledge discovery: an introduction to the biological collections ontology and related ontologies. *PLoS ONE*, **9**, e89606.

White EP, Baldridge E, Brym ZT *et al.* (2013) Nine simple ways to make it easier to (re)use your data. *Ideas in Ecology and Evolution*, **6**, 1–10.

Whitlock MC (2011) Data archiving in ecology and evolution: best practices. *Trends in Ecology and Evolution*, **26**, 61–65.

© 2015 John Wiley & Sons Ltd

Whitlock MC, McPeek MA, Rausher MD, Rieseberg L, Moore AJ (2010) Data archiving. *The American Naturalist*, **175**, 145–146.

Wood DA, Vandergast AG, Barr KR *et al*. (2012) Comparative phylogeography reveals deep lineages and regional evolutionary hotspots in the Mojave and Sonoran Deserts. *Diversity and Distributions*, **19**, 722–737.

Yilmaz P, Kottmann R, Field D *et al*. (2011) Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nature Biotechnology*, **29**, 415–420.

All authors were involved in formulating the ideas for this manuscript, data entry and writing the manuscript. CR instigated the manuscript and performed the R analyses on 'reproducible articles'. LP performed the majority of analyses and coordinated the final text. All the authors are interested in comparative spatial genetics and use repurposed data in their research. LCP, JK and CR are usually found at -27.498 153.012, LL at -36.734 174.703 and SCB at 41.181 -8.602 (WGS84).

---

doi: 10.1111/mec.13254

Data accessibility

An Excel spreadsheet of all the *Molecular Ecology* articles and data sets included in our analyses, along with a read-me file, R script and R data file, are available on Dryad doi:10.5061/dryad.kg943.2

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Use of different marker types across *Molecular Ecology* sections.

Fig. S2. The number of data sets without publically archived data, as a function of genetic marker type.

© 2015 John Wiley & Sons Ltd

Supplementary Data

Figure S1 Use of different marker types across *Molecular Ecology* sections. Data represents the proportion of datasets for each genetic marker within each subsection of ME. A total of 412 data sets from 285 articles were analysed ('Fast Track' (1) and 'From the Cover' (3) articles were excluded due to low sample size). Genetic marker names were abbreviated from the following: microsatellites, mitochondrial sequences, nuclear sequences, chloroplast sequences, amplified fragment length polymorphisms, and single nucleotide repeat data. Marker categories with less than 15 data sets were pooled into "other": EST (2), allozymes (4), microarray (4), next generation sequences (3), proteins (1), RFLPs (1), and RNA sequences (1). The width of each column is proportional to the number of genetic data sets found within each subsection.

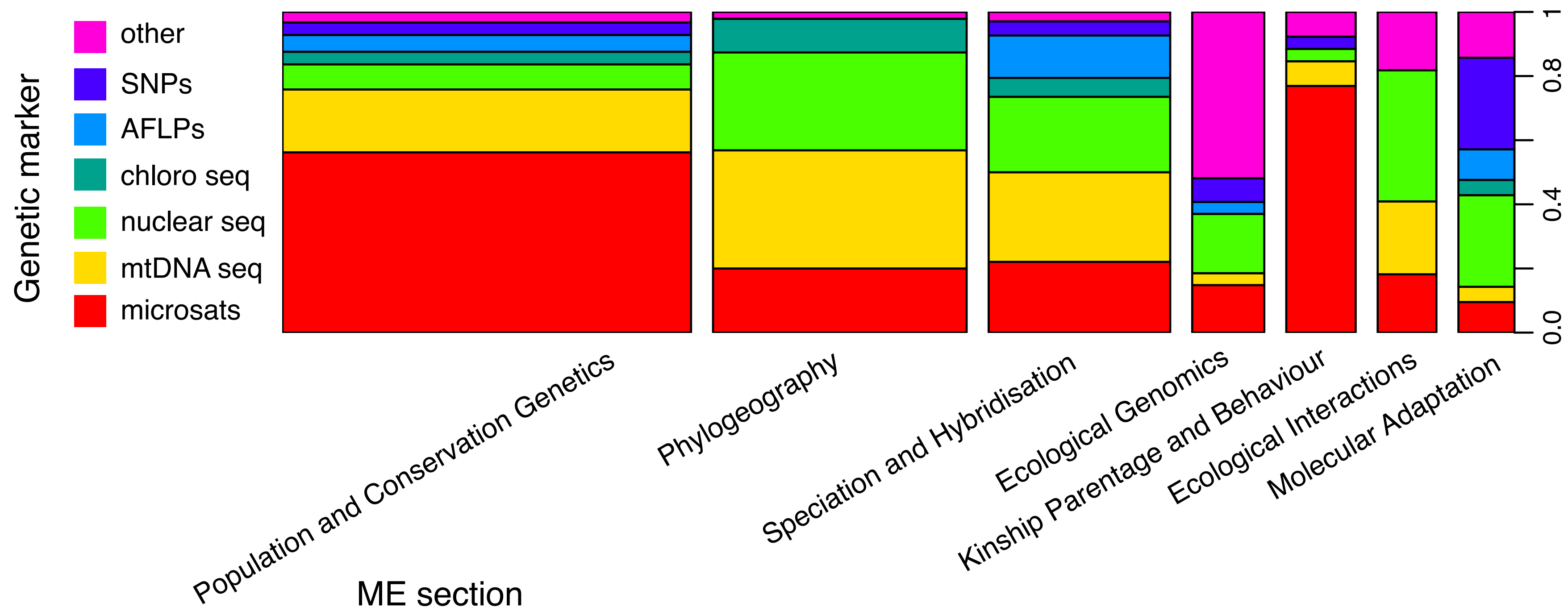


Figure S2. The number of datasets without publically archived data, as a function of genetic marker type. Here we define 'publicly available' as either accessible in public databases or supplied as supplementary data. Non-sequence data included: microsatellites (136), SNP (15), AFLPs (19), RFLPs (1), allozymes (4), proteins (1), microarray (4), EST (2). Sequence data included: mtDNA sequence (91), nuclear sequence (71), chloroplast sequence (21), RNA sequence (1), next-generation sequences (3).