THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Seascape genetics of Indo-Pacific reef organisms: methodologies and case studies for an emerging discipline

Libby Liggins

BSc, MSc

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2014

The School of Biological Sciences

Abstract

Which factors shape population connectivity and the consistency of their influence across species and seascapes is largely unresolved. This thesis takes a comparative genetic approach to investigate the influences of seascape topology (past sea level changes, contemporary oceanography, and geography) and species biology on the genetic connectivity among coral reefs around Australia, the Indo-Australasian Archipelago (IAA), and the wider Indo-Pacific. I focus on a suite of common, co-distributed coral reef fishes and invertebrates that differ in life history characteristics. In the introductory chapter of my thesis I discuss population connectivity in marine systems, the contribution of seascape genetics to marine population connectivity research, and the state of our knowledge in the Indian and Pacific Oceans (Chapter One). Chapters Two and Three review the field of seascape genetics and the methodologies used in seascape genetic studies, respectively. My empirical research (Chapters Four-Six) takes a broad, yet integrated approach to describe and understand the processes underlying seascape genetic patterns around Australia, the IAA, and the Indo-Pacific. First, I investigate historical spatial genetic (mitochondrial DNA, mtDNA) patterns and processes using a comparative, multi-species framework (Chapter Four). I extend typical methods of comparative phylogeography to include: a matrix comparison method that allows the quantitative characterization of spatial genetic patterns; and a multiple regression modeling approach to identify the processes underlying the spatial genetic patterns for each species. I find that despite being subjected to common geographic and seascape influences, spatial genetic patterns differ across four common coral reef fishes (*Acanthurus triostegus*, *Dascyllus trimaculatus*, *Halichoeres hortulanus*, *Pomacentrus coelestis*). Although species with similar dispersal potential (based on egg type and pelagic larval duration) have the most similar spatial genetic patterns, the seascape features and processes (e.g. previous landbridges and oceanographic distances) underlying the genetic patterns are not always shared among species. Second, I characterize the latitude-wide genetic patterns of one species (*P. coelestis*) and extend methods typically used in such investigations by borrowing concepts and measures more commonly used for analyses of community composition (Chapter Five). Genealogical analyses reveal that levels of population genetic diversity in the core of the species range are elevated by the co-occurrence of two cryptic clades. Furthermore, the application of partitioned β-diversity measures and nestedness analyses reveal that differing demographic processes underlie the genetic patterns observed at the northern and southern latitudinal peripheries of the species range. Last, I focus on edge-of-range genetic patterns for two tropical echinoderm species (*Acanthaster planci* and *Tripneustes gratilla*) at a little-studied high latitude peripheral population, Kermadec Islands, New Zealand (Chapter Six). I find surprisingly concordant patterns across species indicating that despite being marginal habitat

for tropical species, the Kermadec Islands populations are maintained by self-recruitment and not immigration over contemporary timescales. Through investigating genetic patterns over a range of species and seascapes, I have evaluated several long-standing spatial genetic hypotheses and their relevance to seascape genetics. I have identified shortcomings in our current understanding of seascape genetic patterns and the limitations of our analytical toolset. As such, I demonstrate how novel approaches can effectively identify the processes underlying spatial genetic patterns in marine systems.

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 General Award Rules of The University of Queensland, immediately made available for research and study in accordance with the *Copyright Act 1968*.

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 journal articles

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: 937-949.

Riginos C, **Liggins L**. 2013. Seascape genetics: populations, individuals, and genes marooned and adrift. *Geography Compass* 7: 197-216.

Liggins L, Treml EA, Riginos C. 2013. Taking the plunge: an introduction to undertaking seascape genetic studies and using biophysical models. *Geography Compass* 7: 172-196.

Torkkola J, Riginos C, **Liggins L**. 2013. Regional patterns of mtDNA diversity in *Styela plicata*, an invasive ascidian, from Australian and New Zealand marinas. *Marine and Freshwater Research* 64: 139-145.

Huelsken T, Keyse J, **Liggins L**, Penny S, Treml EA, Riginos C. 2013. A novel widespread cryptic species and phylgeographic patterns within several giant clam species (Cardiidae: *Tridacna*) from the Indo-Pacific Ocean. *PLoS One* 8: e80858.

Liggins L, Gleeson L, Riginos C. 2014. Evaluating edge-of-range genetic patterns for tropical echinoderms, *Acanthaster planci* and *Tripneustes gratilla*, of the Kermadec Islands. *Bulletin of Marine Science* 90: 379-397.

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

Richards ZT, **Liggins L**. Hermatypic corals and crown-of-thorns starfish of the Kermadec Islands. *Bulletin of the Auckland Museum*: in press.

Technical report

Albert S, Grinham A, Bythell J, Olds A, Schwarz A, Abernethy K, Aranani K, Sirikolo M, Watoto C, Duke N, McKenzie J, Roelfsema C, **Liggins L**, Brokovich E, Pantos O, Oeta J, Gibbes B. 2012. Building social and ecological resilience to climate change in Roviana, Solomon Islands, The University of Queensland. *A report prepared for the Australian Government and Solomon Islands Government Pacific Strategy Assistance Program*.

Publications included in this thesis

Riginos C, **Liggins L**. 2013. Seascape genetics: populations, individuals, and genes marooned and adrift. *Geography Compass* 7: 197-216. – incorporated as Chapter Two.

Contributor	Statement of contribution
Liggins L	Designed the review structure (40%) Wrote and edited the paper (40%)
Riginos C	Designed the review structure (60%) Wrote and edited the paper (60%)

Liggins L, Treml EA, Riginos C. 2013. Taking the plunge: an introduction to undertaking seascape genetic studies and using biophysical models. *Geography Compass* 7: 172-196. – incorporated as Chapter Three.

Contributor	Statement of contribution
Liggins L	Designed the review structure (70%) Wrote and edited the paper (70%)
Treml EA	Wrote and edited paper (10%)
Riginos C	Designed the review structure (30%) Wrote and edited the paper (20%)

Liggins L, Gleeson L, Riginos C. 2014. Evaluating edge-of-range genetic patterns for tropical echinoderms, *Acanthaster planci* and *Tripneustes gratilla*, of the Kermadec Islands. *Bulletin of Marine Science* 90: 379-397. – incorporated as Chapter Six.

Contributor	Statement of contribution
Liggins L	Conducted field sampling (100%) Statistical analyses of data (100%) Wrote and edited the paper (85%) Conducted laboratory work (50%)
Gleeson L	Conducted laboratory work (50%)
Riginos C	Wrote and edited the paper (15%)

Contributions by others to the thesis

Cynthia Riginos contributed to the conception and design of the project; funding; collection of specimens; interpretation of the data throughout the thesis; and writing and editing of Chapters Two, Three, Four, Five, and Six.

Eric A. Treml contributed to the conception and design of the project; funding; collection of specimens; the writing and editing of Chapter Three; and provided the biophysical model data used in Chapter Four.

Hugh P. Possingham contributed to the conception and design of the project; funding; and editing of Chapter Three and Five.

Lachlan Gleeson contributed to the laboratory work of Chapter Six.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

First, thank you to my supervisors. Cynthia Riginos thank you for being a great mentor – pushing me, encouraging me, and being patient. Eric A. Treml, I am grateful for you guiding me into this project. Hugh P. Possingham I appreciated your insightful overview. Thank you also to my committee: Lyn Cook and Margie Mayfield.

I am grateful to several collaborators that have provided research, logistical, and technical advice, have provided samples and data; and collaborated on various projects during my thesis term. In particular I thank David Booth, Will Figueira, Eric D. Crandall, Michelle Gaither, Brian Bowen, Mark McCormick, Paul Barber, Elizabeth Sbrocco, Gert Wörtheide, Nina Yasuda, Linda Tonk, David Harris, Lexa Grutter, Scott Burgess, John Dwyer, Jeff Leis, Fred Allendorf, and Joe Bennett.

Thank you to the various government agencies, non-governmental organizations, research institutes, and other research initiatives for permitting and logistical support: the Coral Triangle Support Partnership (Timor-L'este); the Ministério da Aquicultura e Pescas, Direcção Nacional de Pescas e Aquicultura, Departmento de Quarentena das Pescas; Solomon Island's Government's Pacific Strategy Assistance Program; Roviana Conservation Foundation (Solomon Islands); Government Ministry of Education and Human Resource Development and Ministry of Fisheries and Marine Resources (Solomon Islands); National Research Institute (Papua New Guinea); the Department of Foreign Affairs and Immigration (Papua New Guinea); the Department of Environment and Conservation (Papua New Guinea); the Australian Government Department of Sustainability, Environment, Water, Population and Communities; Australian Customs and border control; the Marine Division of the Australian Government Department of Sustainability, Environment, Water, Population and Communities; the Northern Territory Government Department of Resources, the Western Australia Department of Environment and Conservation; the Ministry of Agriculture and Food, Forests, and Fisheries (Tonga); the Ministry of Natural Resources and Environment (Tuvalu); Ministry of Primary Industries Fisheries Department (Fiji); the Australian Museum Lizard Island Research Station; the Heron Island Research Station; the Great Barrier Reef Marine Park Authority and Queensland Parks and Wildlife Marine Parks; the Queensland Government Department of Primary Industries; the Auckland Museum; the Sir Peter Blake Trust; Tagai State College; the Torres Strait Regional Authority; the New Zealand Navy; Ailan Awareness (Papua New Guinea); National Fisheries College (Papua New Guinea); Freeflow diving (Timor-l'este); the National Evolutionary Synthesis Center (NESCent); and the Australian Fisheries Management Authority. (Specifics of permits for the collections and research work are detailed in the appropriate chapters.) I am particularly grateful for the logistical support that has been provided

by the following individuals: John Aini, Simon Albert, John Kinch, Wayne Lovell, Ann Turner, Michael Phillips, Cameron Beal, Annabel Jones, Vic McGrath, and Andrew Denzin.

I am indebted for the help and fun provided by my fieldwork buddies: J. David Aguirre, Morgan Jimuru, Ian M. McLeod, Anna Mirams, Shane Penny, Rui Pinto, Tane Sinclair-Taylor, Pete Waldie, Michele X. Weber, Stephen, Lavud, Takenda, Mark Preist, Fabio Cortesi, Kurt Davis, Derek Sun, Eva McClure, Roy Pearce, Clinton Duffy, Helen Bostok, Stephen Ullrich, Andrew Berry, Rochelle Constantine, and Tom Trnski. I look forward to the next adventure!

A big thank you to my lab mates for sharing this research experience in the lab and in the field: Carla Meers, Anna Mirams, Jude Keyse, Jenny Giles, Lachlan Gleeson, Jody Shields, Carrie Sims, Penny Mills, Lisa Pope, Tom Huelsken, Dean Blower, Andrew Mather, and James Hereward. James thank you for your endless patience and knowledge in all things lab related - you were a suitable inaugural 'King of the Lab'!

I am grateful for the generous financial support I have received during my PhD from the Australian Government (Australian Postgraduate Award) and Queensland Government (Smart Futures PhD Scholarship). My work has been supported by the following: Australian Research Council (DP0878306, to Cynthia Riginos and Hugh P. Possingham); an Explorer's Club Exploration Fund; the Sea World Research and Rescue Foundation (SWR/1/2012, to Cynthia Riginos and Libby Liggins); a Paddy Pallin Foundation and The Foundation for National Parks and Wildlife Science Grant; an Ecological Society of Australia Student Research Grant; the Lerner Gray Memorial Fund of the American Museum of Natural History; and a Great Barrier Reef Marine Park Authority's Science for Management Award.

I am thankful to the School of Biological Sciences (in particular Gail Walters) and the Graduate School of The University of Queensland for their support. My PhD experience has been made richer via the generous conference travel funding awarded by these institutions, and a special outreach grant from the School of Biological Sciences that enabled us to complement our research in the Torres Strait with outreach activities. In general, I am grateful to the communities of the School of Biological Sciences and The University of Queensland for providing research support and a rich learning environment.

I am grateful for the opportunities and activities I have been involved with during candidature that enriched my science experience. These activities include: the Sir Peter Blake Trust Inaugural Blake Expedition to the Kermadec Islands; the Academy of Technological Sciences and Engineering Young Science Ambassador scheme; the Wonder of Science program; the Wentworth Group of Concerned Scientists; the National Evolutionary Synthesis Center (NESCent) Graduate Fellowship scheme; the NESCent working group, 'Advancing genetic diversity research in the

Indian and Pacific Oceans'; and the Great Barrier Reef Marine Park Authority's Future Leaders Eco Challenge. Thank you to those who have provided and shared these valuable experiences with me.

Thank you to my family and friends for tolerating my absence. I missed you very much! Thank you to Marina and Coco for taking me on as your own. I haven't minded inheriting a few more brothers either! Especially thanks to Dan for looking out for me. Dave, thank you for all of the above and much more.

Keywords

phylogeography, marine landscape genetics, seascape genetics, coral reef community, population connectivity, dispersal, biophysical model, Indo-Pacific, range periphery

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 060302, Biogeography and Phylogeography, 80%

ANZSRC code: 060411, Population, Ecological and Evolutionary Genetics 20%

Fields of Research (FoR) Classification

FoR code: 0603, Evolutionary Biology, 80%

FoR code: 0604, Genetics, 20%

Table of Contents

CHAPTER ONE. General introduction

Many species are distributed throughout the coral reefs of the Indian and Pacific Oceans (Spalding et al. 2007). This pattern indicates that these populations must have been connected at some stage. To what degree these connections are the result of historical and contemporary population connectivity is important for understanding the evolutionary and ecological dynamics of these species and coral reef communities more generally (Palumbi 1994, Caley et al. 1996, Mora et al. 2003, Lester et al. 2005). Whether populations are maintained by on-going dispersal, or are self-sustaining, is also critical for species management and coral reef conservation planning (Mora and Sale 2002, Gell and Roberts 2003, Palumbi 2004, Jones et al. 2009). Such questions concerning the population connectivity of reef-associated organisms are not isolated to coral reefs of the Indo-Pacific, but are applicable to marine systems throughout the world. Most of the world's marine biodiversity is reef-associated organisms (Gray 1997) and therefore, an on-going research priority of the marine sciences is to understand the general scale and determinants of population connectivity in these organisms (Cowen and Sponaugle 2009, Leis et al. 2011, Kool et al. 2012).

This thesis contributes to the study of marine population connectivity via the field of seascape genetics (Selkoe et al. 2008, Riginos and Liggins 2013). The goals of this thesis were to provide empirical case studies and methodologies of study design and data analysis appropriate for seascape genetic studies. In particular, this thesis focuses on the coral reef fauna of the Indian and Pacific Oceans. An introduction to population connectivity in marine systems, the contribution of seascape genetics to marine population connectivity research, and the state of our knowledge in the Indian and Pacific Oceans are included in this chapter (Chapter One). Chapters Two and Three review the field of seascape genetics and the methodologies used in seascape genetic studies, respectively. (Chapters Two and Three are written as partnered papers and are included in this thesis in their published forms.) The contributions of the subsequent chapters to the overall goals of this thesis are outlined at the end of this chapter (Chapter One).

Population connectivity in marine systems

The definition of population connectivity varies depending on the methodological context or applied purpose (Lowe and Allendorf 2010, Kool et al. 2012). In this thesis I refer to population connectivity as the transfer of individuals among populations, entailing both the dispersal and successful recruitment of an individual into the recipient population, whether they go on to reproduce or not (Pineda et al. 2007, Leis et al. 2011). By this definition therefore, population connectivity in marine systems can be affected by factors that influence both the dispersal and recruitment of individuals. These factors can

be categorized as those that are biological and those that are physical, however it is the interaction of the two that ultimately determines population connectivity.

Biological factors that influence marine population connectivity

Although life histories vary widely in the sea, most reef-associated marine organisms have a bi-partite life history whereby the pre-metamorphic stages (i.e. eggs and larvae) are highly dispersive and the post-metamorphic stages (i.e. juvenile and adult) are relatively sedentary and reef associated (Leis 1991). Thus, the early life history stages determine any potential connectivity among adult populations of reef-associated organisms (Kinlan and Gaines 2003). These early life histories vary among species in ways that can influence their relative potential for population connectivity (reviewed in Leis 1991). Species typically have either benthic eggs that are attached to the substrate or pelagic eggs that can disperse while they develop. Larvae that hatch from both egg types disperse in the water column for a period of minutes to months (Shanks 2009). The length of the pelagic larval duration (PLD) is bounded by the minimum required development time for that species (the pre-competency period) and whether larvae can feed while in the plankton, or rely on maternal provisioning (planktotrophic versus lecithotrophic). Thus, these two early life history traits, egg type and PLD, are intuitively expected to influence the dispersal potential of a species, and therefore the scale over which populations are connected. These are hypotheses that have been frequently tested in meta-analyses (discussed in Chapters Two and Four) and are empirically tested in Chapter Four of this thesis.

Other early life history traits and larval behaviors also affect population connectivity, however the nature of their influence on population connections can vary. For example, the swimming ability of a larva may enable it to disperse a greater distance, or contrarily, it may allow it to resist being advected away from its local reef. This hypothetical example also illustrates how larval preference or behaviors are important in determining population connectivity. In one of the pioneering examples in the growing field of marine behavioral larval ecology (Leis 2002), Gerlach et al. (2007) was able to experimentally demonstrate an olfactory preference in the larvae of two fish species; whereas *Apogon doederleini* preferred the water from its natal reef, *Pomacentrus coelestis* was attracted to water sourced from reef habitat, but not its natal reef. Thus, although larval swimming ability, sensory abilities, and behavior are important in establishing population connections in the marine system, their influences likely vary among species.

Once larvae reach a reef, further locally relevant biological factors will influence whether they successfully recruit into the population. The settlement of larvae to the reef involves a series of dramatic ecological, physiological, and behavioral changes (McCormick et al. 1997) and mortality can be extremely high during this period (Almany and Webster 2006). Although mortality during settlement and subsequent recruitment has largely been considered stochastic, there has been recent revival of the idea that this phenomenon may not be random and deterministic processes could contribute (Planes and Lenfant 2002, Hamilton et al. 2008, Marshall et al. 2010). For example, “legacy

effects" have been suggested to bias recruitment in larval fish whereby the survival is greater in larvae that have undergone a less stressful pelagic phase (Shima and Swearer 2010). It has been suggested therefore, that locally derived larvae may have a greater chance of survival due to their shorter or less stressful pelagic larval phase. Alternatively, foreign larvae may be unable to settle successfully because they are not adapted to the local environment (Marshall et al. 2010), or suffer high-density blocking (Hewitt 1993) from the existing conspecific population (reviewed in Waters et al. 2013). Thus, the relationship between a species' dispersal potential, and a successful population connection is subject to many intervening biological factors.

Physical factors that influence marine population connectivity

Physical features of the seascape help determine the dispersal trajectories and recruitment success of the early life history stages of reef-associated organisms. Barriers and geographic distance in the sea can reduce population connectivity (see Fig. 1, Chapter Two for an overview of the physical features of the seascape that affect population connectivity and the temporal scale of their influence). The most obvious barriers to marine connectivity are contemporary landmasses and landmasses that were exposed intermittently during the Pleistocene due to lowered sea levels (~2.5mya-12,000ya). For example, significant genetic differentiation has been observed on either side of where the mainland Australia-Tasmania land bridge existed, as a result of restricted connectivity between those populations during the Pleistocene (Waters et al. 2005, Waters et al. 2007).

Permeable barriers are also common in the sea, whereby dispersal among populations may be restricted but not completely blocked. For coastal or reef-associated species, permeable barriers can be stretches of deep ocean or habitat disruptions that contribute to greater genetic population differentiation among populations than over the equivalent distance along a coastline or continuous reef (e.g. habitat discontinuity, Ayre et al. 2009). In the absence of any barrier, reef-associated marine species are often found to have a pattern of (genetic)-isolation-by-(geographic)-distance (IBD, Wright 1943) indicative of reduced population connectivity among increasingly distant populations. The slope of this IBD relationship is expected to scale with the dispersal potential of a species (i.e. egg type and PLD), however results have varied across studies (discussed in Chapters Two and Four). Collectively, barriers and geographic distance can be referred to as stable configurations that influence population connections (further described in Chapter Two).

In contrast to stable configurations, ocean currents can both increase and restrict population connectivity acting like conveyers (Mora and Sale 2002, e.g. the East Australian Current, Suthers et al. 2011) or forming barriers (Gilg and Hilbish 2003, e.g. the Mindanao and Halmahera Eddies, Barber et al. 2006, Kool et al. 2011). The influence of oceanographic features will sometimes relate to species biology. For example, the strong currents of the Mona Passage between Hispaniola and Puerto Rico (Baums et al. 2006) can be crossed by the larvae of some, but not all species (Shulman and Bermingham 1995). Furthermore, ocean currents are also temporally unstable (e.g. the Torres Strait

has seasonal currents, Wolanski et al. 1988, Margvelashvili et al. 2008). It has been shown that the location of a frontal boundary current for the two weeks prior to sampling influenced the genetic composition of kelp bass recruits (Selkoe et al. 2006). Hence, predicting the influence of ocean currents on marine population connectivity requires the simulation of larval dispersal events within a spatially (and possibly temporally) explicit model of oceanographic currents (e.g. Treml et al. 2008; discussed in Chapter Three).

Locally relevant physical or environmental factors will determine the successful recruitment of larvae into a receiving population. The impacts of these local physical factors may depend on the fitness of a particular individual relative to other conspecifics (i.e. local selection), or they could influence all individuals of a species equally. For example, some reef areas may be low quality habitat, and hence have low carrying capacity and low reproductive output affecting their connectivity with other populations across a species range (Treml et al. 2012). Such population dynamics are the basis of the Core-Periphery Hypothesis (da Cunha et al. 1950, Brussard 1984) that describes the impacts of population size and proximity to the range core on range-wide genetic patterns. The Core-Periphery Hypothesis predicts that population connectivity will be higher in the core of a species range and reduced toward the range periphery (Eckert et al. 2008; discussed in Chapter Five). In extreme cases such as at the edge of a species habitable range, populations could be completely disconnected from the rest of the species range over contemporary timescales. Alternatively, these reef regions may not sustain reproductive populations and therefore, rely on larval dispersal from core regions (Barton 2001; discussed in Chapter Six).

In summary, the early life history traits of reef-associated marine organisms influence their dispersal potential and physical features of the seascape help determine their dispersal trajectories. The interaction of several biological and physical factors will influence whether an individual larva successfully recruits to form a population connection. Thus, while there is the expectation that – species that vary in their early life history characteristics will have varying levels of population connectivity (Chapter Four); and regions that vary in their historical and contemporary seascapes will have different levels of population connectivity (Chapters Five and Six) – there are many factors at play, and these theoretical expectations require empirical evaluation.

Studying marine population connectivity

Direct observation of the dispersal of reef-associated marine organisms is near impossible due to the small size of dispersing eggs and larvae, and often large ocean distances among populations. For this reason, innovative approaches to measure population connectivity of reef-associated marine organisms have emerged and each method contributes uniquely to our knowledge base.

Much of our research into marine population connectivity has focused on individual populations, characterizing whether they are "open" and therefore highly connected to other populations, or "closed" and largely reliant on local reproduction and self-recruitment (Caley et al. 1996, Cowen 2000). The first compelling evidence for self-recruitment in marine populations was via methods of chemically-tagging fish otoliths (Jones et al. 1999) and using natural geochemical signatures in their otolith (Swearer et al. 1999). Adaptation of the chemical marking methods developed by Jones et al. (1999) have subsequently provided insight into the self-recruitment rates of populations across several fish species and seascape settings (e.g. Jones et al. 2005, Almany et al. 2007). Studies have found that it is common for 30-60% of larvae settling on a reef to be progeny of resident adults (Jones et al. 2009), suggesting populations are likely part of a metapopulation.

Applying these same chemical-tagging methods to metapopulation scenarios has been rare (Thorrold et al. 2007, Cowen and Sponaugle 2009) because they are labor intensive. In contrast methods that rely on "natural" tags such as those developed by Swearer et al. (1999) can be more easily applied over large distances and to many individuals. Differences in the elemental chemistry of otoliths have been used to infer the source of the larva across several seascapes by comparing the natal chemical signatures of the otolith core (e.g. Swearer et al. 1999, Hamilton et al. 2008). Importantly these methods have provided evidence for differential survival of larvae on a common reef according to their natal origin, and dispersal pathway (termed "legacy effects", Shima and Swearer 2010). However, the broad applicability of this method relies on there being a distinguishable and stable geochemical signature associated with each population, which is not always the case (Thorrold et al. 2007, Cowen and Sponaugle 2009). Furthermore, methods that depend on otolith-like structures have not been effectively developed for many invertebrates (S. Miller pers. comm.).

Other methods that have been applied to metapopulation scale connectivity include genetic methods. Specifically, parentage analysis and population assignment using individual genotypes based on hypervariable, neutral genetic markers have been used effectively at the single population-scale (e.g. Jones et al. 2005) and within metapopulations, including among marine protected area networks (Planes et al. 2009, Almany et al. 2013). These studies have provided valuable insight into dispersal kernels (capable of recording individual movements in excess of 30km, Almany et al. 2013), rates of self-recruitment, and metapopulation dynamics over short timeframes. In particular, larval dispersal within metapopulations has been highlighted as an important determinant of population viability by these studies, and recommendations for local-scale fisheries management have ensued (e.g. Almany et al. 2013). Nevertheless, these studies also report immigration from outside the focal study region, suggesting the relevance of population connectivity over larger spatial scales. Whereas increasing the spatial and temporal scale of studies that use parentage analysis and population assignment is likely not feasible, other genetic methods and dispersal simulations can be used to understand broadscale population connectivity patterns.

Simulations of dispersal pathways are an effective way of inferring likely population connectivity over large spatial and timescales (Paris and Cowen 2004). Unlike the previously discussed empirical methods that provide a snapshot of connectivity among sub-populations, simulations may be run over seasons and years within a spatially-explicit framework to describe: larval dispersal relevant to population demography and conservation management (Treml and Halpin 2012); and average patterns, variability, and rare events that influence the colonization and evolutionary dynamics of populations (e.g. Kool et al. 2011, Mora et al. 2011). Moreover, biophysical model simulations (e.g. Roughgarden et al. 1988, Treml et al. 2008) that enable the coupling of our biological knowledge of an organism (acquired via empirical studies) with the physical environment of the ocean have contributed to our understanding regarding general scales of population connectivity and traits that have a role in determining these scales (discussed in Chapter Three). For example, dispersal simulations have demonstrated that organisms with shorter PLDs should have more population genetic structure due to reduced population connectivity (Faurby and Barber 2012, Treml et al. 2012). Furthermore, it has been suggested that 95% of larvae disperse less than 155 km and spend only 13 days dispersing, regardless of their specific life history traits (Treml et al. 2012). However, biophysical models are essentially hypotheses of larval dispersal trajectories and require validation (Galindo et al. 2010). Genetic data has been suggested as one way to validate, or cross inform, such model predictions (reviewed in Chapter Three).

In contrast to previously discussed empirical methods for studying population connectivity, inference gained from genetic studies using phylogeography (Avise 2000) and population genetics is distinctly different. Firstly, patterns of genetic connectivity derived using these methods reflect the dispersal *and* successful recruitment of individuals (and their genes). Whereas, the capture of larvae or settlers for the analysis of otoliths, and a focus on the settling/recruiting cohort is necessary for population assignment and parentage analyses, methods that rely on population genetic theory largely use population-wide demographic signatures that are dependent on successful population connections that have been made over time. Secondly, these genetic approaches help reveal the relevance of population connectivity (i.e. migration) relative to other evolutionary processes (selection, mutation, and drift) that influence population demography and evolution. For example, in instances where there is no contemporary population connectivity, leveraging theoretical population genetic understanding of how migration (population connectivity) interacts with the process of drift, and how frequently mutations arise, we can gain an understanding of how long ago a population became disconnected. Thus, genetic studies can generate knowledge regarding historical and contemporary patterns of population connectivity.

The spatial (and temporal) genetic patterns inferred via genetic studies are well-suited to analyses interested in identifying the biological and physical factors that determine population connectivity. “Seascape genetics” is the discipline that formally address how spatially variable structural and environmental features influence genetic patterns of marine organisms and is the

primary discipline used in this thesis (reviewed in Chapter Two). There have been several seascape genetic studies of single species (discussed in Chapters Two), for which there is a wealth of analytical methods available (discussed in Chapters Three). The number of studies conducted in the field of seascape genetics is now large enough to facilitate meta-analyses that help us understand the generality of patterns in the seascape and the underlying biological factors (e.g. Riginos et al. 2011, 2014, Selkoe et al. 2014). For benthic marine organisms, meta-analytical approaches have highlighted the role of the early life history dispersal stage in forming patterns of genetic structure and geographic scales of genetic differentiation (discussed in Chapters Two and Four). However, determining the synergistic influence of biological and physical factors on connectivity patterns requires a multi-species, spatially explicit approach that has been rare (Dawson 2014; but see Chapter Four).

Ultimately the progression of our knowledge regarding marine population connectivity and the biological and physical factors that underlie patterns will rely on approaches over small scales that can form a mechanistic understanding, and those over broad scales that can identify the generality of patterns and processes though time. Genetic methods can be used at both these ends of the spectrum from parentage analysis and population assignment right through to phylogenetics. A particular strength of seascape genetics is that it generates data that can be used alongside other forms of inference, such as behavioral observations (Buston et al. 2009), demographic information (Lowe and Allendorf 2010), life history traits (Riginos et al. 2011, 2014), biophysical models (Crandall et al. 2012, 2014 – Appendix Two), and other spatial predictors. In particular, an opportunity exists to use multi-species investigations to better understand the role of species traits in spatially explicit (i.e. co-sampled) investigations, and in different seascape contexts.

Marine population connectivity of the Indian and Pacific Oceans

The Indo-Pacific Ocean is the largest biogeographic region in the world (Spalding 2007). Shallow reef marine biodiversity reaches its peak at the confluences of the Indian and Pacific Oceans in the Indo-Australasian-Archipelago (IAA), that comprises the "Coral Triangle" (including much of Indonesia, Malaysia, the Philippines, Brunei, Timor-Leste, Papua New Guinea, the Solomon Islands, and intervening seas) and the Timor Sea, the Arafura Sea and the Torres Strait in the north of Australia (Tittensor et al. 2010). Species richness incrementally decreases eastward across the Pacific Ocean and westward across the Indian Ocean from the IAA (Veron et al. 2009). There have been three major hypotheses posited to explain these biodiversity patterns: the Centre of Origin, meaning speciation is higher in the Coral Triangle than elsewhere (Briggs 1999, Mora et al. 2003); Centre of Accumulation, meaning the Coral Triangle accumulates species that have arisen elsewhere (Ladd 1960, Jokiel and Martinelli 1992); and the Centre of Overlap, meaning the Coral Triangle is the overlap point of Indian and Pacific Ocean species (McMillan and Palumbi 1995, Randall 1998). It has been proposed that patterns in genetic diversity should also mimic species-level diversity gradients (Vellend and Geber

2005). In the context of the Indo-Pacific, it has been further suggested that phylogeographic patterns and population genetic studies should help disclose the prevalence of one hypothesis over another (Drew and Barber 2009). However, there is little consensus in the genetic diversity patterns and phylogeographic histories in this region (reviewed by Carpenter et al. 2010) and it is likely that each of the hypothesized processes have contributed to the high biodiversity of the IAA (Bellwood et al. 2012, Bowen et al. 2013).

Several suture zones delineate species boundaries within the IAA according to historical continental affiliations, and are thought to affect intraspecific lineages in similar ways (as suggested by Avise 1992, Rocha et al. 2007; and found by DeBoer et al. 2014). Most of the population genetic structuring in this region however, has been attributed to the climate oscillations of the Pleistocene (~2.5mya - 12kya, reviewed in Bellwood et al. 2012) as the shallow seas in the IAA exposed parts of the Sunda and Sahul Shelves (Voris 2000). There were two intermittent passages between the Pacific and Indian Oceans for the movement of marine biota during the Pleistocene: the Torres Strait and the approximate path of the modern Indonesian Through Flow (ITF). Marine connection via the ITF was likely more persistent than the Torres Strait throughout the Pleistocene, due to the deep oceanic trench between the Sahul and Sunda Shelves. For more than 100,000 years the Torres Strait was an intermittent land bridge connecting northern Australia and Melanesia (Chappell and Shackleton 1986). A consistent oceanic connection across the Torres Strait linking the Coral Sea and the Arafura Sea was re-established after sea levels rose to present levels 7,000ya, flooding the land bridge (Voris 2000, Reeves et al. 2008).

Biophysical modeling of larval dispersal throughout the IAA (Kool et al. 2011, Treml and Halpin 2012, Treml et al. 2012) and the Pacific Ocean (Treml et al. 2008) suggests the importance of contemporary connectivity corridors for populations of benthic marine organisms in the Indo-Pacific. Major oceanic currents likely drive dispersal through the Solomon Islands, along the Great Barrier Reef, through Micronesia, and along the ITF (Treml et al. 2012). Models also predict less intuitive population connections, such as between Indonesia (and Timor-Leste) and the Kimberley Coast of Australia (Treml et al. 2012). Based on biophysically derived larval dispersal alone, genetic diversity is forecast to be higher in the central part of the IAA as propagules from various peripheral regions accumulate there (Kool et al. 2011). By comparison, northern Australia is forecast to have low genetic diversity (Kool et al. 2011). Unfortunately no individual forecast of population connectivity or genetic diversity patterns based on contemporary dispersal patterns throughout the Indian and Pacific Oceans is available. Indeed, the large range sizes of many Indo-Pacific reef-associated species precludes comprehensive studies of population connectivity based on simulations and empirical data.

Marine genetic studies in the Indo-Pacific have generally consisted of either broadscale sampling over both ocean basins (e.g. Horne et al. 2008, Gaither et al. 2009), or finer scale sampling within the IAA or Coral Triangle alone (e.g. Timm and Kochzius 2008, Barber et al. 2002a; reviewed in

Keyse et al. 2014). In the case of broader-scale sampling, there has been little attempt to resolve the specific geographic location of breaks within the IAA, but other important patterns have emerged. For example, Drew et al. (Drew et al. 2008) examined the phylogeographical relationships across the range of five fishes, extending from the Coral Triangle to Fiji, in the South West Pacific. The Fijian populations were all genetically distinct from those in the Coral Triangle. Furthermore, Drew and Barber (2009) revealed a westward cladogenesis of a common reef fish (*Pomacentrus moluccensis*) originating in Fiji and passing through the Coral Triangle into the Indian Ocean. These studies and others, suggest the Western Pacific may also be a "Centre of Origin" or diversification for some species (Frey and Vermeij 2008, Williams and Duda 2008, Bowen et al. 2013).

Although there already exist many phylogeographic studies throughout the Indo-Pacific, the IAA and wider Indian and Pacific Oceans are seldom considered within a single study. It is likely that population connectivity of benthic marine organisms in the Indo-Pacific has been impacted by historical events within the IAA, patterns of contemporary larval dispersal, and geography of the seascape. Furthermore, it is likely that certain biological traits of species have had a role in determining which of these physical factors have been most important in shaping their past and present population connectivity. Robust answers to questions relating to historical and contemporary population connectivity, and the roles of biological and physical factors in determining population connectivity, ultimately require systematic sampling across multiple species from the same locations.

Structure of this thesis

This thesis contributes empirical case studies and methods of study design and analysis for the field of seascape genetics and our understanding of population connectivity in Indo-Pacific coral reef organisms. Following this chapter, the thesis consists of two review chapters (Chapters Two and Three), three empirical data chapters (Chapters Four, Five, and Six), and a concluding chapter (Chapter Seven). Chapters Two - Six have been prepared as papers for publication; hence there is some repetition of the conceptual framework and methodological detail through the thesis. Each of these papers have co-authors, thus, I use "we" throughout in order to acknowledge their input (but see 'Declaration by Author', page IV). Chapters Two, Three, and Six are published, and Chapters Four and Five are in preparation for peer review. Several published papers I have co-authored and that relate to this thesis are included as appendices to the main thesis, and are referenced as such. An overview for each of the core thesis chapters is provided below.

Chapter Two reviews notable papers in the emerging field of seascape genetics, highlighting general themes and biological traits of species and seascape features that affect spatial genetic patterns in marine systems. Similarities to, and differences from, (terrestrial) landscape genetics are discussed, and future directions for the discipline of seascape genetics are recommended.

Chapter Three has been published alongside Chapter Two. This chapter reviews the study design considerations and the methodologies commonly used in the disciplines of phylogeography, population genetics, and landscape genetics that are applicable to seascape genetics. This review caters to early career researchers or scientists that may be unfamiliar with marine systems, spatial genetic studies, and/or biophysical models. We emphasize the utility of certain methods in the context of marine systems, and in particular the applicability of biophysical models.

The empirical data chapters of this thesis (Chapters Four, Five, and Six) build on previous studies in the field of seascape genetics (described in Chapter Two) and use several of the recommended methodologies from Chapter Three including those related to study design, explicit hypothesis testing, integrating biophysical models with genetic data, quantitative and multivariate analyses, and the cross-disciplinary application of relevant concepts and measures. The aims of the empirical data chapters were to test spatial genetic hypotheses in marine systems across various seascapes and species using purpose-designed and innovative methods.

In **Chapter Four** four coral reef fishes (*Acanthurus triostegus*, *Halichoeres hortulanus*, *Pomacentrus coeletis*, and *Dascyllus trimaculatus*) are co-sampled across a common seascape to observe comparative spatial genetic patterns and to identify the underlying historical and contemporary seascape features and processes responsible for the patterns. We explicitly test the hypotheses that species of similar dispersal potential (based on egg type and PLD) will have similar spatial genetic patterns, and that the effect of historical seascape features will be more evident in the spatial genetic patterns of low dispersal species.

Chapter Five uses extensive sampling of one species (*P. coelestis*) to investigate how spatial genetic patterns vary across the range of a species. This study tests the applicability of the core-periphery hypothesis to a tropical marine species. We extend the basic expectations of the core-periphery hypothesis in two ways. First, we include considerations of population history that may be particularly relevant to tropical marine organisms that have cryptic admixture of lineages at the junction of the Indian and Pacific Oceans. Second, we use the latitudinal extremes (replicated in the northern and southern hemispheres) of this coral reef fish to test the generality of peripheral patterns in tropical marine organisms. We suggest two likely demographic scenarios for the range periphery of a tropical marine species and describe how these would be reflected using conventional genetic measures and alternative measures more commonly used in community ecology.

With the intention of understanding the generality of peripheral genetic patterns in certain seascapes, **Chapter Six** focuses on one peripheral location in the range of two coral reef species (*Acanthaster planci* and *Tripneustes gratilla*). This study investigates whether the populations of the sub-tropical Kermadec Islands in the South Pacific are likely self-sustaining via local retention or are sustained by immigration from the core of the species range.

Chapter Seven draws together the major findings of the chapters and summarizes the contributions of this thesis to our understanding of population connectivity in marine systems and the field of seascape genetics. Finally, avenues for future research are proposed.

CHAPTER TWO. Seascape genetics: populations, individuals, and genes marooned and adrift

Published in Geography Compass 7: 197–216. 2013

Cynthia Riginos and **Libby Liggins**

School of Biological Sciences, The University of Queensland

Abstract

Seascape genetics is the study of how spatially variable structural and environmental features influence genetic patterns of marine organisms. Seascape genetics is conceptually linked to landscape genetics and this likeness frequently allows investigators to use similar theoretical and analytical methods for both seascape genetics and landscape genetics. But, the physical and environmental attributes of the ocean and biological attributes of organisms that live in the sea, especially the large spatial scales of seascape features and the high dispersal ability of many marine organisms, differ from those of terrestrial organisms that have typified landscape genetic studies. This paper reviews notable papers in the emerging field of seascape genetics, highlighting pervasive themes and biological attributes of species and seascape features that affect spatial genetic patterns in the sea. Similarities to, and differences from, (terrestrial) landscape genetics are discussed, and future directions are recommended.

Landscape and seascape genetics – genetics in spatially heterogeneous environments

The question of how spatially arrayed environmental and habitat features influence microevolutionary processes has a long history in population genetics (Epling and Dobzhansky 1942, Wright 1943). Recently there have been calls to explicitly integrate spatial ecological information with population genetic data, in an endeavor coined "landscape genetics" (Manel et al. 2003) where associations between specific landscape features and genetic variation can be statistically evaluated (Storfer et al. 2007). Thus, landscape genetics implicitly melds ecological and evolutionary outlooks in seeking to understand how spatial factors influence genetic changes over both space and time.

Although some authors have emphasized recent events as the primary focus of landscape genetics (Manel et al. 2003), we advocate a more flexible perspective whereby the relevant temporal scale will be determined by the spatial factor(s) of interest, the temporal stability of those spatial factors, and the dispersal ability of the organism(s). Different methods, of course, will be more suitable for some time scales than others (Balkenhol et al. 2009, Anderson et al. 2010, Bohonak and Vandergast

2011). Similarly, landscape genetics can encompass questions across a hierarchy of biological organization from specific genes or loci, to individuals, and populations. To date, most terrestrial and marine landscape genetic studies have focused on selectively neutral processes such as the identification of barriers or resistance to gene exchange, although spatial sources of selection fall within the scope of landscape genetics as well (Manel et al. 2003, Holderegger and Wagner 2008, Manel et al. 2010, Storfer et al. 2010).

In the last 10 years, there have been a profusion of papers in landscape genetics, with several marine examples (albeit in a much smaller proportion than for terrestrial studies: Storfer et al. 2010) and recognition that spatial processes could and should be investigated in marine "seascapes" (Galindo et al. 2006, Selkoe et al. 2008). Reviews have focused primarily on terrestrial examples (Manel et al. 2003, Storfer et al. 2007, Holderegger and Wagner 2008), with the exception of Selkoe et al. (2008) and a brief overview by (Hansen and Hemmer-Hansen 2007) both of which emphasized the importance of oceanographic currents as a primary distinguishing feature between seascapes and (terrestrial) landscapes. Since those reviews, there have been an increasing number of studies that self-identify with "seascape" or "marine landscape" genetics; a search of the Web of Science (Dec. 2011) for the key words "seascape genetics" and "marine landscape genetics" yielded 48 studies (after removing non-marine and non-genetic hits), 37 of which were from 2008 onwards.

Thus, this review updates a quickly moving field from the earlier reviews. In addition, our definition of seascape genetics differs in emphasis. Whereas we agree that ocean currents are an important component of marine landscapes (as articulated by Galindo et al. 2006, Selkoe et al. 2008), we also contend that the relative influence of other spatial factors on marine genetic variation is little known and worthy of integrated investigation. Furthermore, full consideration of spatial structuring factors depends on the spatial and temporal scales of seascape features as well as the life history of the specific organism. For many marine organisms, their extensive dispersal ability may necessitate seascape genetic studies to encompass large geographic areas, yet some seascape attributes may also vary over short time periods. This juxtaposition of spatial and temporal scales creates challenges for the field. Although the terms "seascape genetics" and "marine landscape genetics" are relatively new, several older papers address similar issues, therefore, we attempt to refer both to new developments and classic studies in this field. In the companion paper to this review (Liggins et al. 2013) – Chapter Three) we discuss aspects of study design and genetic analysis relevant to seascape genetics, with an expanded description of biophysical models that are increasingly complementing genetic surveys.

A fluid lifestyle

Marine organisms live in a dense and viscous moving fluid, which transports nutrients, food, gametes, and/or individuals depending on species life history (Thorson 1971, Levinton and Haefner 2002) (Carr et al. 2003, Dawson and Hamner 2008). Marine organisms have a wide diversity of life histories,

and frequently different life stages make use of distinctly different environments. At the extremes, there are organisms with entirely benthic lifestyles and direct development of their young (including but not limited to seahorses, some gastropods, some echinoderms), whereas other organisms are entirely planktonic (many diatoms, dinoflagellates, copepods, krill) or entirely pelagic (including cetaceans and many fishes). Many life histories include both benthic and pelagic stages. For instance, seaweeds and kelp have benthic and floating stages: as adults they grow attached to the benthic substrate, but frequently fragment and floating pieces can drift large distances before attaching elsewhere or contributing gametes to other populations. Most animals which are benthic-associated as adults also have planktonic (pelagic) larvae, and these animals vary widely in their planktonic larval duration (PLD), ranging from minutes to months (Shanks 2009).

Complementing the wide variety of life histories and dispersal abilities of marine organisms, oceans and seas are spatially heterogeneous with respect to many important environmental variables (temperature, nutrients; Fig. 1), but these environmental conditions may change rapidly in relationship to relatively static features such as coastline configuration. Thus, for pelagic animals, static seascape features may appear to move relative to their frame of reference; they may track variable environmental features, for example, keeping themselves within a distinct water mass. In contrast, for relatively sedentary benthic organisms, their frame of reference will be the static benthos, but the environmental conditions in which they are immersed may shift rapidly due to water movements (i.e. considerable movement on an environmental axis, but no movement on a geographic axis).

Environmental factors and life history traits can interact synergistically to influence genetic patterns. For animals that are benthic as adults yet pelagic as larvae, the larval experience can potentially affect the adult population structure. For instance, PLD may be spatially variable within a species, because higher ambient temperatures are expected to raise metabolic rates and thus reduce the duration of the larval stage (O'Connor et al. 2007) perhaps affecting genetic structuring among locations (David et al. 2010). Similarly, the open water environment experienced by larvae can influence their survival (Shima and Swearer 2010) and even juvenile post-settlement survival (reviewed by Marshall and Morgan 2011), thus potentially affecting the genetic patterns formed across the seascape.

There has been much speculation regarding the consequences of high vagility, particularly planktonic larval movements, on genetic patterns of adult populations. For the many marine species that disperse extensively, population genetic theory predicts high genetic variability and low differentiation across large spatial scales (Waples 1998, Hellberg et al. 2002, Faurby and Barber 2012). Accordingly, many (but certainly not all) marine animals appear to be characterized by fairly low population genetic structure (Palumbi 1992, Ward et al. 1995, Kinlan and Gaines 2003). Therefore, population genetic structure may be very low and, as a consequence, difficult to detect

empirically (Waples 1998). Thus, geneticists may struggle to interpret weak and possibly unreliable genetic signals in a seascape context.

A related and prevailing hypothesis has been that higher levels of genetic differentiation should be found among populations of animals that have a short PLD, versus those that have a long PLD. Whereas recent analyses of published genetic data sets (invertebrate and fishes) have found PLD to be a weak or poor predictor of genetic differentiation (Weersing and Toonen 2009, Riginos et al. 2011, Selkoe and Toonen 2011), these post-hoc analyses ignore the effects of geographic location and evolutionary differences among species that may be correlated with other life history traits that affect genetic structuring (Dawson 2012) and differences in effective population sizes among species that will also affect estimates of genetic differentiation (Faurby and Barber 2012). Indeed some studies of co- distributed taxa have found the expected relationship between PLD and genetic differentiation (Waples 1987, Doherty et al. 1995, Riginos and Nachman 2001). The PLD - genetic differentiation relationship remains an area of active study and debate (see Weersing and Toonen 2009, Riginos et al. 2011, Selkoe and Toonen 2011, Dawson 2012 for more extensive discussions). A firm conclusion that emerges however, is that species with direct development (PLD = 0) have high genetic structure (Weersing and Toonen 2009, Kelly and Palumbi 2010).

Another frequent observation that may arise from planktonic larval dispersal and temporally variable environments is that adult genetic patterns can fluctuate without obvious regard to space, in a phenomenon termed chaotic genetic patchiness (Johnson and Black 1982, and see Selkoe et al. 2008 for a comprehensive discussion). A number of explanations have been proposed to explain this phenomenon, including: high variability in reproductive success leading to differing parents contributing to distinct sets of larvae (Hedgecock 1994, Planes and Lenfant 2002) that may be transported in groups by temporally variable currents, selection on larvae during their pelagic stage (Johnson and Black 1982) and selection on juveniles after settlement to the benthos (Vigliola et al. 2007). Note that these competing explanations involve factors that are spatially variable, yet potentially predictable (see also discussion in Selkoe et al. 2008). Indeed Selkoe et al. (2006) showed that the location of a frontal boundary current for the two weeks prior to sampling influenced the genetic composition of kelp bass recruits. Hence, for kelp bass, an understanding of the temporally variable seascape can help explain genetic patterns.

In summary, the marine realm encompasses spatial elements whose temporal stabilities differ widely (Fig. 1), and marine organisms span a wide array of life histories and lifestyles (Carr et al. 2003). In many instances there are parallels between marine and terrestrial environments (Dawson and Hamner 2008), and some seascape genetic studies are similar in approach to those applied to terrestrial ecosystems, particularly when investigating effects of relatively static features on benthic organisms. Yet when temporarily variable features or highly dispersive organisms are the foci of study, then the application of methods developed from terrestrial landscape genetics may be

inappropriate or require modification. The combined considerations of spatio-temporal instability and high dispersal are motivating creative approaches for studying spatial marine genetic patterns that are distinct from terrestrial landscape genetics.

Empirical investigations of seascape features

Seascape genetic studies have predominantly focused on how physical factors facilitate or restrict connections between populations or individuals, using genetic response variables such as F-statistics or estimates of gene flow. *F*-statistics (F_{ST} and related measures), estimate the proportion of genetic variation partitioned among populations relative to the total variation among all individuals and populations. Few studies have considered effects of seascape attributes on genetic variability (see Table 1 for some exceptions). Seascape factors that are likely to influence the genetic structure of marine species differ widely in their grain size and also in their permanence (Fig. 1). Both the relevant spatial and temporal scales for seascape variables should guide study design (see Liggins et al. 2013 – Chapter Three), and in general, studies emphasizing large spatial scales and relatively static landscape features have primarily used population-level sampling and associated methods, whereas studies interested in smaller spatial scales and temporally unstable variables have been more likely to employ individual- focused methods.

In the following subsections, we first describe studies that have investigated relatively stable factors such as geographic distance and topography (stable for thousands of years or more) followed by those that examine unstable factors, especially currents (varying by years, months, days, or hours), including attempts to summarize temporally variable factors as geo-referenced seascape attributes. Obviously these examples represent the tails of spatial and temporal continua (Fig. 1); designing studies that can estimate the effects arising from spatial and temporal processes that vary over many orders of magnitude is a challenge (see Chapter Three for further discussion). In addition, the scale of investigation may determine the relevant stability; for example, major oceanographic currents and their boundaries may be present in the same location on the scale of 100-1,000's of kilometers but at a finer sampling scale (10-100s km) their direction and magnitude of flow may shift over months and days.

The majority of seascape genetic studies to date have chosen to use genetic markers that are commonly assumed to be neutrally evolving (not subject to strong selection) such as microsatellites or mitochondrial (mt)DNA sequences. Because strong selection can alter allele frequencies for selected loci relatively rapidly (at least rapidly in an evolutionary sense), historical patterns (reflecting genetic drift and migration among populations) can be obscured. Thus, the ideal genetic markers for inferring demographic processes, such as isolation or migration among populations and changes in population size, are those that are neutral. For those investigators interested in non-neutral evolution, marine animals also are good study systems for investigating the effect of environmentally-mediated selection

Table 1. Seascape genetic studies that have quantitatively tested for the effects of multiple factors

Seascape factors examined	Significant seascape factors[a]	Habitat configuration	Response variable(s)[b]	Type(s) of analysis	Loci	Common name	Scientific name	Reference
Only including relatively stable factors								
Habitat continuity, Euclidean distance	Habitat continuity, Euclidean distance	linear coastline	F_{ST}	Multiple regression	Msat	Giant kelp	*Macrocystis pyrifera*	(Alberto et al. 2011)
Salinity, temperature (at spawning), Euclidean distance	salinity, temperature	enclosed basin (Baltic Sea)	F_{ST}	partial Mantel tests	Msat	herring	*Clupea harengus*	(Jorgensen et al. 2005)
Salinity, sand	salinity, sand	enclosed lagoon	first principal component (from PCA) based on genetic distance	linear models (GAM)	Pgi allozyme	cockle	*Cerastoderma galucum*	(González-Wangüemert et al. 2009)
Stormwater, wastewater, Euclidean distance	stormwater, wastewater	linear coastline	F_{ST}, D_{est}	single variate regressions, multivariate regressions, canonical redundancy analysis	mtDNA, Msat	bat star	*Patiria miniata*	(Puritz and Toonen 2011)
Sandy coastline, open water, biogeography, Euclidean distance	sandy coastline, open water, biogeography, Euclidean distance	semi enclosed basin (Gulf of California)	D	partial Mantel tests	mtDNA	Cortez triplefin	*Axoclinus nigircaudus*	(Riginos and Nachman 2001)

Table 1. continued

Including unstable factors								
Habitat continuity, minimal oceanographic transport time	habitat continuity, minimum oceanographic transport time	linear coastline	F_{ST}, D_{est}, assignment tests	multiple regression	Msat	giant kelp	*Macrocystis pyrifera*	(Alberto et al. 2011)
SST[c] (mean and StDev), coastline topography, habitat area	SST StDev	linear coastline with two distant oceanic locations	Local genotypic autocorellation	linear models	Msat	barrens forming urchin	*Centrostephanus rodgersii*	(Banks et al. 2007)
SST[c] (mean and variance), coastline topography, area of recent range expansion	SST StDev, and coastline topography, area of recent range expansion	linear coastline including recent range expansion zone, with two distant oceanic locations	AR, H, F_{ST}, D, genotypic autocorellation	partial Mantel tests, GESTE, modeling	Msat	barrens forming urchin	*Centrostephanus rodgersii*	(Banks et al. 2010)
Distance by habitat (depth), distance by current, Euclidean distance	distance by habitat (depth)	continental shelf, offshore banks	F_{ST}	partial Mantel tests	Msat	tusk fish	*Brosme brosme*	(Knutsen et al. 2009)
Chlorophyll, turbidity, SST[c], Euclidean distance	Chlorophyll, turbidity, SST	linear coastline	F_{ST}	Multiple Mantel tests, partial Mantel tests	mtDNA, Msat	Franciscana dolphin	*Pontoporia blainvillei*	(Mendez et al. 2010)
Kelp cover, SST[c], flow centrality[d]	kelp cover, SST, flow centrality	linear coastline with nearshore islands	AR, H, local F_{ST}, G_{ST}'	multiple regression (modified linear mixed model)	Msat	Kellet's whelk	*Kelletia kelletii*	(Selkoe et al. 2010)
Kelp cover, SST[c], flow centrality[d]	kelp cover, SST, flow centrality	linear coastline with nearshore islands	AR, H, local F_{ST}, G_{ST}'	multiple regression (modified linear mixed model)	Msat	spiny lobster	*Panulirus interruptus*	(Selkoe et al. 2010)

Table 1. continued

Kelp cover, SST[c], flow centrality[d]	kelp cover, SST, flow centrality	linear coastline with nearshore islands	AR, *H*, local F_{ST}, G_{ST}'	multiple regression (modified linear mixed model)	Msat	kelp bass	*Paralabrax clathratus*	(Selkoe et al. 2010)
Oceanographic distance, Euclidean distance	Oceanographic distance	linear coastline with nearshore islands	F_{ST}, D_{est}	Mantel tests	Msat	Kellet's whelk	*Kelletia kelletii*	(White et al. 2010)

[a] Significant in at least some analyses
[b] AR, *H*, and *D*, represent Allelic richness, Heterozyosity, and genetic Distance; see original paper for specific methods of calculation and transformations. D_{est} and G_{ST}' are measures of differentiation that correct for allelic variability.
[c] SST = Sea Surface Temperature
[d] Flow centrality reflects the hub position in a network based on biophysical predicted connections.

on genetic variation (Schmidt and Rand 2001, Nielsen et al. 2009a), whereby the biological units of interest are not individuals or populations, but rather are the genetic loci themselves. We also review some of these gene focused studies in a following section.

Stable configurations – distance and topography

Just as on land, geographic distance, barriers, and habitat isolation in the sea can reduce genetic exchanges and, through the action of genetic drift and sometimes selection, lead to divergence between populations and individuals. If, i) populations are evenly spaced, as in a lattice formation, ii) gene exchange is inversely proportional to geographic distance between populations or individuals, that is, migration occurs in a stepping-stone manner, and iii) sufficient time has elapsed for an equilibrium between migration and genetic drift to be established, then an isolation-by-distance (IBD) pattern is expected (Slatkin 1993, Rousset 2000) and often found for marine species (Selkoe and Toonen 2011). Although IBD analyses traditionally use straight-line (Euclidean) geographic distances, other ecologically relevant distance measures, such as shortest over-water distance, shortest habitat path, and oceanographic distance (derived from an ocean circulation model that may or may not be parameterized with biological attributes of the organism, i.e. a biophysical model) may capture physical and/or biological processes affecting gene exchange and, therefore, be better predictors of genetic differentiation than Euclidean geographic distance. Thus, in seascape studies, Euclidean distance can form a simple model against which more sophisticated models can be compared, or Euclidean distance can be included as a covariate along with other predictive variables so that the independent contributions of each variable are estimated (see Table 1 for examples).

Distance in and of itself can modify gene exchange, but other types of barriers or habitat discontinuities can also obstruct movement. Obviously landmasses form barriers and, depending on their configuration, landmasses may be complete barriers to movement. More complex cases are permeable barriers that reduce but might not preclude gene exchange. For coastal or reef species, large stretches of open water can contribute to greater divergence relative to comparisons along coastlines (Doherty et al. 1995, Ayre and Hughes 2004). Features along coastlines can also influence genetic patterns, as found for the large freshwater outflow of the Amazon River on some reef dwelling surgeonfishes (Rocha et al. 2002) upwelling locations for rockfish (Johansson et al. 2008) large stretches of sand for some rocky reef fishes (Riginos and Nachman 2001, Johansson et al. 2008) and invertebrates (Ayre et al. 2009) and even hotspots of wastewater and stormwater pollution for a bat star (Puritz and Toonen 2011). For a fish restricted to deep continental shelves (100-1,000 m), deep trenches (>1,000 m) were similarly found to significantly contribute to genetic differentiation (Knutsen et al. 2009).

In some cases, the effect of partial barriers is modulated by the size of that barrier. For example, for the rock-associated Cortez triplefin, the IBD slope (Euclidean geographic distance vs. F_{ST}) for populations separated by sand was steeper than the slope for populations separated by rock reef

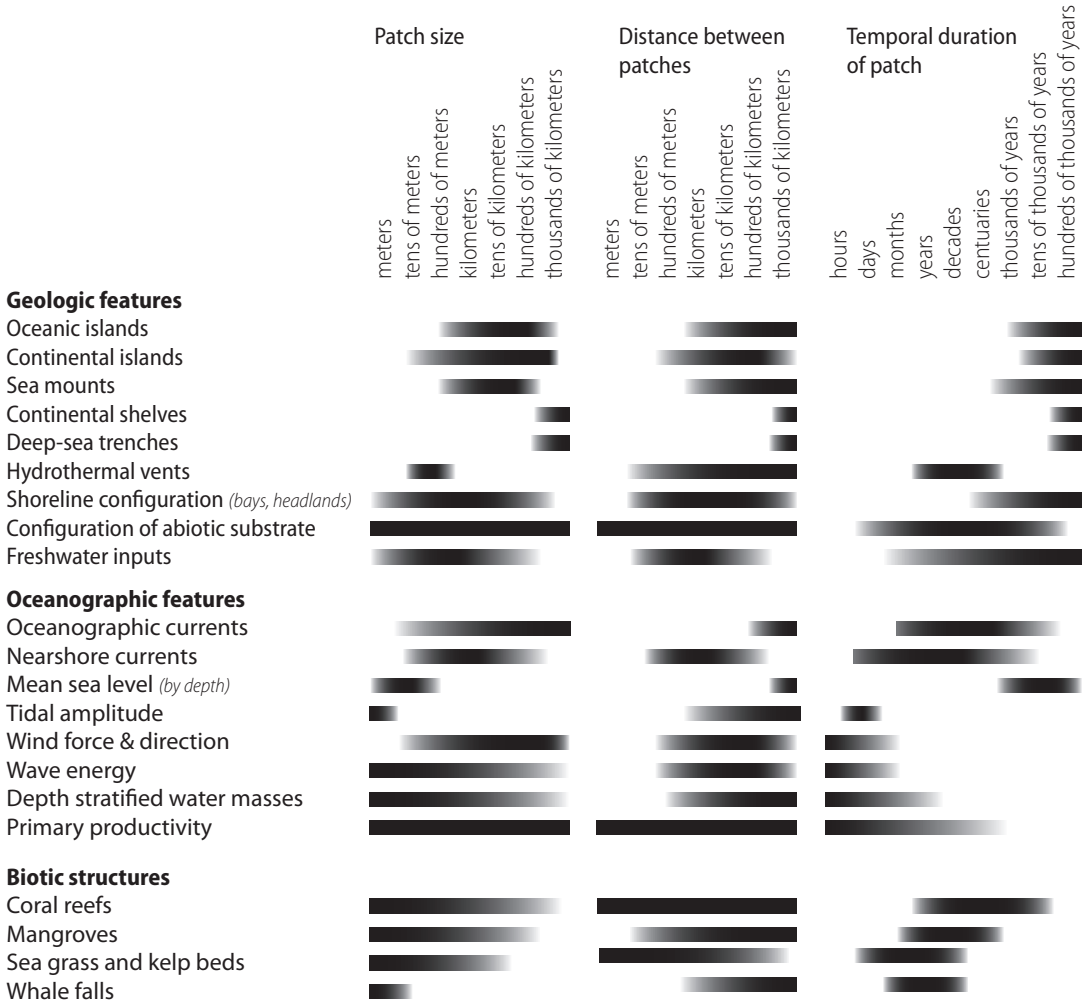


Figure 1. Spatial and temporal variability of seascape features that may affect genetic differentiation. Patch size indicates the size of the feature. Patch size and distance between patches are aspects of environmental granularity. The temporal duration of a patch indicates the stability or permanence of a feature in a particular configuration. The intensity of shading is intended to convey the relative frequency of a particular size, distance, or duration. Relative frequencies were informed by soliciting expert opinions from geologists, oceanographers, evolutionary biologists, and ecologists.

(Riginos and Nachman 2001). Similarly, the degree of habitat discontinuity (estimated using satellite image analysis in a GIS - geographic information system) in combination with geographic distance were better predictors of genetic differentiation between giant kelp populations than either factor alone (Alberto et al. 2011). For golden jellyfish living in land-surrounded marine lakes, the genetic distance between lake and reef lagoon populations was correlated with geographic distance between each lake and its nearest lagoon (Dawson 2005). The effect of distance can also vary with spatial scale; for example, IBD was found within Pacific archipelagos, but no linear relationship with distance was found among archipelagos for a surgeonfish (Planes et al. 1996). Conversely, for the crown-of-thorns starfish, IBD was found among Pacific locations (spanning large stretches of open ocean) but not within the more continuous reefs of Japan and the Philippines (Yasuda et al. 2009). The potential for interactions between distance and other spatial variables highlights the utility of considering multivariate approaches (Table 1) that can evaluate the relative importance of differing variables and their interactions.

Unstable configurations – currents

Advances in remote sensing and oceanographic modeling have greatly enhanced the resolution with which currents, particularly nearshore currents relevant to coastal larvae, can be inferred in both space and time. Knowledge of water movements can be combined with species-specific biology to create complex biophysical models that yield spatially explicit dispersal probabilities (see Chapter Three). Earlier examinations of major (offshore) oceanographic currents and genetic patterns found modest correlation (Palumbi et al. 1997) or highlighted disparities (Benzie 1999) whereas some recent investigations, especially using more precise biophysical models to predict genetic patterns, have been compelling and are discussed below.

Most studies with oceanographically-informed predictions attempt to correlate the simulated movement of individuals or water masses against empirical genetic data, often using qualitative comparisons. For example, Cowen et al.'s (2006) biophysical model of larval dispersal for reef fish defined four largely distinct regions of population connectivity within the Caribbean that qualitatively matched genetic patterns previously described for a goby (Taylor and Hellberg 2003) and a coral (Baums et al. 2005). Similarly, Baums et al. (2006) found that simulated coral larvae could not disperse across an ocean passage between the eastern and western Caribbean, which coincided with a genetic break revealed using empirical genetic data.

Some investigators have taken the connectivity matrices produced by biophysical models and used them to project the development of population genetic structure forward through time, allowing simulated genetic patterns and empirical genetic patterns to be compared directly. The first application of this method found that simulated genetic patterns for a broadcast spawning coral, *Acropora cervicornis*, in the Caribbean were in agreement with the broadscale observed genetic patterns (Galindo et al. 2006). Similarly, Kool and co-workers found that biophysically-predicted

genetic structure in coral reef populations in the Caribbean (Kool et al. 2010) and coral reef communities in the Indo-West Pacific (Kool et al. 2011) matched observations of genetic structure qualitatively assessed across many species and specifically for the coral *Montastraea annularis* (Foster et al. 2012). Discordances between modeled data (either coupled or uncoupled to genetic simulations) and observed genetic data can highlight regions where dispersal alone does not capture connectivity (such as when post-settlement processes are important) and shortcomings in the model (Galindo et al. 2010, Foster et al. 2012), but discordances also may reflect the inherently differing time scales between modeled biophysical data and empirical genetic patterns of real organisms.

Other studies have sought to quantify the relationships between different oceanographic predictors and empirical data. For instance, connections based on oceanographic models were found to better predict F_{ST} values between population pairs as compared to Euclidean distances in linear regression analyses for a whelk (White et al. 2010) and a giant kelp (Alberto et al. 2011). For the whelk, an oceanographic distance, indicating the likely distance that larvae would travel, was the better predictor of pairwise F_{ST} values (White et al. 2010). In the case of the giant kelp, the best explanatory model was one that combined minimum transport time and habitat continuity (Alberto et al. 2011). Rather than inferring genetic metrics independently for each set of population pairs, Crandall et al. (2012) derived coalescent estimates of gene flow simultaneously among all sampled populations and demonstrated that biophysically-based stepping-stone models of gene flow were a better fit to the data than a series of other migration models. (Coalescent analyses are based on computationally intensive simulations whereby the observed data are compared against the parameter space that could produce such data; see Hey and Machado 2003 for a general review).

A key aspect of predictions derived from oceanographic measurements is that if currents are important for transporting propagules and influencing genetic patterns, then genetic exchange should be asymmetric (Wares et al. 2001), and genetic methods that can detect asymmetric migration (Marko and Hart 2011) should outperform genetic response variables that provide a single bidirectional value between populations such as F_{ST}. For instance, Wares et al. (2001) developed a test based on genealogical patterns of mtDNA haplotypes and demonstrated statistically significant north to south dispersal for two barnacle species and an urchin, matching the prevailing southward current. Similarly, dispersal inferred from assignment tests using nuclear microsatellites was in the same direction that sea beet seeds were expected to float (Fievet et al. 2007) and coalescent estimates of directional gene flow for Antarctic icefishes (Matschiner et al. 2009, Papetti et al. 2012) and neritid snails (Crandall et al. 2012) matched the Antarctic circumpolar current flow and dispersal predicted by biophysical models, respectively. Other studies, however, have not found matches between the directionality of currents and directionality of asymmetrical genetic estimates of migration (e.g. kelp, *Laminaria digitata*, Billot et al. 2003). Furthermore, for the giant kelp *Macrocystis pyrifera*, oceanographic models had a closer fit to symmetric measures of F_{ST} than to assignment tests that have directionality (Alberto et al. 2011). Oceanographic predictions of movement are averaged over time,

typically over many years; if dispersal is highly variable in the short-term, as appears likely, short time windows assessed by assignment tests that use patterns of linkage disequilibria to identify migrants from the present or previous 1-2 generations may be mismatched with time-averaged predictions. Aside from assignment tests, most genetic statistics are based on allele frequency differences among populations that arise over evolutionary time frames (tens to hundreds of thousands of years or more). Thus, it is equally surprising that biophysical models based on present-day models of currents (rather than Pleistocene averages) and empirical genetic patterns are concordant in many instances.

Spatially explicit summaries of temporally variable conditions

Because the transitory nature of currents precludes associating single values to specific geo-referenced locations, several studies have taken summary values, such as mean sea surface temperature (SST), mean or maximum current speed and direction, chlorophyll, and turbidity (Mendez et al. 2010, White et al. 2010, Alberto et al. 2011, Knutsen et al. 2009). For example, Mendez et al. (2010) examined the predictive power of regional (time averaged) differences in SST, chlorophyll, and turbidity for Franciscana dolphins and found these environmental differences to be better predictors than Euclidean geographic distances for microsatellites, but not mtDNA.

Others have summarized variability as a landscape attribute. For example, in a study of the barrens-forming urchin, Banks and co-workers estimated the "protectedness" of embayments, as measured by the amount of coastline within 20 km of a site, and found that within protected sites there was greater genotypic autocorrelation (similarity) among individuals (Banks et al. 2007) and greater differentiation between sites (Banks et al. 2010), consistent with larvae being retained in protected bays. Banks et al. (2007) also showed that genotypic autocorrelation among local individuals was positively associated with variability in SST. SST variance should be correlated with variance in currents because the region in question is one where eddies are periodically shed from a major oceanic current; thus, the positive association between SST variance and within population genetic autocorrelation was consistent with fewer cohorts recruiting to locations with highly variable currents (Banks et al. 2007). In another example, Selkoe et al. (2010) took biophysically-derived migration probabilities and rated populations of whelks, lobsters, and kelp bass based on the centrality of their position relative to an inferred network of oceanographic connections, with the prediction that central populations would be more diverse and less genetically differentiated than peripheral populations. However, support for this hypothesis was modest and mixed across species, with the best evidence found in kelp bass.

Although currents have received the greatest consideration, other spatial attributes are also dynamic. For instance, in the deep sea, hydrothermal vent communities are only stable for years to decades, with vent longevity reduced by high rates of tectonic spreading. Coykendall et al. (2011) found that genetic diversity of rift tubeworms was inversely correlated with tectonic spreading rates, as would be expected if young vents have been more recently settled or received fewer colonizing

settlers over the life of the local vent than older vent tubeworm populations. These studies of urchins, whelks, lobsters, kelp bass, and rift tubeworms illustrate innovative ways in which temporally dynamic properties of the seascape can be summarized as predictive variables.

Comparisons among species

Comparisons among species can reveal the relative importance of biological traits on genetic patterns and also identify spatial commonalities among species. Yet, comparative landscape genetic studies are rare (Storfer et al. 2010), presumably because the difficulty of working with a single species is compounded with the inclusion of multiple species. Species differing in dispersal-related traits have been evaluated in a landscape context only in a few studies. For instance, in a comparison of corals across the same reef systems, the neighborhood distance (reflecting dispersal) was greater for the broadcast spawning *Acropora tenuous* as compared to brooding *Seriotopora hystrix* (Underwood et al. 2009). For ten species of co-distributed intertidal invertebrates, an extensive (300 km) sandy coastline was found to be a persistent barrier to gene flow for most species with planktonic developing larvae, but not for two of the species with direct developing larvae, with the conclusion that habitat generalists are less affected by this barrier (Ayre et al. 2009). The idea of accounting for shared landscape features can be scaled up to include many species, but including many species may necessitate spatially-implicit rather than spatially-explicit evaluations if species are not co-distributed. For instance, several analyses of published datasets have examined the effect of PLD in an IBD framework (Kinlan and Gaines 2003, Selkoe and Toonen 2011). This type of approach was expanded for reef fishes whereby spatially implicit factors included both biogeographic boundaries and Euclidean over water distance to show that the dispersal-related trait of egg type (benthic vs. pelagic) was an additional contributor to species level F_{ST} (with biogeographic boundaries and over water distance significant as well: Riginos et al. 2011).

Although the aforementioned studies made explicit contrasts among species traits, one could similarly ask whether specific seascape features or locations contributed to common spatial genetic patterns among species. This type of question is well established in the instance of assessing specific features such as biogeographic boundaries (as in Lessios et al. 1998, Ayre et al. 2009), but less so for identifying emergent commonalities across species. An exception is the approach used by Selkoe et al. (2010), where they tested for spatially consistent patterns among three species along the California coast and found correlations between site-specific heterozygosity and between site differentiation (Table 1). In the future, the identification of specific geographic areas that are associated with either greater or less than expected genetic differentiation across multiple species could inspire testable hypotheses regarding the general influence of specific features in marine landscapes.

Adaptive differentiation, seascape genetics, and fitness

The high dispersal lifestyle typified by many marine organisms is frequently cited as a challenge for the evolution of local adaptation (Slatkin 1985, Lenormand 2002), because locally

adapted variants may be swamped by migration and individuals may fail to disperse to an environment that best suits their genetic makeup. Yet, empirical examples of allele frequency shifts in response to environmental gradients are well documented (reviewed by Schmidt et al. 2008) including the classic example of selection induced by salinity on the *Lap* locus in mussels, whereby the *Lap94* allele diminished in frequency during summer months among new settlers in low salinity Long Island Sound (across tens of kilometers, Koehn et al. 1980b). Because the biochemical properties of the Lap94 protein are physiologically disadvantageous in low salinity conditions relative to protein products of other *Lap* alleles (Koehn et al. 1980a, Koehn and Siebenaller 1981), there is a clear functional link between the Lap locus, fitness of individual mussels, and spatial genetic patterns. Not only was selection on *Lap* found to vary spatially with salinity, selection was also seasonally dependent, highlighting the importance of time and seasonality, as well as dispersal, for understanding spatial patterns induced by selection.

Another example of locus-specific selection tied to environmental variables comes from cod and the *Pan I* locus. Patterns of nucleotide variation indicated that selection has affected *Pan I* evolution (Pogson 2001, Pogson and Mesa 2004) and subsequent surveys have revealed geographically complex patterns of *Pan I* polymorphism. In the northeastern Atlantic, mean June SST was a significant predictor of *Pan I* allele frequencies among adults when controlling for Euclidean distance in partial Mantel tests (Case et al. 2005) whereas water depth was a strong predictor of *Pan I* allele frequencies among Icelandic populations (Pampoulie et al. 2006). For juvenile cod within a Norwegian fjord, the strongest predictors of *Pan I* allele frequencies were temperature, salinity, and depth (Case et al. 2005). Genome scans in cod have also identified additional loci that show correlations in their allele frequencies with temperature, salinity, latitude, and longitude among Atlantic locations (Moen et al. 2008, Nielsen et al. 2009b).

As is the case with all spatial genetic studies, correlations between landscape or seascape features and genetic loci (such as *Pan I* and depth, temperature, salinity etc.) identify candidate loci for environmental mediated selection. That is, correlations are identified but causation cannot be inferred. For instance, the surveyed marker may be physically linked to the target of selection or statistical associations among loci could arise as a consequence of demographic history. A convincing case for causation would require demonstrating a mechanistic link (Feder and Mitchell-Olds 2003, Lowry and Willis 2010). For *Lap* in mussels, the biochemical properties of the proteins resulting from different alleles are well resolved providing a functional explanation for the spatial genetic patterns. For *Pan I* in cod, the patterns of nucleotide variation of that locus corroborate selection, but further understanding of the function of the Pan I gene product would be necessary to link selection at this locus to a specific environmental attribute.

On a small spatial scale (tens of meters), several recent studies have highlighted strong genetic discontinuities between adjacent populations segregating by intertidal height (periwinkles:

Johannesson et al. 2010) or by depth (gorgonians: Mokhtar-Jamai et al. 2010 and Prada et al. 2008; bird's nest coral: Bongaerts et al. 2010). In addition to the caveats about correlation mentioned above, even if genotype-by-environment fitness differences can be demonstrated (as in Bongaerts et al. 2011), it will be important to consider whether endogenous reproductive incompatibilities also maintain genetic divergence, with multiple loci segregating with environmental features indicative of such a situation. If the adjacent populations have some degree of inherent reproductive isolation, then for a given locus or linkage group, distinguishing between seascape (depth) induced selection and endogenous selection will be difficult (Bierne et al. 2011). In such situations, a seascape feature such as depth may be detectably contributing to spatial segregation between adjacent ecotypes, but identifying which loci may be specifically under selection due to local adaptation will be extremely challenging.

Despite the many interesting and spatially replicated gradients that occur in the marine environment (see Schmidt et al. 2008), studies investigating selection in a seascape genetic framework are relatively uncommon. Expansion of seascape genetics to a genomic perspective (Nielsen et al. 2009b) will open up avenues of investigation. Increasingly, genomic methods and next generation sequencing are facilitating population genomics of non-model organisms (Luikart et al. 2003, Hohenlohe et al. 2010) and extensions into "landscape genomics" (Joost et al. 2007, Nielsen et al. 2009b, Manel et al. 2010). Population genomic methods could reveal candidate genes for selection associated with seascape features, as with cod (see previous and (Nielsen et al. 2009b) and thus provide hypotheses which could be verified experimentally and functionally.

Future directions: competing seascape factors and analytical challenges

A fundamental challenge for seascape genetics is to develop predictive models that adequately describe both (relatively) static and dynamic seascape features. A consequence of this challenge has been an emphasis on simple landscape attributes such as geographic distance between sites (i.e. spatially implicit features) and less attention to precise geo-referenced spatial data (Alberto et al. 2011). Certainly several recent studies have made use of geo-referenced data and GIS type approaches (Banks et al. 2007, Alberto et al. 2011, Puritz and Toonen 2011), but many approaches familiar to terrestrial landscape ecologists are often less useful to seascape ecologists due to the high temporal variability of many seascape features (currents, SST, salinity; Fig. 1); the dynamic nature of currents and associated seascape features are difficult to summarize in a spatially explicit manner. In particular, concepts of landscape resistance used for least cost path and isolation-by-resistance analyses (McRae 2006) are not easily translatable to flowing aqueous habitats and fail to accommodate the likely asymmetries in migration. Connectivity models assuming symmetric migration can make misleading predictions if migration is asymmetric (Vuilleumier and Possingham 2006). Similarly, when symmetric metrics of population genetics (such as F_{ST}) are applied to situations with asymmetric gene flow,

estimates of gene flow can be highly inaccurate (Wilkinson-Herbots and Ettridge 2004, Marko and Hart 2011). Not surprisingly, then, seascape genetics has seen strong development and usage of biophysical models that incorporate asymmetric movements (Galindo et al. 2006, Treml et al. 2008, Kool et al. 2010, Selkoe et al. 2010, White et al. 2010, Kool et al. 2011, Foster et al. 2012), with published models for the Caribbean, the California coast, and parts of the Pacific and Indian Oceans. This usage and formulation of biophysical models will certainly continue and expand in geographic coverage.

At present biophysical models generate estimates of potential dispersal, whereas realized genetic connectivity depends on larvae reaching their destination and surviving to reproduce. The larval and settlement periods are characterized by high daily mortality (Almany and Webster 2006), some of which undoubtedly results in natural selection. Also, emerging research in marine biology indicates that larval environments influence both planktonic larval dispersal (O'Connor et al. 2007, Shima and Swearer 2010) and juvenile survival (reviewed by Marshall and Morgan 2011). Note that such fitness consequences are not necessarily due to selection on heritable traits (natural selection). Thus, for benthic animals with pelagic larvae, the effects of the open water and post-settlement environments on individuals as well as genotypes and allele frequencies are likely to be far more complex than predicted by physical impediments to larval movement. Better functional understanding of how larval experiences shape lifetime fitness may inspire hypotheses, including those formulated by biophysical models, that account for larval or post-settlement deaths resulting from relevant environmental characteristics. Such hypotheses, informed by larval biology, will be well-suited for testing using a seascape genetics approach. Therefore seascape genetics can contribute to a fuller understanding of which factors influence realized connectivity.

Clearly the challenges of understanding dispersal through water sets seascape genetics apart from terrestrial landscape genetics. Another contrast between the two fields is that terrestrial landscape genetics has embraced individual level sampling and analyses, whereas seascape genetic studies have primarily used population-level sampling (that is, collecting many individuals from each sampling location with typically <15 locations total). The geographic scale over which marine organisms are likely to disperse combined with the spatial scales of seascape features (Fig. 1) make sampling hundreds of individuals evenly (or at random intervals) along thousands of kilometers logistically challenging. We only know of one seascape genetic study that can truly be described as using individual-based sampling over large geographic distances: a survey of harbor porpoises in Europe where sampling was fairly even with respect to the Atlantic coastline (Fontaine et al. 2007). But individual approaches have resolved sources of variation across relatively small scales for low dispersal organisms (as in Underwood et al. 2009, David et al. 2010), and individual-based methods have successfully been applied to studies with population level sampling (Jones et al. 2005, Selkoe et al. 2006, Banks et al. 2007, Buston et al. 2009, Banks et al. 2010, Saenz-Agudelo et al. 2011). Where logistically feasible, seascape genetics as a field could benefit by incorporating more individual-level

approaches, both in sampling and analyses. In particular, a situation in which individual approaches would be highly informative would be locus-specific investigations of selection over steep environmental clines, such as depth, intertidal exposure, or salinity.

An aspiration for all areas of landscape genetics is to move beyond exploratory studies and embrace experimental designs aimed at testing *a priori* hypotheses (Storfer et al. 2010). This necessitates planning the sampling strategy and analyses such that competing factors can be statistically evaluated (Selkoe et al. 2008, Chapter Three). In reviewing studies that were self-identified as seascape or marine landscape genetic, we found many studies that qualitatively evaluated factors (typically "barriers" of one kind or another). Studies that quantitatively evaluated multiple seascape features, however, were rare (and summarized in Table 1). In our opinion, greater utilization of multivariate approaches and model testing (see Balkenhol et al. 2009 for a recent review) would enhance our understanding of which factors influence genetic variation in marine species.

Conclusions

Seascape genetics is a rapidly developing field of inference. Although many advances from (terrestrial) landscape genetics are relevant and should be embraced by marine-focused investigators, the dynamic fluid medium of seas and oceans also necessitates novel approaches and methods of analysis that will continue bringing together investigatory teams with expertise in both genetics and oceanography. Whereas the ability to predict water movements at levels of resolution relevant to the dispersal of populations and individuals is very exciting, additional seascape factors are likely to also impact spatial genetic patterns. Because genetic differentiation and variability depend on survival to reproductive age and not just dispersal, investigations of factors influencing survival (whether selection on specific traits or effects of larval exposures) will complement dissections of dispersal-affecting seascape features. Finally, we encourage investigators using spatial outlooks to design their studies such that competing seascape features can be quantitatively assessed.

Acknowledgements

We thank M Beger, ED Crandall, AJ Richardson, KA Selkoe, SE Swearer, and EA Treml, for comments, especially their opinions regarding spatial and temporal variability of seascape features. The comments of JA Kupfer and two anonymous reviewers also substantively improved the manuscript. LL was supported by an Australian Postgraduate Award from the Australian Government and a Queensland Government Smart Futures PhD Scholarship. Many of the ideas discussed here grew out of work funded by the Australian Research Council (DP0878306).

CHAPTER THREE. Taking the plunge: an introduction to undertaking seascape genetic studies and using biophysical models

Published in Geography Compass 7: 173–196. 2013

Libby Liggins[1], Eric A. Treml[1,2] and Cynthia Riginos[1]

[1]School of Biological Sciences, The University of Queensland

[2]Department of Zoology, University of Melbourne

Abstract

The field of seascape genetics aims to evaluate the effects of environmental features on spatial genetic patterns of marine organisms. Although many methods of genetic analysis and inference appropriate to ''marine landscapes'' derive from terrestrial landscape genetics, aspects of marine living introduce special challenges for assessing spatial genetic variation. For instance, marine organisms are often highly dispersive, so that genetic patterns can be subtle, and the temporal variability of the marine environment makes these patterns difficult to characterize. Tools and techniques from oceanography can help describe the highly connected and dynamic nature of the marine environment. In particular, models incorporating physical oceanography and species attributes in realistic simulations (e.g. biophysical models) can help us understand this complex process and formulate spatially explicit biologically-informed predictions of gene flow. Thus, researchers embarking on a seascape genetic study need a solid understanding of marine organisms and spatial genetics perhaps combined with knowledge of physical oceanography and ecological modeling. Although some researchers may acquire proficiency in all of these areas, seascape genetic studies incorporating biophysical modeling are likely to bring together groups of investigators with complementary expertise. This preliminary guide is intended to be a starting point for a reader new to either seascape genetics or biophysical models.

What's in a bit of water?

Seascape genetics is a sub discipline of ''landscape genetics'' focused on marine habitats and species. Landscape genetics draws upon methods from landscape ecology and geography to characterize spatial factors and uses methods derived from population genetics as metrics of genetic variation (Manel et al. 2003, Storfer et al. 2010). The majority of landscape genetic studies to date have been in terrestrial habitats (Storfer et al. 2010), but there has been increasing interest in applying these concepts to marine organisms and seascapes. Examples of seascape genetic studies include those

interested in the historical colonization and migration patterns among marine populations, contemporary patterns of dispersal among populations, and patterns of neutral and adaptive genetic variation in relation to past or present features of the seascape (see Riginos and Liggins 2013)– Chapter Two for a review of studies). Here, we focus on the methodologies that apply to such investigations.

The most challenging element of the marine environment for traditional spatial genetic techniques is ocean dynamics (Galindo et al. 2006, Selkoe et al. 2008). The fluidity of the ocean and the strength of ocean currents can lead populations to be highly connected by dispersive individuals, and disconnected by ocean barriers, in non-intuitive ways. In addition, the physical template of the ocean, including salinity, light, temperature and currents, fluctuates through time (see Fig. 1, Chapter Two). Furthermore, life histories of marine organisms often differ from terrestrial organisms (Strathmann 1990) and can vary substantially among marine organisms, so that inter-population genetic exchange can occur via gametes, embryos, larvae and/or adults, in any combination. These various modes of exchange are differentially impacted by ocean dynamics; for example, ocean currents are a dispersal vector for some passive species (and in some locations and seasons only) and for other life history strategies (e.g. strong-swimming or benthic) a current may be either irrelevant or act as a barrier.

Many seascape genetic studies have benefited from considering their genetic data alongside physical oceanographic models (e.g. White et al. 2010, Alberto et al. 2011; similar to the utility of wind models in the study of gene flow via wind-dispersed pollen or seeds in land plants, see Levin et al. 2003). However, certain life history features and behaviors of marine organisms are also important determinants of dispersal (Strathmann 1990, Kingsford et al. 2002, Gerlach et al. 2007) prompting the use of biologically-informed models similar to those used for terrestrial animals and already used for some marine mammals (e.g. Austin et al. 2004). Thus, coupled biological-physical models (hereafter biophysical models) incorporating ocean circulation data have emerged as away to help generate seascape genetic hypotheses, produce biophysical data to correlate with genetic data, and examine the mechanisms underlying genetic patterns.

This preliminary guide is intended to supplement the accompanying review ('Seascape genetics: populations, individuals, and genes marooned and adrift', Riginos and Liggins 2013 – Chapter Two) in which distinguishing features of the marine environment are described, and influential seascape genetic studies are discussed. In the present paper, we focus on experimental design and methodologies: we first identify considerations specific to seascape genetics in the context of present-day methods and analytical developments in the field of landscape genetics. In the second part of this paper, we introduce biophysical models, review their use in combination with genetic data, and highlight relevant technical considerations in their use. (Note that important terms are explained in the appended glossary).

Undertaking seascape genetic studies

The central purpose of seascape genetic studies is to infer associations between spatial environmental features and genetic variation that may be neutral (shaped by processes of mutation, genetic drift and gene flow) or adaptive (influenced by natural selection). Although the focal environmental features may differ from those of terrestrial landscape studies, many methods of analysis and inference will be the same. Thus, general reviews of landscape genetics are recommended (e.g. Holderegger and Wagner 2008, Manel et al. 2003, Storfer et al. 2007). Here we review aspects of experimental design of seascape genetic studies aimed at investigating neutral genetic patterns and point readers to key references for further reading.

A priori *identification of relevant seascape features and study design*

The design and field sampling approach of any spatial genetic study is very important, yet these elements are frequently overlooked and/or underestimated (see Storfer et al. 2007; for an extended discussion of approaches in landscape genetics; see Table 1 for a summary of considerations relevant to seascape genetics studies). Which seascape features are sensible to investigate should be informed by knowledge of the organism's biology, ecology and the geography of its range (see Fig. 1 for biological characteristics of marine organisms that may be relevant, some of which are discussed below). In turn, the geographic nature and length of time for which spatial features persist should guide the choice of genetic markers and methods of analysis (Anderson et al. 2010; and see Fig. 1 in Chapter Two for examples of genetically relevant seascape features).

In order to rigorously evaluate the effect of any geographic feature (marine or terrestrial) on genetic structure, sampling should be spatially broad, so that the genetic patterns revealed as a consequence of that feature, can be assessed within the broader 'genetic-background'. However, undertaking an optimal sampling design across a seascape is possibly more challenging than on land, given that many parts of the marine environment are under studied or are inaccessible (to most investigators). The consequence is that comprehensive sampling across a species' entire geographic range tends to be rare. Inferences from marine datasets often must assume there are unsampled populations and lineages, which has implications for their analysis and interpretation (see Slatkin 2004, Schwartz and McKelvey 2008).

When deciding on the arrangement of sampling across a seascape, spatial autocorrelation (Legendre 1993) or the correlation among environmental features simply due to distance (e.g. depth, latitude, longitude, temperature etc.) needs to be accommodated (as is the case for any spatial analysis). While the influence of these distance related patterns can sometimes be analytically partitioned (see partial Mantel tests and multivariate approaches below), these complications would be more easily dealt with within a strategic sampling design (e.g. stratified random; for discussion on the effects of sampling design see Schwartz and McKelvey 2008).

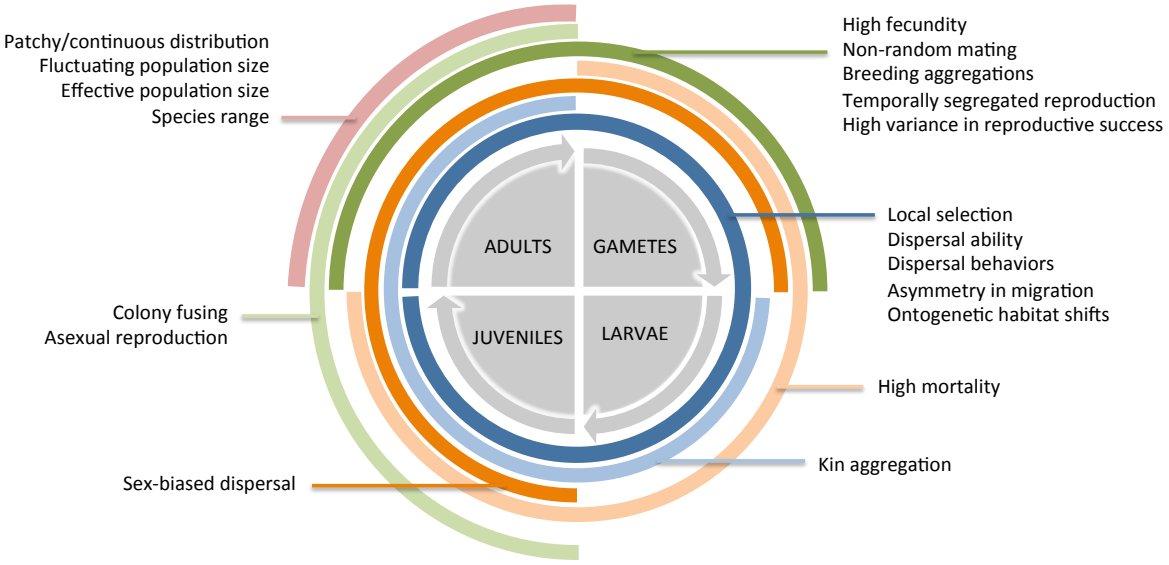


Figure 1. Biological characteristics of marine organisms that should be considered in the design, sampling, and analysis of a seascape genetic study. (Note that many of these characteristics are not exclusive to marine organisms, but are also common in terrestrial organisms). Colored bars depict the life stages at which the biological characteristics are relevant (central grey circle: gametes, larvae, juveniles, adults). Local selection during any life stage (dark blue) can influence the interpretation of genetic patterns and genetic measures that assume marker neutrality (the influence of high dispersal ability, ontogenetic habitat shifts, dispersal behavior and the analytical challenges posed by asymmetry in migration are discussed in the main text). Kin aggregation during the larval, juvenile and/or adult stages (light blue), can compromise sampling design unless samples are taken sufficiently distant from each other. The ability of some colonial organisms to undergo asexual reproduction (i.e. more than one individual with the same genotype) and colony fusion during juvenile and/or adult stages (i.e. one individual with several genotypes; light green) means physically identified "individuals" are not necessarily representative of genetic "individuals" complicating sampling design and violating the Hardy-Weinberg Equilibrium (HWE) model assumption of "random mating". Fluctuating population size (pink) is common in many marine invertebrates and violates the assumption of "constant" and "large population size" in many population genetic models (such as HWE and the island model of migration; the influence of effective population size, species range and patchy or continuous distributions on sampling are discussed in the main text). High mortality during the gametic, larval and/or juvenile life stages (light orange) encourages appropriate targeting of life stages, given the study objectives. Non-random mating during the adult and gametic life stages (dark green) violate the assumptions of population genetic models; high variance in reproductive success can lead to small effective population sizes that violate the HWE assumption of "large population size" and can lead to temporal shifts in the genetic composition of age-classes (discussed in the main text, as is the influence of high fecundity, breeding aggregations and temporally segregated reproduction on sampling efforts and analysis). The influence of sex-biased dispersal during the juvenile, adult and/or gametic phases (dark orange) is discussed in the main text.

Attributes of species biology will influence marker choice and what is considered to be a sufficient sample size. Many marine organisms are likely to have large effective population sizes resulting from high vagility and fecundity (Hellberg 2009). The process of genetic drift is inversely related to effective population size (Wright 1931). The large population size of many marine organisms, thus, contributes to high levels of genetic diversity within populations and low genetic differentiation between populations (Hedrick 1999) a situation in which statistical tests of population structure are prone to both Type I and Type II error (Waples 1998). These characteristics may necessitate sampling more individuals per population (to capture the asymptote of intra-population genetic diversity) and/or the use of more unlinked molecular markers (Waples 1998), permitting analyses based on linkage disequilibria, which are more sensitive to recent isolation between populations than frequency based metrics (see below).

One important caution is that marine or terrestrial studies relying on a single marker (such as mitochondrial DNA, mtDNA) or few markers, must assume that the marker studied is representative of population processes throughout the genome; thus, selective neutrality is implicitly assumed and the large variance in coalescence times among loci is not taken into consideration. In the case of haploid mtDNA (or chloroplast DNA, cpDNA, for algae or sea plants) only properties of female lineages are actually being estimated which may be inappropriate if there is sex-biased dispersal (as in some marine turtles, e.g. Casale et al. 2002). For extensive discussions regarding the appropriate choice of markers, see Manel et al. (2010) and Bohonak and Vandergast (2011).

Focal age-class is important to consider when sampling marine organisms that change their dispersal behavior, or shift their habitat use over their life time, to avoid obtaining misleading genetic patterns (see Fig. 1). The ecologies of age-classes often vary in the marine realm: pelagic larvae develop into benthic adults in many species; and in other species, juveniles will disperse widely and remain solitary whereas adults form social or breeding groups (e.g. leopard seals, Forcada and Robinson 2006). In these examples, sampling one age-class will provide a different genetic pattern to sampling the other age-class. In some species ontogenetic shifts may be very subtle, for example, larval cardinal fish settle onto rubble and sand areas away from continuous reef sites where adults reside (Finn and Kingsford 1996). These behaviors can confound the interpretation of spatial genetic patterns and co-varying environmental factors if age-class is not taken into account (also see Goldberg and Waits 2010).

The spatial genetic patterns of marine animals are also known to shift over time (Johnson and Black 1982, Planes and Lenfant 2002, Selkoe et al. 2006), potentially due to temporally variable seascape features including currents and/or certain reproductive strategies (e.g. high variance in reproductive success, Fig. 1). Although this phenomenon is also reported in terrestrial systems (mammals, Scribner et al. 1991; plants, Hossaert-McKey et al. 1996; insects, Guillemaud et al. 2003) it appears to be particularly prevalent in the marine system (Toonen and Grosberg 2011). The

occurrence of such temporal genetic shifts encourage consistent timing of sampling (i.e. consistent age-group and timing across locations), and suggest it may be appropriate to sample each population several times, so that the temporal fluctuations in allele frequencies may be captured and accounted for when describing spatial genetic patterns.

Individuals or populations?

A fundamental consideration in planning a study is whether individual- or population based sampling and methods of analysis will be employed (Anderson et al. 2010). Individual based analyses use multilocus genotypes to cluster individuals into groups (as in Pritchard 2000) and to infer relatedness and parentage (as in Gerber et al. 2003). Population level analyses typically rely on allele frequencies, although some are based on allelic richness (as in Petit et al. 1998 and Diniz-Filho et al. 2012, see below). In practice many studies, especially those based on multilocus microsatellite genotypes, will use both individual- and population-based analyses. However, sampling efforts (individual- or population-based) will often be determined by the focal organism's habit.

Many marine habitats are patchy, therefore sampling strategies familiar to terrestrial systems can also be appropriate to marine systems. In general, if the relevant spatial scale for analysis is small and the species of interest does not have a clumpy distribution but is fairly continuously distributed, then individual-based sampling may be more appropriate. Even over larger distances, individual-based analytical approaches are attractive in that adults and juveniles can be assigned to interbreeding groups based on the genotypes of sampled individuals using parentage analysis and assignment tests (as in Saenz-Agudelo et al. 2009). This ability to delineate groups based on Hardy-Weinberg expectations and linkage disequilibria is useful for species that only aggregate during breeding, but have been sampled outside of the breeding period, or sympatric populations that are segregated by spawning time only.

The primary drawback of individual-based analyses is that multiple unlinked markers (such as microsatellites) are required, whereby power is enhanced by sampling many individuals for many loci (typically 10 or more; for discussion on the effects of sample size, number of markers, and allelic richness see (Landguth et al. 2012b). Historically, marker development has been financially and technically restrictive, however recent developments in sequencing technologies has made marker development for non-model organisms, such as many marine organisms, considerably more accessible.

For studies interested in longer time frames and sampling over larger distances (which may be necessitated given the patch sizes and distances between patches for seascape features, see Fig. 1 in Chapter Two) population-level analyses based on allele frequencies or genealogies can be appropriate and informative. Population-level studies often necessitate sampling fewer individuals per location relative to individual-based methods so that more locations can be included, which can provide

Table 1. A summary of considerations relevant to the design and sampling of a seascape genetics study, common genetic measures, and some methods for the analysis of genetic patterns alongside geographic and environmental data.

Design	**Sampling**	**Genetic measures and patterns**	**Do spatial features predict genetic patterns?**
Seascape feature/s *(see Fig. 1 in Chapter One)* What is the relevant spatial and temporal scale? **Marine organism/s** *(see Fig. 1 in this paper)* What features of the species biology and ecology are relevant to the question? What is the geographic range of the species? **Molecular marker/s** Single or multiple loci? Are the loci linked? What is the mode of inheritance? Are they likely to evolve in an approximately neutral manner or are they likely to be under selection? Are they variable enough for the focal timescale?	**Consider:** Spatial scale of sampling Spatial arrangement of sampling Replication of seascape features Potential for spatial autocorrelation Individual or population based sampling Number of populations to sample Appropriate sample size (total and per population) Appropriate age class for question Potential for unsampled populations and parts of species range Aspects of the species biology and ecology (see Fig. 1)	**Geneology** Networks Phylogenetic trees **Diversity – ''within''** Measures of genetic diversity: *He*, allelic richness, allelic diversity, and similar descriptors **Differentiation – ''between''** Measures of genetic distance, differences in allele frequencies, fixation and differentiation indices: F_{ST}, D, D_{est}, and similar indices **Gene flow** Assignment tests Coalescent estimates	**Explorative** *Using geo-referenced genetic data:* phylogeography, clustering, barrier detection **Correlative** *Using genetic and environmental measures, such as remotely sensed data, biophysical data, measures derived via least-cost-path analysis, resistance surfaces, circuit theory and for graph theory:* ''within''- standard univariate or multivariate methods ''between''- Matrix comparison methods, Mantel tests, partial Mantel tests, and other multivariate approaches that accommodate non-independence, i.e. isolation-by-distance, isolation-by-environment, isolation-by-resistance, etc. **Hypothesis testing** A priori *hypotheses can be informed using methods such as:* Habitat/niche modeling, biophysical modeling and simulations *The fit of the hypothesis to the observed genetic data can be assessed using univariate and multivariate methods such as those above*

greater power for testing the effects of specific seascape features. Population-level analyses also tend to be more flexible with regards to number and type of genetic markers employed and those based on allele frequencies or coalescence will reflect historical averages, and will be less sensitive to recent population subdivision than assignment test methods based on multilocus genotypes.

Genetic variation within and between spatial units

Many summary statistics from population genetics may be used as response variables in a seascape genetic study. Genetic response variables can be divided into two main groups: those that provide a metric within a single spatial unit (an individual, population or geographic region) and those that reflect a contrast between spatial units (within and between metrics can be based on individual genotypes or population attributes). For instance, measures of genetic diversity, such as allelic diversity, allelic richness, and heterozygosity are typically expressed as values tied to an explicit spatial unit. In contrast, genetic distances and indices of genetic differentiation (e.g. F_{ST}, Wright 1943; Nei's D, Nei 1972; D_{est}, Jost 2008) express the difference between spatial units and hence are not linked to a single location but to two or more locations. The differences between populations described by genetic distances primarily reflect the duration of isolation (where genetic drift causes divergence between populations) and the magnitude of isolation (influencing gene flow since separation).

There are many differentiation indices and some are more appropriate than others in certain instances (see Bird et al. 2011 for suggestions). For example, gene flow can be inferred as the inverse of F_{ST}, however the underlying population model assumes no mutation or change in population size (for a complete discussion of factors affecting and appropriate interpretation of this metric see Whitlock and McCauley 1999). Classic population genetic models often have underlying assumptions that will be unrealistic for many marine species (Selkoe et al. 2008) and could lead to the erroneous interpretation of genetic patterns (see Karl et al. 2012 for a discussion on this topic). Recently it has become clear that the inherent variability of genetic markers can also bias some differentiation indices, and this phenomenon is of particular concern in seascape genetics because marine organisms typically display high genetic diversities (see Meirmans and Hedrick 2011 for an overview and suggested solution).

In study species where gene flow should be substantial and/or past changes in population size are likely, assignment tests or coalescent methods are more appropriate estimators of genetic connections than those based on genetic distances or differentiation indices (see Marko and Hart 2011 for a recent review of approaches). These methods can infer directionality of gene flow and therefore may be more useful for evaluating asymmetric processes such as transport by currents and investigating source-sink dynamics. However, assignment tests and coalescent approaches are often computationally intensive, and like most applications of genetic distance metrics, usually assume selective neutrality.

The most common metrics used to describe genetic variation *within* and *between* are discussed above, but there are other approaches that bridge this distinction and may allow a more nuanced understanding of genetic patterns in the marine system. For instance, Foll and Gaggiotti (2006) have developed a population-specific metric of F_{ST} that estimates how distinct a population is relative to all others. The authors were able to use this metric to relate local F_{ST} to environmental factors using a linear model (discussed below). Petit et al. (1998) have also suggested a method for partitioning the allelic richness of a population to reflect richness due to novel alleles (divergence), versus those shared with other populations.

Borrowing theory and measures from other fields of science could provide novel approaches to characterizing genetic patterns in the sea. For example, Diniz-Filho et al. (2012) presented an approach, whereby differences in allelic richness between populations are partitioned into "turnover" (alleles found in one population and not the other) and differences in richness (when one population's diversity is nested within the other) in a similar manner to the partitioning of beta-diversity in ecological investigations (e.g. Carvalho et al. 2012). In a very different approach, Dyer and Nason (2004) pioneered a method using graph theory to derive conditional genetic distances (*c*GD) among populations that take into account the relationship each population has with every other population in the study (for an introduction to graph theoretic representations of genetic data see Garroway et al. 2008, Dale and Fortin 2010). In their method, a population graph is constructed, in which nodes (populations) are only connected by edges when there is genetic covariance (based on the *c*GDs) between them. One of the several advantages associated with using such ecological or graph theoretic approaches is that it allows access to a well-established suite of analyses.

Evaluating relationships between environmental predictors and genetic patterns

Seascape genetics, like landscape genetics, is concerned with patterns resulting from dispersal and gene flow (or conversely, barriers that restrict gene flow) and habitat characteristics that modify movements and successful immigration (quality, available space, predators, etc.). The interdisciplinary nature of landscape genetics offers spatially explicit and quantitative methods to assess the relationship between genetic patterns and environmental features and can complement other well-established approaches such as phylogeography (Avise 2000) and habitat or niche modeling (Elith and Leathwick 2009).

A common null hypothesis for genetic patterns established over space has been isolation-by-distance (IBD, Wright 1943, Slatkin 1993, Rousset 2000), where a positive relationship is expected between geographic distance and genetic differentiation for continuously distributed populations (or individuals) that approximate an equilibrium between migration and genetic drift. Indeed, some marine organisms exhibit an IBD pattern (Selkoe and Toonen 2011), but others have an irregular pattern of genetic differentiation (i.e. "chaotic patchiness"), suggesting seascape features unrelated to Euclidean distance also influence genetic patterns (Selkoe et al. 2008, Chapter Two).

In terrestrial studies, ecological distances are often modeled by weighting the cost (resistance) of traversing various habitats or features and using least-cost-path analyses (as in Spear et al. 2010) or isolation-by-resistance (McRae 2006) within a geographic information system (GIS). In marine studies, simple over water distance is a commonly used least-cost-path approach, and the increasing availability of remote sensing tools allows investigators to categorize or rank some seascape attributes to create resistance surfaces. Therefore some simplified least-cost-path and isolation-by-resistance approaches can be easily adapted to the seascape. But other, more ecologically meaningful estimates are difficult to implement within a dynamic ocean environment (Galindo et al. 2010). For example, circuit theory (McRae 2008) which takes every possible path among populations into consideration simultaneously, may be inappropriate in systems where dispersal is likely to have directionality. In these instances, biophysical models that use ocean dynamics to capture asymmetries in dispersal, offer a more appropriate method (see next section).

Although ideally a spatial genetic study has *a priori* hypotheses regarding structuring factors (e.g. distance, barriers and/or habitat and environment characteristics), many studies evaluate their genetic patterns against qualitative predictions or use a series of statistics to examine predictions in turn. For example, clustering methods and barrier detection methods identify natural groupings based on genotypes or allele frequencies (see Guillot et al. 2009 for a review of methods) and are frequently used to infer spatial locations of genetic discontinuities; from these putative barriers, potential causes are sometimes qualitatively assessed, largely in a *post hoc* manner (Anderson et al. 2010, Holderegger and Wagner 2008, Storfer et al. 2010). Whereas such clustering approaches are useful for data exploration and hypothesis generation, their utility for testing alternative or multiple causative factors is limited (and can be misleading, Cushman and Landguth 2010). Some studies employ multiple approaches in series, for example, testing for IBD and also using an analysis of molecular variance (AMOVA, Excoffier et al. 1992) to test the effect of specific barriers, or in combination with isolation-by-environment (IBE) analyses to assess the influence of habitat or environmental characteristics (e.g. Fontaine et al. 2007). These methods have limited utility as they do not consider geographic distance (which can thought of as a source of spatial autocorrelation for IBE analyses), potential barriers and other relevant environmental characteristics in combination.

More rigorous approaches use multivariate frameworks and model testing to infer the importance of specific factors (Balkenhol et al. 2009, Cushman and Landguth 2010, Storfer et al. 2007). For genetic response variables that are tied to a specific location (within variables), linear models can be used, but caution should be taken as spatial autocorrelation may exist between these sites where within variables are measured (see Beale et al. 2010, for example approaches of accounting for this). Similar care should be exercised when using statistics relying on the relationship between locations. These measures violate the statistical assumption of independence, as each single locational unit will be involved in multiple pairwise comparisons. This issue of non-independence has been long recognized in testing for IBD and is often resolved by the use of Mantel tests (Mantel 1967) where

significance is assessed via permutation (although there are other methods, e.g. the maximum-likelihood population-effects model, Clarke et al. 2002).

Partial Mantel tests (Smouse 1986) are often used to accommodate multiple predictive factors; however, this method has been suggested to have low power and may be misleading (Balkenhol et al. 2009, Legendre and Fortin 2010, but see Cushman and Landguth 2010). While several other multivariate methods have been proposed in landscape genetics (reviewed by Storfer et al. 2007, Balkenhol et al. 2009, Thomassen et al. 2010), many approaches are tailored to certain study systems and thus their methods are not easily transferred. This is particularly the case for seascape genetic analyses where the system may be data depauperate or biologically dependent on very different processes. Nonetheless, the literature of landscape genetics and spatial statistics are inspiring innovative seascape genetic approaches (see Table 1 in Chapter One for empirical examples of multivariate seascape assessments).

Correlation, not causation

Most landscape genetic studies look for biologically informed correlations but are not able to test for causation. To demonstrate a mechanistic link between a putative spatial factor and genetic variation (either neutral or selective) requires evaluation, such as direct observation of dispersal and reproduction, reciprocal transplantation, common garden experiments, or functional genomics (Feder and Mitchell-Olds 2003, Lowry and Willis 2010). Within the scope of current landscape genetic techniques however, any strong correlation will be bolstered by sound explanation of the organismal biology and its relation to the environment (comparative approaches can also lend strength to weak patterns, see Selkoe et al. 2010). Hence, making an informed choice of organism, sampling design, and environmental variables or seascape features will enhance any seascape genetic study. In addition, using modeling and simulation approaches such as those described in the next section, one can test their understanding of the mechanisms that lead to the observed spatial genetic patterns (also see Epperson et al. 2010, Thomassen et al. 2010, Landguth et al. 2012a, for simulation approaches in landscape genetics).

Using biophysical models in seascape genetic studies

Many seascape features can be quantified using methods similar to those of terrestrial landscape genetics (described in the above section). However, seascape genetic studies can also draw from a complementary set of tools: biophysical models provide methods to simulate individual dispersal, population connectivity, and even genetic patterns across the seascape. In this section we describe the methods of biophysical modeling and their relevant outputs, highlight ways in which genetic and biophysical data may be used together, and point out some considerations for their use.

Methods of biophysical modeling

In a marine context, biophysical models are used to simulate the movement of individuals or propagules by incorporating ocean dynamics derived from hydrodynamic models forced by winds, tides, solar radiance, freshwater inputs, and other characteristics that may influence organismal movement, growth, behavior and survival. These physical factors are then coupled with biological attributes of the focal organism, such as (but not limited to) pelagic larval duration (PLD, Siegel et al. 2003), dispersal behavior (Dekshenieks et al. 1996), mortality (Possingham and Roughgarden 1990), and growth rate (Lett et al. 2010).

There are many approaches, methods, and techniques to model dispersal and population connectivity in the marine system. Each method should incorporate ocean features and the species' biology within a spatially realistic framework (North et al. 2009). This framework can be accomplished in many ways, ranging from a simple oceanographic distance model of marine population connectivity, which assumes this distance is an adequate proxy for the "real" dispersal process (e.g. White et al. 2010) up to a full spatially- explicit and coupled biophysical model where the individual virtual larvae have unique biological attributes and behavior (e.g. Paris and Chérubin 2007). Adding complexity (or realism) to these models may include incorporating post-settlement processes such as mortality, density-dependence, recruitment, maturity, fecundity, and other population (metapopulation) dynamics. A primary challenge is deciding what level of complexity (and spatial resolution) is appropriate for the study questions.

Common biophysical approaches include modeling the dynamics of an entire cohort of swimming virtual larvae as a cloud or plume within a complex ocean (Mora et al. 2011, Treml et al. 2008, 2012) and modeling larvae as individual particles swimming/floating in a dynamic ocean (Cowen et al. 2000, Kool et al. 2010, Mitarai et al. 2009). Individual based models (Grimm and Railsback 2005) are becoming more popular and accessible where individual larvae can be assigned properties and behavior allowing them to interact within a simulation environment. These models are also well suited to a wide range of life histories in the sea, from benthic-reef associated organisms (e.g. coral) to highly social, pelagic organisms (e.g. dolphins).

The versatility of biophysical models has been harnessed to inform population dynamics (Possingham and Roughgarden 1990), fisheries stock structure (North et al. 2009), and to estimate larval dispersal distances and patterns (e.g. mussels, Gilg and Hilbish 2003; reef fish, Cowen et al. 2006). To date, biophysical models have been used alongside seascape genetic studies exclusively to model larval dispersal and population connectivity in organisms that have a bi-partite life history. These modeling techniques are particularly attractive for studies that focus on organisms with pelagic gametes or larvae (Leis et al. 2011) as there is often no direct way to observe and quantify propagule movement among populations due to their small size and potentially large distances travelled (but see

Jones et al. 2005 for a coupled genetic and physical tracking approach). Here, we review some of the promising methods in which these sources of data can be jointly examined.

Coupling biophysical models and seascape genetic data

The output of biophysical models relevant to seascape genetic questions can include the dispersal pathway of an individual (or individuals), a species' or population's dispersal kernel describing the probability of dispersal with distance from a source, and various connectivity matrices describing the pairwise dispersal characteristics. Common connectivity matrices include the probability matrix, the population transition matrix (Caswell 1989, Cowen 2006), a migration matrix (Bodmer and Cavalli-Sforza 1967), source distribution matrix (Cowen et al. 2007), or a matrix of proportional immigration (Cowen et al. 2006). Which model output is of most interest to geneticists depends on the question and the focal timeframe. For example, genetic assignments tests, or parentage based analyses across one generation or one reproductive event can be compared to modeled dispersal estimates that also represent single dispersal events (Epperson et al. 2010). If the question is concerned more with the long-term average, and/or rare events (near the "tail" of a dispersal kernel) such as those measured using mtDNA, then averaging over many model simulations, modeling a cloud of larvae, and/or explicitly tracking rare potential events may be important.

There is no straight-forward method to quantitatively test the fit of genetic data and biophysical model outputs. Dispersal matrices derived from biophysical models are pair-wise and directional, thus intuitively they can be compared with pairwise genetic matrices of gene flow (also directional). For genetic measures that are site-specific (*within* measures), and between measures that are typically symmetrical (e.g. pairwise F_{ST} produces a triangular matrix), often the biophysically derived data and/or empirical genetic data will be converted to allow comparison.

One useful framework for investigating site-specific, pairwise, aggregate, or network-wide emergent properties of both genetic and biophysically derived data is with graph theory. The structure of complex dispersal or connectivity matrices can be represented as a network allowing graph metrics and properties to be easily calculated (e.g. node centrality, degree, flow patterns and modularity; see Treml et al. 2008, Urban et al. 2009, Treml et al. 2012). For example, Selkoe et al. (2010) used a network representation of their biophysically derived connectivity matrix to calculate a site-specific flow metric, termed "eigenvector centrality" for comparison with various measures of genetic diversity. In another approach Kininmonth et al. (2010) created a network representation of their pairwise F_{ST} matrix based on ten microsatellite loci of a brooding coral in an attempt to identify genetic "communities" (i.e. highly connected modules within the network) across the Great Barrier Reef using graph theory. Further, the authors used a network representation of hydrodynamic and distance based models for the same region to cross inform the designation of "communities" based on the likely dispersal of coral propagules.

Another promising means to assess the fit of genetic data with the simulations of a biophysical model is by using the connectivity matrices produced by the biophysical model to explain or predict the empirical genetic data. For example, Kool et al. (2010, 2011) used a matrix approach (based on the matrix model of migration developed by Bodmer and Cavalli-Sforza 1967) to project a time-averaged connectivity matrix forward in time. The projected genetic patterns (genetic diversity and genetic differentiation) could then be compared qualitatively with the empirical genetic patterns over the same seascape (as in Foster et al. 2012). (Also see Galindo et al. 2006, 2010, for a slightly different method).

In high dispersal species it may be preferable to focus on the inter-population genetic measures of gene flow (estimated from coalescent or assignment method approaches as discussed previously), rather than indices of differentiation, to maintain the directionality of relationships. For example, Crandall et al. (2012) used a probabilistic coalescent framework to model gene flow of neritid snails in the Pacific based on the asymmetric connectivity probability matrix of their biophysical model. This method enabled the authors to evaluate the relative performance of the biophysically derived matrix in describing patterns of gene flow relative to other hypothesized and classic population models.

Generally, biophysical simulations and genetic patterns across a common seascape have been in agreement; however there have been cases of inconsistency (Foster et al. 2012, Galindo et al. 2010). Discrepancies between a biophysical model (or any simulation) and observed genetic data can highlight where other processes not already captured may be operating or where the genetic assumptions and/or the model assumptions are violated.

Considerations when using biophysical models with genetic data

While the use of biophysical models and computer simulation approaches in landscape and seascape genetics is very promising (Balkenhol et al. 2009, Epperson et al. 2010), there are some considerations for their use that warrant highlighting. Some of these points are outlined below.

First, modeling methods rely on having reputable physical and biological data (Gallego et al. 2007, Metaxas and Saunders 2009). Ocean circulation models are now available for most of the world's oceans, but these are often not well resolved at small spatial scales and along complex coastlines (Cowen and Sponaugle 2009), where tides and complex topographies dominate flows. The advancement of satellite data acquisition means there is a multitude of contemporary and comprehensive environmental data available to be used in modeling approaches, however, attention to the biological parameters within these models has been less rigorous (but see Connolly and Baird 2010). This disparity is understandable given that biological parameters are difficult to quantify across different environments and species, and often require a combination of observation and experimentation that can be labor intensive (Metaxas and Saunders 2009).

Second, aligning the resolution and spatio-temporal scale between the biophysical model and genetic model/data is essential. Obviously, a biophysical model with resolution defined by a 10 km grid is not suitable, for example, to inform any genetic relationship between seagrass beds separated by 2 km. Likewise, a biophysical model that is based on contemporary ocean currents may inaccurately simulate genetic relationships that have formed over thousands of years (such as those investigated using population sampling and measured using mtDNA markers; but see Crandall et al. 2012).

Third, deciding which parameters are appropriate for inclusion and using the appropriate level of model complexity for the application is important (Gallego et al. 2007, Hannah 2007). Where a biophysical model of larval dispersal *per se* is used alongside genetic data, one is implicitly interested in the degree to which larval dispersal is driving genetic patterns across the seascape (Kool et al. 2010). In most cases, realized connectivity may also depend on habitat quality, variation in reproduction, population density and local selection (see discussion of environmental variables, above), in addition to the dispersal process. While modeling of post-settlement survival is rare (Hinckley et al. 1996), the addition of selection to biophysical models (as differential mortality or fecundity as functions of the underlying environment and individual genotypes; Epperson et al. 2010) may make them more realistic (Balkenhol et al. 2009). However, modeling approaches can quickly become computationally expensive, lose statistical power and gain uncertainty as more variables are introduced.

Fourth, biophysical models should include some level of model evaluation and parameter validation (Hannah 2007). Physical oceanographic measurements can be used to validate the physical parameters and processes and while it has been suggested that the integrated bio-physical portion can be "validated" using empirical genetic data (Galindo et al. 2006, Hellberg et al. 2002) this may not be appropriate (or possible) due to the mismatch between genetic model assumptions and biophysical modeling assumptions. Genetic data has its own inherent assumptions and inconsistencies particularly over large timeframes (but see assessment of error using probabilistic coalescent frameworks in Crandall et al. 2012). To build confidence in a biophysical model, the sensitivity of predicted and observed data matches to changing a variety of parameters should be explored (Treml et al. 2012). There are also several methods available to aid in model selection (see Hartig et al. 2011), such as Approximate Bayesian Computing (Beaumont 2010) and pattern-oriented modeling (Grimm et al. 2005, Grimm and Railsback 2012). However, as with all modeling approaches, the user must also be aware that, strictly speaking, the biophysical models cannot be validated. The model is necessarily a simplification of the natural system with a finite number of parameters and processes, and therefore will not "truthfully" represent the system (Oreskes et al. 1994). As a result, model evaluation, as opposed to "validation", is a key component in integrating models and empirical data.

Lastly, biophysical modeling is quite technical and requires substantial expertise in physical oceanography and marine ecology, and available/adequate computing resources. Although biophysical models are being made more accessible (see Condie et al. 2005, Roberts et al. 2010), their use requires sound data, appropriate parameters, and thorough understanding of the inherent assumptions, and what the model output represents.

Conclusions

Undertaking a seascape genetic study requires an understanding of the focal organisms' biology and how it is likely to interact with the seascape features of interest (see Fig. 1). Using well-considered sampling design (i.e. spatial arrangement, individuals or populations) and genetic methods (i.e. choice of marker, methods of analyses) is invaluable in teasing apart the relative influences of competing seascape features (see Table 1). Methods of analysis are becoming more rigorous and spatially explicit offering better opportunities for the interpretation of the genetic patterns, however taking an informed approach to any seascape genetic question and study design cannot be underestimated.

Seascape genetics is increasingly forming a discipline distinct from ''landscape genetics'' (Chapter One). The biology of marine organisms and the marine environment inspire questions that are often distinct from questions asked of terrestrial systems, some of which require the development of new techniques and the melding of different areas of expertise. The topic of greatest marine-focused development relevant to landscape genetics has been biophysical modeling. Despite there being several challenges remaining, the integration of biophysical modeling and seascape genetics is exciting not only for the knowledge that it will generate about the marine system, but because it encourages collaboration and mutual understanding between experts and across disciplines.

Glossary

A Molecular Analysis Of Variance (AMOVA): a method of partitioning hierarchical genetic diversity into groupings that is analogous to an analysis of variance: proportions of variance are expressed according to hierarchies, such as within and among populations, within and among groups etc. (Excoffier et al. 1992).

Alleles: different forms (polymorphisms) of the same marker (locus). Allelic diversity: a measure of how many alleles are in a population for a locus. It can be expressed as a number or a proportion.

Allele frequencies: the frequency of an allele in a population is called the allele frequency. Genetically distinct populations will differ in their composition or frequency of alleles.

Allelic richness: the number or proportion of alleles within a population with a correction for sample size bias (El Mousadik and Petit 1996).

Assignment tests: a statistical approach that assigns an individual to the sampled population from which its genotype is most likely to be derived.

Bayesian: a field of statistics that combines data with prior information about parameter values in order to derive posterior probabilities of different models or parameter values.

Benthic: marine organisms that live on, in or attached to the sea floor.

Biophysical modeling: couples physical and biological data to model a biological system. Bi-partite life history: many marine organisms have a bi-partite life history, whereby they have a planktonic dispersive early stage (as gametes, eggs and/or larvae), after which they "settle" (and metamorphose) to resume a benthic juvenile and adult stage.

Bottleneck (genetic): a population bottleneck occurs when the effective population size, *Ne*, decreases substantially. A bottleneck causes an immediate decrease in genetic diversity, promoting stochastic genetic drift.

Coalescence (times): the event (or timing) of common ancestry for two alleles found in the present day population. For example, mtDNA of two siblings coalesces in the previous generation as they both received their copies from their mother.

Chloroplast DNA (cpDNA): cytoplasmic elements containing a circular genome. They share many properties with mtDNA, including maternal transmission, but are only found in algae and plants.

Circuit theory: electrical circuit theory that can be used to model connectivity. Models based on circuit theory have the advantage of being able to evaluate the contributions of multiple dispersal pathways, simultaneously considering redundancy in a connection and also increasing connectivity via multiple pathways (McRae 2008).

Dispersal kernel: a probability density function describing how far individuals disperse from their place of origin. The mode has demographic relevance, whereas the tail is relevant on an evolutionary level (Paris and Chérubin 2007).

Effective population size (N_e): an index of how many individuals are passing on their genetic material. It represents the size of the ''ideal population'' that would lose variation or "drift" at the same rate observed in the real biological population. N_e is an important parameter in population genetics and can be used to model and predict drift and changes in genetic diversity.

Euclidean geographic distance: the shortest path between two geographic points; may account for the curvature of the earth.

F_{ST}: F_{ST} describes the proportion of genetic variation that is attributable to variation between populations relative to the total variation among all populations (Wright 1943). There are many specific methods for calculating F_{ST} and its derivatives (see Meirmans and Hedrick 2011).

Gene flow: the spread of alleles/genes from one population to another resulting from migrant individuals moving among populations. In the absence of selection and drift, gene flow would eventually homogenize allele frequencies across populations. Mathematically gene flow is expressed as N_em, the product of the effective population size (N_e) and migration rate (m). N_em is the number of migrant individuals per generation.

Genealogy: a genealogy portrays the ancestral relationship between individuals and is usually presented in a tree-like form. Common usages of genealogies include tracing the inheritance of alleles across generations, or a genealogy can represent a summary of evolutionary changes for a locus whereby splitting events on the tree represent mutation creating a new variant.

Genetic admixture: when interbreeding occurs between two genetically differentiated populations, the resultant population and individuals are considered ''admixed''. Admixture is a common source of linkage disequilibria.

Genetic distance: a measure of genetic distinctness between populations, which should be proportional to the amount of time since the populations diverged (assuming the loci are neutral and there has no gene flow following divergence). There are many specific methods for calculating genetic distance, for instance by changing the weightings among mutational models (Nei and Kumar 2000).

Genetic drift: changes in allele frequencies caused by random effects of sampling when gametes are passed from one generation to the next. Genetic drift is the main process leading to neutral genetic structure and is inversely proportional to effective population size; small populations experience greater genetic drift.

Genetic variation or genetic diversity: a measure of heritable attributes, generally but not exclusively at the DNA level (i.e. single nucleotide polymorphisms, haplotypes, alleles, etc.). Genetic variation is quantifiable within individuals (diploid heterozygous individuals have different alleles at the same loci), among individuals, among populations, and among species.

Genomics: the study of the function of genes and the structure and evolution of genomes.

Genotype: the precise combination of alleles (typically across many loci) found in an individual.

Graph theory: a body of mathematical knowledge in which structural units are depicted as nodes with relationships between them depicted as links. Graph theory provides a flexible framework that can clarify the relationship between structures and processes, including the mechanisms of configuration effects and compositional differences.

Habitat/niche modeling: the process of using algorithms to predict the distribution of species in geographic space based on a mathematical representation of their known distribution in environmental space (i.e. habitat/niche).

Haplotype: a stretch of DNA that may include multiple polymorphic sites. Most frequently, the term is used to refer specifically to DNA sequences from haploid markers (such as mtDNA or cpDNA). Technically, however, the two copies (alleles) of nuclear diploid DNA sequences are also haplotypes.

Haplotype diversity: a measure of how many haplotypes are in a population for a locus. May be expressed as a number or a proportion.

Hardy-Weinberg Equilibrium (HWE): a population genetic principle stating that allele and genotype frequencies reach equilibrium within a single generation and remain constant, assuming that a population is large (unlimited), has random mating, no genetic drift, no selection, no gene flow, and no mutation. HWE is often an assumption underlying population genetic analysis, and is a useful null model often used to infer assortative mating, selection, or migration.

Heterozygosity (*He*): a measure of genetic diversity. Within an individual, heterozygosity is the proportion of loci with two different alleles. Within populations, heterozygosity can be expressed as observed or expected heterozygosity. Observed heterozygosity (*Ho*) is the proportion of individuals in the population that are heterozygous (for the locus or loci of interest). The expected heterozygosity (*He*) is the heterozygosity that would be observed if there was complete random mating; it is calculated from allele frequencies rather than observed genotypes.

Island model of migration: a population genetic model that combines the effects of gene flow and genetic drift. The model assumes: there is an infinite number of populations, populations each have a constant size (*N*), individuals migrate among populations at a constant rate of m, every individual is equally likely to migrate, there is no mutation or selection and every population is in a migration-genetic drift equilibrium (Wright 1931).

Isolation-by-distance (IBD): the pattern of local genetic differences that can accumulate under geographically restricted dispersal. It is based on stepping stone model of migration, and assumes that

migration between populations occurs at the same rate as genetic drift within populations (e.g. a migration-drift equilibrium). IBD results from the expectation that genetic distance will increase with geographic distance (Wright 1943).

Isolation-by-environment (IBE): the pattern of local genetic differences that can accumulate due to the local environment. A pattern of genetic differences between populations that are correlated with environmental differences in their habitats is referred to as an IBE pattern (Mendez et al. 2010).

Isolation-by-resistance: the pattern of local genetic differences that can accumulate due to the landscape resistance, which results from different landscape elements filtering gene flow in differing ways (McRae 2006).

Landscape genetics: a field of study that quantifies the effect of landscape and/or environmental characteristics on gene flow or spatial genetic variation (Storfer et al. 2007).

Least-cost-path analyses (LCP): with least-cost-path analysis, connectivity values are based on the path of least resistance between any two landscape elements.

Lineage: an evolutionary lineage is a group of species, or populations or gene variants that form an exclusive line of descent, often represented in a phylogenetic tree or network.

Linear model: a mathematical model describing the effect of one or more predictive variables on an observed response variable. In land- or seascape genetics, environmental features usually comprise predictive variables and a genetic metric is the response variable.

Linkage disequilibria: a pattern found when alleles from different loci do not assort independently. Linkage equilibrium (the independence of loci) is assumed in many population genetic methods. Linkage disequilibrium can be caused by physical linkage (i.e. loci are close together on the genome) or by demographic events such as bottlenecks followed by rapid expansion, mixing between previously isolated groups, extreme drift in small populations, and selection.

Locus (plural: loci): a locus is any region of the genome. The term is vague with regards to size and function of the genomic region in question.

Mantel tests: a permutation-based statistical test describing the correlation between two distance or dissimilarity matrices (Mantel 1967). A partial Mantel simultaneously accounts for the effects of other distance matrices (Smouse 1986).

Marker (genetic): a genetically heritable and variable trait or locus. Microevolutionary processes: evolutionary processes including mutation, selection, gene flow, and genetic drift that lead to a change in allele composition and allele frequencies within a population over time.

Microsatellite (loci): simple DNA sequence repeats, typically of 2–6 base pairs motifs. Both alleles are discernable for diploid organisms and microsatellites are well-known for their high rate of mutation making them well-suited for inferring demographic processes over relatively recent timescales.

Migration: in population genetics, migration (m) is the rate (i.e. proportion of the total population) at which migrants are exchanged among populations.

Mitochondrial DNA (mtDNA): cytoplasmic elements that contain a small circular genome and are transmitted from mothers to their progeny. Genetic studies of animals frequently focus on the mtDNA because it is easy to sequence and also maternal transmission reduces the total copies in a population relative to nuclear genes (as mtDNA is both haploid and only persists through female lineages). The smaller population size of mtDNA relative to nuclear loci means that genetic drift acts more efficiently changing allele frequencies in populations more rapidly (albeit over evolutionary time scales of tens of thousands or more years).

Mutation: an alteration to the genome, which creates new alleles. The rate and process of mutation varies by locus. A point mutation alters the nucleotide sequence for a single nucleotide, insertions add new nucleotides, and deletions remove nucleotides.

Neutral locus: a locus evolving without the influence of selection. Although selection cannot be statistically detected for many loci and neutrality is therefore assumed in these instances, in reality it is unclear whether any locus is ever absolutely neutral. Most population genetic methods of inferring gene flow assume (approximate) neutrality.

Neutral population processes: gene flow, genetic drift and mutation are neutral population processes. If there is no selection and loci are neutral, only drift and gene flow affect the fate of a new allele created by mutation.

Parentage analysis: a statistical approach that assigns an individual to parents or parental populations based on their genotypes.

Pelagic larval duration (PLD): the length of time a larva spends in the pelagic environment after hatching and before settling onto a reef and/or metamorphosing.

Pelagic: marine organisms (or their gametes) that live in the open sea, or the water column away from the benthos.

Population genetics: a field of biology that studies the genetic composition of populations, and the changes in genetic composition that result from microevolutionary processes.

Population (genetic) structure: a descriptor of the tendency for individuals within a population to be more genetically similar than individuals from different populations. Species biology and geographical features influence the degree of population genetic structuring found across geographic space (via microevolutionary processes).

Phylogeography: a field of study concerned with the microevolutionary and geographical processes governing the geographic distributions of genealogical lineages within and among closely related species (Avise 2000).

Recruitment: the process and phase by which "settlers" in a marine population successfully recruit into the adult population.

Seascape genetics: an area of study that evaluates the effects of spatially variable structural and environmental features on genetic patterns of marine organisms; equivalent to marine landscape genetics.

Selection: the non-random survival or mortality (of individuals) associated with a specific heritable trait(s). Various statistical tests can infer the imprint of selection on specific loci.

Self-recruitment: also called local replenishment; recruitment into a population from itself.

Settlement: the process of pelagic larvae transitioning into a benthic life style. Metamorphosis occurs during the settlement process.

Source-sink dynamics: a metapopulation construct where a "source" is a population in which the net export of individuals is greater than the net import of individuals; the reverse is a "sink".

Unlinked markers: loci that are not physically linked. Using unlinked markers increases the statistical power to infer past population history and locus-specific selection.

Acknowledgements

We thank HP Possingham, ED Crandall, JA Kupfer and one anonymous reviewer for comments that substantially improved the manuscript. LL was supported by an Australian Postgraduate Award from the Australian Government and a Queensland Government Smart Futures PhD Scholarship. Many of the ideas discussed here grew out of work funded by the Australian Research Council (DP0878306, to CR), the Sea World Research and Rescue Foundation (SWR/1/2012, to CR and LL), a Paddy Pallin Foundation and The Foundation for National Parks and Wildlife Science Grant, an Ecological Society of Australia Student Research Grant, the Lerner Gray Memorial Fund of the American Museum of Natural History, a Great Barrier Reef Marine Park Authority's Science for Management Award, and an Explorer's Club Exploration Fund (to LL).

CHAPTER FOUR. The roles of historical and contemporary seascape processes in forming spatial genetic patterns of four Indo-Pacific reef fishes

In preparation for Journal of Biogeography

Libby Liggins[1], Eric A. Treml[1,2], Hugh P. Possingham[1,3] and Cynthia Riginos[1]

[1]School of Biological Sciences, The University of Queensland

[2]Department of Zoology, University of Melbourne

[3]Department of Life Sciences, Imperial College London

Abstract

Aim To determine what historical and contemporary seascape processes have shaped the spatial genetic patterns of four coral reef fishes, and to identify any common responses among species related to species dispersal potential (based on egg type and pelagic larval duration, PLD).

Location The Indian and Pacific Oceans, with a particular focus on the Indo-Australian-Archipelago.

Methods Four coral reef fish species of varying dispersal potential were sampled throughout the Indian and Pacific Oceans (*Dascyllus trimaculatus*, *Pomacentrus coelestis*, *Hailchoeres hortulanus*, and *Acanthurus triostegus*). We characterized spatial genetic patterns (based on mtDNA control region) using AMOVA-based clustering, measures of genetic differentiation, and genetic diversity. The correlation of spatial genetic patterns among species was addressed using the congruence among distance matrices (CADM) method and multivariate analyses of variance. The historical and contemporary seascape processes associated with patterns of genetic differentiation for each species were inferred using multiple regression of distance matrices (MRDM) and stepwise model selection.

Results Species with similar PLDs (*P. coelestis* and *H. hortulanus*) had correlated patterns of genetic differentiation (Φ_{ST}) across their co-sampled ranges. MRDM indicated that the common influence of oceanographic distance was underlying their correlated genetic patterns. Species with pelagic eggs (*A. triostegus* and *H. hortulanus*) also had correlated patterns of genetic differentiation (D_{est}), however a common underlying seascape process could not be inferred. Additionally, a common influence of the Torres Strait and the Lydekker's/Weber's line was inferred for the genetic patterns of differentiation for *P. coelestis* and *A. triostegus*, despite their differing dispersal potential.

Main conclusions The dispersal potential of species (based on egg type and PLD) did not consistently predict which species had similar spatial genetic patterns. Furthermore, the association of historical versus contemporary processes with the spatial pattern of genetic differentiation was not associated with the dispersal potential of the study species. Our study suggests a focus on patterns of genetic

structure alone will fail to recognise the role of processes that lead to subtle spatial patterns of genetic differentiation that may be shared among species. Last, our study characterizes the complementary strengths of genetic differentiation measures (Φ_{ST}, F_{ST}, and D_{est}) across a common seascape.

Introduction

The distribution of genetic diversity and genetic structuring across a species range reflects the combined effects of historical and contemporary physical processes (Avise 2000, Hellberg 2007). Whether co-distributed species are influenced by such processes in similar, or predictable ways, is an area of great interest (e.g. Alvarez et al. 2009, Fortuna et al. 2009). In some compelling cases, multiple species have concordant patterns of genetic structure that also align with certain physical features (e.g. Avise 1992, Thiel-Egenter et al. 2009, DeBoer et al. 2014). In situations of phylogeographical concordance (*sensu* Avise and Ball 1990) there is often a clear indication of what process underlies the patterns. When there is no overwhelming concordance among species, inferring what processes are responsible, and which processes have had a common influence on genetic patterns across species, is challenging. However, if patterns vary among species in a way that can be predicted based on a known interaction between their biological traits and the physical processes of interest, we are then enabled to predict patterns in species not otherwise represented in genetic surveys.

Understanding which species traits are important determinants of spatial genetic patterns in marine systems has received much attention. For benthic marine organisms, meta-analytical approaches have highlighted the role of the early life history dispersal stage in forming patterns of genetic structure and geographic scales of genetic differentiation (the isolation-by-distance relationship, IBD, Slatkin 1993). Specifically, fish species with benthic eggs tend to have greater population genetic structure across their range than species with pelagic eggs (global F_{ST}; Riginos et al. 2011, 2014). Furthermore, the length of a species' pelagic larval stage (pelagic larval duration, PLD) has been observed to have a weak positive relationship with the slope of their IBD relationship (Siegel et al. 2003, Selkoe and Toonen 2011, but see Bradbury et al. 2008, but no relationship with genetic structure (i.e. when direct developers excluded: Weersing and Toonen 2009, Riginos et al. 2011, Kelly and Palumbi 2009). Nonetheless, simulations that incorporate the early life history attributes of such organisms clearly support an inverse relationship between PLD and genetic structure (Faurby and Barber 2012) and evolutionarily-significant levels of migration, that would lead to population genetic structure (Treml et al. 2012).

The spatially implicit design of many meta-analytical studies and contemporary focus of simulation studies provide little information about how historical processes affect spatial genetic patterns. In contrast, empirical genetic studies have provided very detailed accounts of the effect historical processes can have on the spatial genetic patterns of marine organisms. Studies have demonstrated that co-distributed species sometimes have spatial genetic structure coincident with

contemporary oceanographic features (e.g. inter-island channels, Toonen et al. 2011) despite differences in their dispersal potential. Other studies have advocated a role of dispersal related traits in determining whether historical processes have impacted spatial genetic patterns (e.g. Sherman et al. 2008) and suggest genetic patterns of high dispersal species are more associated with contemporary processes than historic features of the seascape (Pelc et al. 2009). These studies suggest that the simultaneous consideration of species' dispersal potential, and both historical and contemporary seascape processes, is necessary for understanding the processes underlying the spatial genetic patterns of marine organisms.

Although comparative phylogeography aims to identify phylogeographically concordant breaks across species, the common influence of a physical process across species may not necessarily manifest as shared spatial genetic patterns. For example, the slope of the Isolation-by-(oceanographic)-distance (IBoD) relationship does not visually manifest on a map, yet is still an important spatial genetic pattern in response to a physical feature (i.e. oceanographic distance). A study by Crandall et al. (2012) demonstrates this point in a multi-species comparison: despite qualitative differences in the population genetic structure of four marine gastropod species, the investigators found that a biophysical model of larval dispersal among South Pacific archipelagos provided the best explanation for the spatial genetic patterns of all species. Thus, subtle patterns of genetic differentiation can be informative when analyzed in a quantitative framework (also demonstrated by Selkoe et al. 2010).

In this study we focus on spatial genetic patterns across the tropical Indian and Pacific Oceans, and in particular the juncture of the oceans around the Indo-Australasian-Archipelago (IAA). The IAA is a tropical marine biodiversity hotspot (Roberts et al. 2002) and several of the hypotheses explaining this phenomenon, would also predict high genetic diversity and phylogeographic structure (Bowen et al. 2013). Several suture zones delineate species boundaries (including the lines drawn by Lydekker 1896 and Weber 1902, Fig. 1) according to historical continental affiliations, but may also affect intraspecific lineages in similar ways (as suggested by Rocha et al. 2007, Avise 1992, and found by DeBoer et al. 2014). Most of the population genetic structuring however, has been attributed to the climate oscillations of the Pleistocene (~2.5mya - 12kya, reviewed in Bellwood et al. 2012) as the shallow seas in the IAA exposed parts of the Sunda and Sahul Shelves (Voris 2000). During this time pockets of shallow benthic marine habitat became isolated and connectivity among the oceans was restricted to a narrow passage of water between the continental shelves (Fig. 1). Since ~7kya there has also been a point of contact for marine animals between the Indian and Pacific oceans via the Torres Strait (Voris 2000, Reeves et al. 2008). Biophysical models of dispersal suggest the potential for high population connectivity through the junction of the Pacific and Indian Oceans via the Torres Strait, the Indonesian Through Flow that runs between the continental shelves (Fig. 1), and several smaller currents within the IAA (Kool et al. 2011, Treml et al. 2012). Thus, contemporary dispersal may have a large role in reconnecting populations that were isolated during other periods in history.

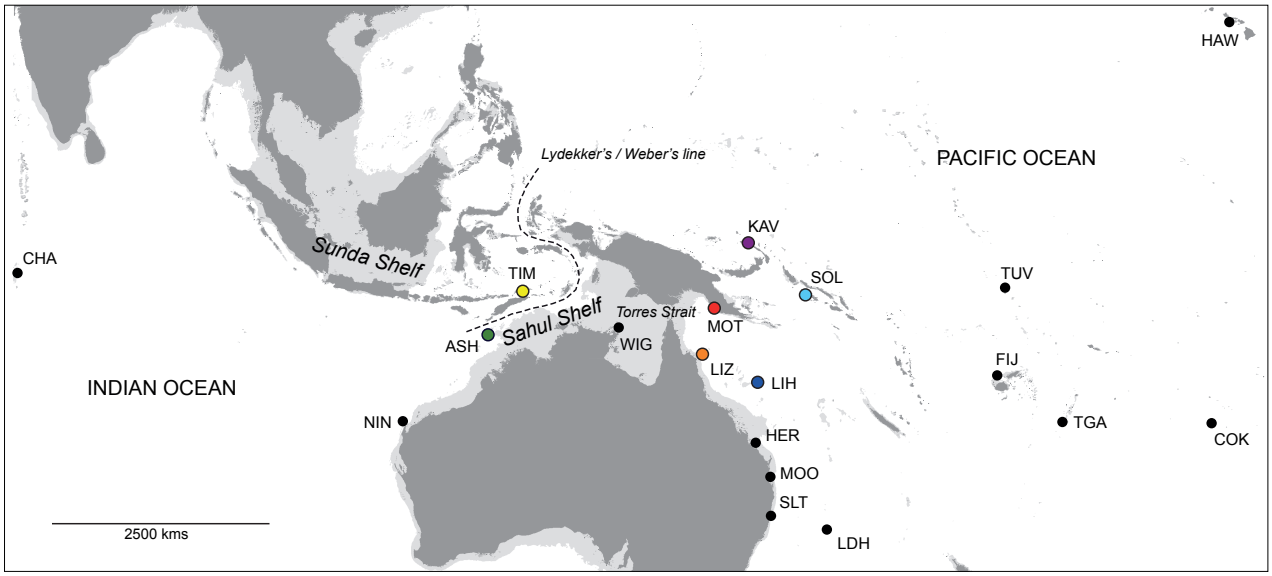


Figure 1. Map of the study area including sampled locations (listed in Table 1.) and geographic features. Dark gray represents present day landmasses; light gray represents greatest landmass extent during the Last Glacial Maximum including the Sunda Shelf and the Sahul Shelf. Colored points denote locations for which all four study species were co-sampled.

Table 1. Sampled locations for each species and related genetic diversity and demographic statistics: nucleotide diversity, π; haplotype diversity, $H_{diversity}$; effective number of haplotypes, $H_{effective}$; number of private haplotypes, $H_{private}$; Tajima's *D*; Fu's F_s; and demographic Tau, τ. Significance of demographic statistics denoted by asterisks: **P* < 0.05, ***P* < 0.01, ****P* < 0.001. τ was only estimated for locations that were co-sampled for every species.

Species/Location		Longitude, latitude	n	π	$H_{diversity}$	$H_{effective}$	$H_{private}$	*D*	F_s	τ
Dascyllus trimaculatus										
CHA	Chagos archipelago	72.01°E, 6.08°S	18	0.0239	0.9608	10.80	12	-0.30	-1.09	-
NIN	Ningaloo reef	113.98°E, 21.73°S	15	0.0144	0.9429	8.33	9	-1.06	-4.42*	-
ASH	Ashmore reef	123.13°E, 12.24°S	17	0.0154	0.9853	13.76	10	-1.21	-7.98	7.11
TIM	Timor-Leste	126.42°E, 8.30°S	16	0.0142	1.0000	16.00	8	-1.59*	-13.11***	4.37
LIZ	Lizard Island	145.49°E, 14.69°S	15	0.0120	0.9619	9.78	5	-0.89	-5.53**	5.68
MOT	Motupore Island	147.24°E, 9.44°S	16	0.0123	0.9667	10.67	7	-1.17	-6.39**	5.06
KAV	Kavieng	150.78°E, 2.57°S	16	0.0137	1.0000	16.00	8	-1.84*	-13.17***	2.75
LIH	Lihou reef	151.53°E, 17.62°S	6	0.0138	0.9333	4.50	3	-1.07	-0.33	5.93
SOL	Solomon Islands	157.41°E, 8.17°S	17	0.0141	0.9853	13.76	9	-1.12	-8.60***	7.12
FIJ	Fiji	177.67°E, 16.73°S	5	0.0109	1.0000	5.00	1	-1.19	-1.72	-
TGA	Tonga	175.20°W, 21.18°S	15	0.0149	0.9810	11.84	8	-0.26	-6.11**	-
COK	Cook Islands	159.78°W, 21.24°S	9	0.0369	0.8333	3.86	5	1.79	2.32	-
Pomacentrus coelestis										
NIN	Ningaloo reef	113.98°E, 21.73°S	24	0.0150	0.9964	22.15	19	-1.47	-22.06***	-
ASH	Ashmore reef	123.13°E, 12.24°S	32	0.0151	0.9839	21.33	21	-1.04	-23.38***	4.88
TIM	Timor-Leste	126.42°E, 8.30°S	19	0.0198	0.9649	11.65	12	-1.20	-6.94**	4.31
WIG	Wigram Island	136.57°E, 11.76°S	16	0.0096	0.8917	6.10	8	-1.81*	-3.59*	-
LIZ	Lizard Island	145.49°E, 14.69°S	16	0.0141	0.9417	8.53	7	-0.95	-4.59*	2.87
MOT	Motupore Island	147.24°E, 9.44°S	15	0.0112	0.9714	10.71	8	-1.36	-8.53***	3.07
KAV	Kavieng	150.78°E, 2.57°S	17	0.0110	0.9485	9.32	9	-1.55*	-7.08***	1.19
LIH	Lihou reef	151.53°E, 17.62°S	11	0.0087	0.9636	8.07	4	-1.01	-4.58**	2.25
HER	Heron Island	151.93°E, 23.49°S	20	0.0087	0.9474	10.00	10	-1.76*	-12.89***	-
MOO	Mooloolaba	153.12°E, 26.64°S	19	0.0136	0.9591	10.94	8	-1.49	-8.07***	-

Table 1. Continued

SLT	Solitary Islands	153.15°E, 30.32°S	14	0.0100	0.9341	7.54	7	-1.74*	-5.43**	-
SOL	Solomon Islands	157.41°E, 8.17°S	16	0.0063	0.8250	4.41	8	-1.99*	-4.85**	2.45
LDH	Lord Howe Island	159.07°E, 31.52°S	18	0.0114	0.9281	8.10	6	-1.11	-6.10**	-
Halichoeres hortulanus										
CHA	Chagos archipelago	72.01°E, 6.08°S	29	0.0091	0.9089	8.17	12	-1.37	-9.26***	-
ASH	Ashmore reef	123.13°E, 12.24°S	15	0.0189	0.9905	13.24	8	0.22	-7.01**	1.55
TIM	Timor-Leste	126.42°E, 8.30°S	5	0.0234	1.0000	5.00	2	0.69	-0.57	13.57
LIZ	Lizard Island	145.49°E, 14.69°S	15	0.0208	0.9905	13.24	9	0.38	-6.53**	11.50
MOT	Motupore Island	147.24°E, 9.44°S	15	0.0194	0.9905	13.24	6	1.35	-6.85**	10.13
KAV	Kavieng	150.78°E, 2.57°S	16	0.0237	0.9830	12.80	9	0.04	-4.89*	11.04
LIH	Lihou reef	151.53°E, 17.62°S	16	0.0196	0.9750	11.64	8	0.24	-5.71**	10.30
SOL	Solomon Islands	157.41°E, 8.17°S	15	0.0162	0.9619	9.78	7	-0.58	-4.09*	2.51
FIJ	Fiji	177.67°E, 16.73°S	5	0.0150	1.0000	5.00	1	-0.45	-1.22	-
TGA	Tonga	175.20°W, 21.18°S	16	0.0168	0.9667	10.67	8	-0.39	-4.82*	-
Acanthurus triostegus										
NIN	Ningaloo reef	113.98°E, 21.73°S	18	0.0046	0.8366	4.77	3	-1.07	-1.30	-
ASH	Ashmore reef	123.13°E, 12.24°S	15	0.0022	0.7048	2.92	1	-0.95	-1.06	2.88
TIM	Timor-Leste	126.42°E, 8.30°S	15	0.0058	0.7714	3.57	2	-0.57	0.60	7.75
LIZ	Lizard Island	145.49°E, 14.69°S	15	0.0069	0.9143	6.82	5	-0.68	-1.71	9.99
MOT	Motupore Island	147.24°E, 9.44°S	7	0.0038	0.8095	3.27	0	-0.77	0.75	0.74
KAV	Kavieng	150.78°E, 2.57°S	15	0.0051	0.8190	4.25	3	-1.31	0.24	1.06
LIH	Lihou reef	151.53°E, 17.62°S	15	0.0037	0.8857	5.77	4	-1.37	-2.78	3.09
SOL	Solomon Islands	157.41°E, 8.17°S	15	0.0047	0.8667	5.23	0	-1.22	-0.92	8.47
TUV	Tuvalu	179.20°E, 8.52°S	16	0.0065	0.8250	4.41	3	-0.82	-1.61	-
FIJ	Fiji	177.67°E, 16.73°S	11	0.0045	0.8727	4.84	3	-1.22	-1.10	-
TGA	Tonga	175.20°W, 21.18°S	7	0.0067	0.9524	5.44	2	-0.18	-0.93	-
COK	Cook Islands	159.78°W, 21.24°S	30	0.0045	0.8345	5.17	6	-1.38	-1.41	-
HAW	Hawaii	158.00°W, 21.41°N	11	0.0053	0.9455	7.12	8	0.24	-1.86	-

Comparing Indian and Pacific Ocean populations, genetic studies have suggested that whereas some taxa have genetic differentiation across the IAA (e.g. Lavery et al. 1996, Bay et al. 2004, Gaither et al. 2011), others do not (e.g. Lessios et al. 2003, Klanten et al. 2007, Horne et al. 2008). Moreover, the position of genetic breaks within the IAA often varies greatly across species (reviewed in Carpenter et al. 2010). However, these empirical results are difficult to compare quantitatively. Each study has their own sampling design, and generally focus on only one static spatial predictor of the genetic patterns (such as the entire IAA, e.g. Gaither et al. 2011, Horne et al. 2008; or the Torres Strait, e.g. Mirams et al. 2011 – Appendix One, Lukoschek et al. 2007), or describe locations of genetic breaks and subsequently infer the underlying processes (Barber et al. 2006, Timm and Kochzius 2008, DeBoer et al. 2014, but see Giles et al. 2014). Yet, the impacts of historical and contemporary processes on spatial genetic patterns rarely manifest as distinct genetic breaks. Furthermore, discordant spatial genetic patterns may not necessarily discount the common influence of a physical process on the genetic patterns of species.

Here we assess congruence in the genetic patterns (genetic diversity, structure, and differentiation) of four coral reef fish species through the IAA. We address which seascape features and processes in this region have a common influence on their patterns of genetic differentiation. Seldom have several seascape features and processes, both historical and contemporary, been considered in a seascape genetic study (but see Crandall et al. 2014 – Appendix Two) and rarely in an empirical multispecies comparison (reviewed in Riginos and Liggins 2013 – Chapter One). We expect that egg type will correlate with geographic structure established over time (Riginos et al. 2011, 2014) and that contemporary patterns of gene flow among populations likely vary in predictable ways across species (related to PLD, Faurby and Barber 2012, Treml et al. 2012). These expectations underlie five sequential null hypotheses based on the egg type and PLD of species (hereafter conferring low dispersal or high dispersal). We predict that there will be greater genetic structure across the range of low dispersal species than for high dispersal species (H_o1). We expect that species of similar dispersal potential will have similar spatial patterns of genetic structure and differentiation (H_o2). We anticipate that similar patterns of genetic structure and differentiation will reflect the common influence of certain seascape features and processes (H_o3). Specifically, for low dispersal species we anticipate patterns of genetic differentiation will reflect the impact of historical seascape features predominantly, whereas we expect the impact of contemporaneous seascape processes will be more evident in the patterns of genetic differentiation of high dispersal species (H_o4). Last, we anticipate that all four species will have higher genetic diversity closest to the junction of the oceans based on an expected correlation between species diversity and genetic diversity (Vellend and Geber 2005; H_o5). Our approach tests the underlying assumption of many spatial genetic studies, that spatial genetic structure and genetic differentiation as compared across species reflect something meaningful about the physical processes that interact with demographic and evolutionary processes.

Methods

Study design

The four reef fish species we examined vary in two of the early life history traits predicted to affect dispersal potential: egg type and pelagic larval duration (PLD). We expected genetic structuring and genetic differentiation would be greater in the two low dispersal species: the domino damselfish (*Dascyllus trimaculatus*, benthic eggs, 23 PLD; lowest dispersal potential) and the neon damselfish (*Pomacentrus coelestis*, benthic eggs, 27 PLD); than the two higher dispersal species: the checkerboard wrasse (*Halichoeres hortulanus*, pelagic eggs, 30 PLD) and convict surgeonfish (*Acanthurus triostegus*, pelagic eggs, 60 PLD; highest dispersal potential). Previous genetic surveys for *D. trimaculatus* (Bernardi et al. 2001, Bernardi et al. 2003, Leray et al. 2010), *P. coelestis* (Gerlach et al. 2007, Liu et al. 2008, Liu et al. 2010, Liu et al. 2012, Liggins et al. in prep. – Chapter Five), and *H. hortulanus* (Drew and Barber 2012) are consistent with these expectations; however significant genetic structure has been reported over short distances for *A. triostegus* (Planes 1993, Planes et al. 1998, Fauvelot and Planes 2002).

The four fish species were collected from 19 locations using pole spears and hand nets while on SCUBA or snorkel (Table 1, Fig, 1). At seven of these locations all four species of fish were collected (ASH, TIM, LIZ, MOT, KAV, LIH). At the remaining locations species were only collected if they were present in high abundance. Given the extent of our sampling we attempted to avoid two caveats common to comparative phylogeographic studies. First, we focus our comparative analyses on locations where all four species have been co-sampled (hereafter co-sampled range), to avoid any affect of unbalanced sampling design (Dawson et al. in press). Second, for each species we examine genetic patterns over the widest possible range (hereafter broad-range), thus providing geographic context for the focal genetic patterns (Rocha et al. 2007, Liggins et al. 2013 – Chapter Three, Bowen et al. 2014).

Laboratory methods

Total genomic DNA was extracted from the collected tissues using a salt extraction method (modified from Aljanabi and Martinez 1997). For all species, we targeted the mitochondrial control region to infer genetic patterns. However, a suitable length of the control region proved difficult to amplify for *A. triostegus*, so we used ATPase subunit 6 and 8 amplified using ATP8.2 and CO3.2 as per Lessios and Robertson (2006). The mitochondrial control region of *D. trimaculatus* and *H. hortulanus* was amplified using CR-A and CR-E (Lee et al. 1995). Amplification protocols followed those detailed in Mirams et al. (2011 – Appendix One). Amplicons were purified using Exonuclease I and Antarctic Phosphatase following the Exo-SAP protocol (New England Biolabs) and sequenced by Macrogen (Korea) via capillary electrophoresis. Sequences were manually checked and edited in CodonCode Aligner v3.7.1.2 (CodonCode Corporation). Sequences of *D. trimaculatus* were combined with control region sequences already published by the authors (JF18156-JF18183; Appendix One), and all sequences for *P. coelestis* were taken from earlier publications by the authors (JF718094-JF718155,

Appendix One; KJ779110-KJ779112, KJ779115-KJ779168, KJ779175-KJ779243, KJ779296-KJ779325, KJ779358-KJ779376; Chapter Five). Sequences were aligned, trimmed, and translated into amino acid sequences using the vertebrate mitochondrial code to ensure they were not of nuclear origin using Se-Al v2.0a11 (Rambaut 1996). All new sequences are available on GenBank (*D. trimaculatus*, KJ779398-KJ779534, *H. hortulanus*, KJ779535-KJ779681; *A. triostegus*, KJ779749-KJ779871).

Genetic patterns

To observe the genealogical relationships among locations and infer phylogeographic structure we constructed minimum spanning networks for each species in PopART (http://popart.otago.ac.nz). Data files were then reduced to haplotype identity and Genodive's (v. 2.0b23, Miermans and Van Tienderen 2004) K-clustering method was used to determine how many significant groups (k) were found across each species broad-range, and co-sampled range. This analysis addressed whether there was greater genetic structure in low dispersal species, than in high dispersal species (H_o1). K-Means clustering divides populations into a number of groups (k) using an Analysis of Molecular Variance (AMOVA, Excoffier et al. 1992) - based, simulated annealing approach, without regard to their relative geographic position. Thus, this method simultaneously conducts a clustering analysis, and AMOVA, without requiring *a priori* designation of groups. From 2 to N-2 groups were selected using the Calinski-Harabasz (1974) pseudo-*F*-statistic with 100,000 steps and 20 repeats.

Pairwise relationships among populations were described using Φ_{ST} (based on the Tamura-Nei distance as selected by the Bayesian Information Criterion in jModeltest v. 2.1.4, Darriba et al. 2012, 10,000 permutations) and F_{ST} (based on haplotype identities in Arlequin 3.5, Excoffier and Lischer 2010). We also calculated D_{est} – an estimation of Jost's *D* (Jost 2008) using Genodive. All three measures were used as they each have the potential to reveal different characteristics of a dataset (see Discussion).

Congruence among species in genetic patterns

The similarity in the patterns of genetic differentiation and structure among species was tested using the congruence among distance matrices method (CADM, Legendre and Lapointe 2004) in ape v3.0-8 (Paradis et al. 2004; executed in R v2.15.3, R Core Team). This method is similar to the Mantel test, but can be used for two or more distance matrices. The overall coefficient of concordance reported is Kendall's *W*, a test statistic that varies from 0, indicating no congruence to 1, indicating high congruence (Kendall and Smith 1939; using the function CADM.global). Pairwise similarity among matrices within the comparison is then assessed *a posteriori* using Mantel's correlations (r_M, using the CADM.post function). The analysis was repeated for distance matrices based on each of the genetic measures in turn (k-clusters, F_{ST}, Φ_{ST}, and D_{est}), for the co-sampled range of all the species. Distance matrices for the k-clusters were binarised matrices based on group membership, where populations of the same cluster were assigned an inter-population distance of 0, and populations of different clusters

were assigned a distance of 1. No transformation of data to account for species or gene region differences was necessary as the analysis reduces pairwise values to a rank order prior to testing the coefficient of concordance through permutation (10,000 permutations). The Holm (1979) method was used to correct p–values following multiple testing (a less conservative method recommended by Legendre and Lapointe 2004). We had no prior expectation for the overall level of congruence among species for patterns of genetic differentiation and structure; however, we did expect that the *a posteriori* tests would reveal that species pairs that had the highest correlation, were of similar dispersal potential (H_o2).

We used multivariate analysis of variance (MANOVA) to analyze differences in genetic diversity and demographic history at the seven locations where all species were co-sampled (using the R base stats package). 'Location' and 'species' were used as factors, and π, $H_{\text{effective}}$, $H_{\text{diveristy}}$, Tajima's D, Fu's F_s, τ, and H_{private} were our response variables. We expected that values would differ among species, but that the patterns across species would be congruent. Specifically, we hypothesized that there would be a pattern of higher genetic diversity in locations closest to the junction of the Indian and Pacific Oceans (i.e. highest values in TIM, ASH, MOT), and that values for Tajima's D, Fu's F_s and τ would decrease away from the this region, indicative of more recent colonization, population expansion, and no population admixture (i.e. lowest values in KAV, SOL, LIH; H_o5).

Seascape predictors of genetic differentiation

We used multiple regression of distance matrices (MRDM; Legendre et al. 1994) to examine the association of historical and contemporary seascape features with patterns of genetic differentiation across the sampled ranges of the four fish species. The MRDM method has been identified as one of the best for predicting genetic differentiation (Balkenhol et al. 2009). This method appropriately facilitates the simultaneous analysis of geographic distances and barriers as predictors for pair-wise genetic distance matrices (Paquette and Lapointe 2009) and it has been demonstrated to be effective in removing the influence of historical covariates so that contemporary influences may also deduced (Dyer et al. 2010).

Several seascape features and processes were used to form predictor matrices in the multiple regression models, including: i) phylogeographic structure, ii) disjunction related to the Lydekker/Weber's line, iii) the Torres Strait, iv) contemporary least-cost oceanographic distances, and v) larval dispersal distances derived from species-specific biophysical models of dispersal. Phylogeographic structure (i) was only included in the model if haplotype networks revealed a distinct lineage was restricted to one or more geographically proximal locations. The rationale behind the incorporation of phylogeographic structure was not to test the effect of phylogeographic structure, but to remove its effect, so that predictor matrices relevant to the residual variance could subsequently be identified. Phylogeographic structure and the geographic boundaries (Lydekker/Weber's line, ii, and the Torres Strait, iii, Fig. 1) were represented as binarised matrices in the analysis (as described above).

The last two predictor matrices were weighted distances. First, contemporary least-cost oceanographic distances (iv) were estimated using marmap v0.5 in R (Pande and Simon-Bouhet 2013) by restricting the paths between locations to the habitable ocean (i.e. less than 35°S). Second, we used a biophysical model of larval dispersal (Treml et al. 2012) to quantify the relative dispersal strength among the co-sampled locations for all species. This dispersal model includes coral reef habitat (Spalding et al. 2001), oceanographic data describing sea surface currents for three years (ROMs, Wang et al. 2005), and several biological parameters describing the dispersal characteristics of each species, as follows: *D. trimaculatus*- benthic eggs, seasonal spawning, 23 day PLD, pre-competency period of ~1.5 days; *P. coelestis*- benthic eggs, annual spawning, 27 day PLD, pre-competency period of ~1.5 days; *H. hortulanus*- pelagic eggs, seasonal spawning, 30 day PLD, pre-competency period of ~1.5 days; *A. triostegus*- pelagic eggs, annual spawning, 60 day PLD, pre-competency period of ~5 days. The model outputs the probability that larvae released in one location survive and settle in every other recipient location, summarized as a 1002 × 1002 source-reef by destination-reef matrix. This dispersal probability matrix was converted to a migration matrix representing the proportion of settlers to every reef patch that came from all upstream larval sources (see Treml et al. 2012 for model details and sensitivity analysis). The migration matrix, M, was converted using $\log(M^{-1})$ to transform the values to be the same rank-order as geographic distance (high proportion of settlers then have a short distance) required for many network-based algorithms. This inverse dispersal strength matrix was used as a proxy for larval dispersal distance (v).

Several MRDM models were fitted for each species. First, models were constructed across the broad-range for each species. This was done separately for each of the pairwise genetic differentiation measures used (F_{ST}, Φ_{ST}, D_{est}). Second, models were constructed for the co-sampled range (again separately for each genetic differentiation measure). The latter was conducted so that models could be compared across species. Analyses proceeded by fitting a model via the stepwise approach (Legendre et al. 1994). Terms were added to the model only if their significance in the model did not exceed a Bonferroni corrected p–value ($\alpha = 0.05$). We anticipated that for species identified as having congruent patterns of genetic differentiation and structure (in the previous CADM analysis), their final MRDM models would attribute their patterns of genetic differentiation to the same seascape features (H_o3). We hypothesized that the effect of historic seascape features (ii, iii) would be most evident in the patterns of genetic differentiation for the low dispersal species (*D. trimaculatus* and *P. coelestis*); whereas, we expected that the effect of more contemporary seascape features (iv, v) would be most evident in the patterns of genetic differentiation of the higher dispersal species (*H. hortulanus* and *A. triostegus*; H_o4).

Results

Genetic patterns

The final datasets included sequences from: 165 individuals of *Dascyllus trimaculatus* (375bp, 12 locations); 237 *Pomacentrus coelestis* (336bp, 13 locations); 147 *Halichoeres hortulanus* (359bp, 10 locations); and 190 *Acanthurus triostegus* (830bp, from 10 locations; Table 1). The haplotype networks revealed that all four species had high levels of genetic diversity that was shared across many geographic locations (Fig. 2). Phylogeographic structure was evident in *A. triostegus*, equating to the suggested *A. triostegus sandvicensis* subspecies of Hawaii (HAW; Randall 1956, Planes and Fauvelot 2002). *D. trimaculatus* also had geographic and genetically distinct lineages in both the Chagos archipelago (CHA) and the Cook Islands (COK). In COK this distinct lineage coexisted with a common haplotype shared with other Pacific locations (MOT, LIZ, SOL, FIJ, TGA). These lineages likely correspond to the clades described by Leray et al. (2010; the Indian Ocean clade in CHA, and both the Tuamotu + Society Islands clade and the widespread Western Central Pacific clade in COK). Matrices representing the phylogeographic structure (i) of *A. triostegus* and *D. trimaculatus* were included in subsequent MRDM analyses for all of their sampled locations.

The geographic arrangement of k-clusters fell into two groups for the broad-range: the clusters for *A. triostegus* and *D. trimaculatus* corresponded to their described phylogeographic structures (k = 2, F_{CT} = 0.135, $P < 0.0001$; k = 3, F_{CT} = 0.050, P = 0.015); whereas for both *P. coelestis* and *H. hortulanus* only two clusters were identified, delineated by the Torres Strait (k = 2, F_{CT} = 0.037, P = 0.002; k = 2, F_{CT} = 0.033, $P < 0.0001$, Fig. 3). The number and geographic arrangement of clusters varied once the analysis was reduced to the co-sampled range for all species, except *P. coelestis* (k = 2, F_{CT} = 0.04, $P <$ 0.0001). Although *H. hortulanus* maintained two clusters, their geographic arrangement varied, so that the New Guinea landmass divided the geographic clusters (k = 2, F_{CT} = 0.016, $P < 0.0001$). *A triostegus* and *D. trimaculatus* had many more clusters and their geographic correlates were more difficult to infer qualitatively (k = 4, F_{CT} = 0.032, P = 0.005; k = 5, F_{CT} = 0.011, P = 0.018). The genetic variance and the number of clusters found in each species across the co-sampled locations, provided some support for our hypothesis that low dispersal species would have greater genetic structuring than the higher dispersal species (H_o1): *D. trimaculatus* had the highest number of significant clusters, and *P. coelestis* had the greatest variance described by clusters.

Congruence among species in genetic patterns

Quantitative analysis of genetic structure and differentiation (Appendix 1) across species further confirmed that there are large differences among species. Analyses based on the k-clusters for the co-sampled ranges produced an overall coefficient of concordance of W = 0.24 (X^2 = 19.48) and was not significant (adjusted P_{perm} = 0.5963). Furthermore, the *a posteriori* pairwise tests revealed that none of the patterns of genetic structuring were correlated among species. Analyses of congruence among distance matrices based on measures of genetic differentiation for all 4 species showed similar levels of non-significant concordance as the aforementioned test based on genetic clusters (F_{ST}: W =

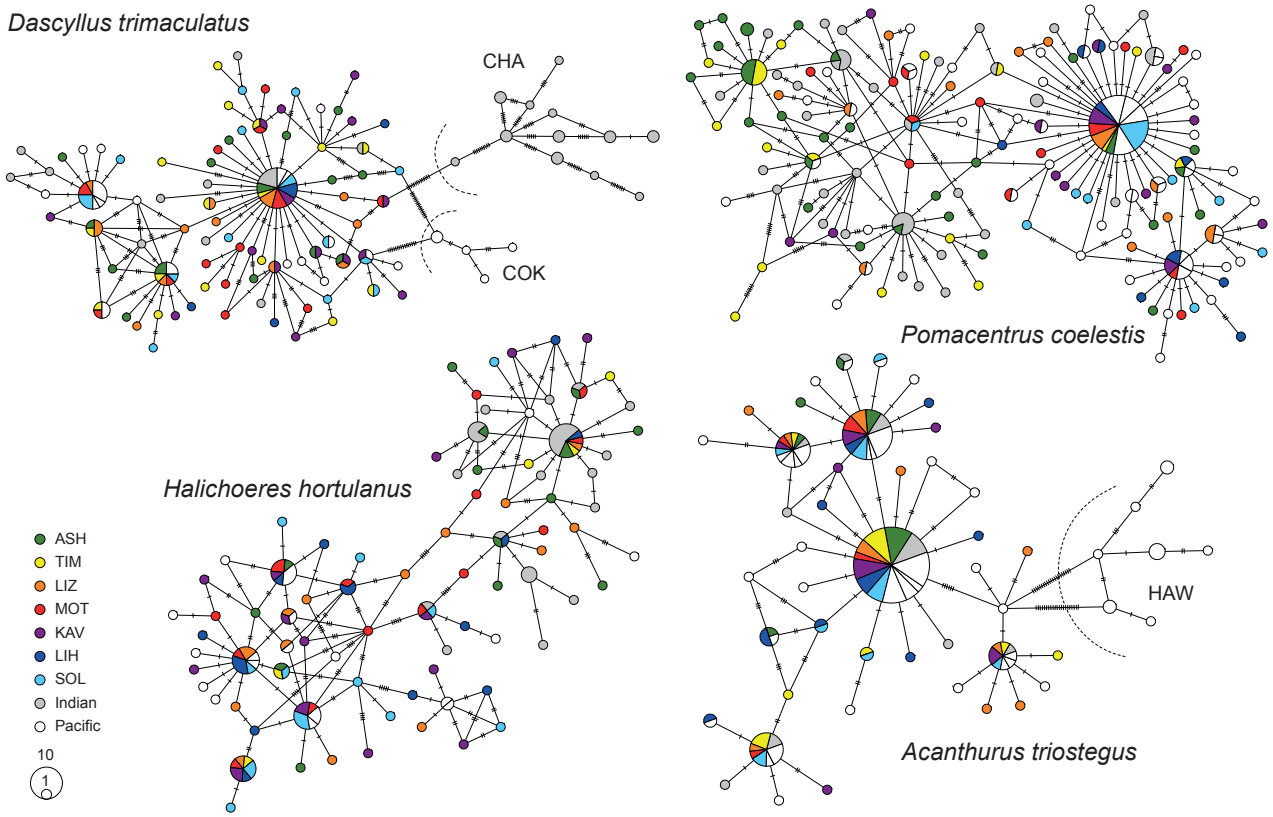


Figure 2. Minimum spanning haplotype networks for the study species. Frequency of haplotypes is indicated by the key (bottom left). Colors indicate location of origin (left). Only the seven populations for which species were co-sampled are represented in the color key, all other locations are represented as Indian Ocean (grey; including CHA, NIN, WIG) or Pacific Ocean (white; including HER, MOO, SLT, LDH, TUV, FIJ, TGA, COK, HAW). Dashed arcs delineate spatially segregated clades indicative of phylogeographic structure relevant to subsequent analyses (*Dascyllus trimaculatus* and *Acanthurus triostegus* only). Location codes are as listed in Table 1.

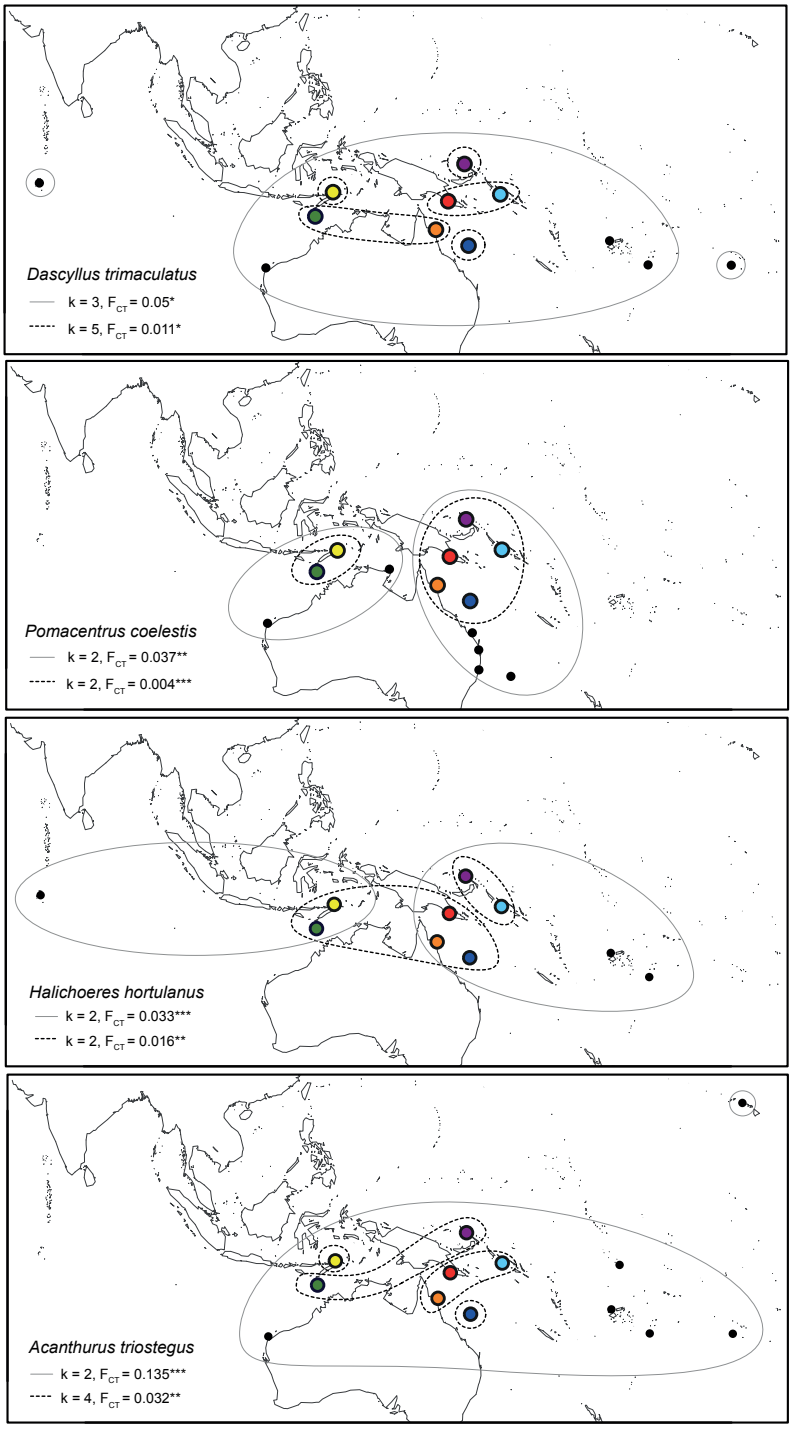


Figure 3. Map of sampled locations for each study species. Significant genetic clusters (k) identified across all the sampled locations are represented by the grey lines on each species map. Significant clusters based only on the co-sampled locations are represented by the black dotted lines on each species map. Significance of the clustering is denoted by asterisks: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

0.15, X^2 = 11.86, adjusted P_{perm} = 0.7830, Φ_{ST}: W = 0.22, X^2 = 17.82, adjusted P_{perm} = 0.6070, D_{est}: W = 0.23, X^2 = 18.59, adjusted P_{perm} = 0.5488). However, the *a posteriori* pairwise tests revealed that some species had significantly concordant patterns across the co-sampled range, and these were species of similar dispersal capacity. The highest Mantel's correlation was found for the geographic patterns in Φ_{ST} of the two mid-range dispersers, *H. hortulanus* and *P. coelestis* (r_M = 0.85, adjusted P_{perm} = 0.0018). This correlation was also evident when genetic differentiation was based on F_{ST}, although this was only marginally significant (r_M = 0.70, adjusted P_{perm} =0.0562). In contrast, for measures based on D_{est} the geographic patterns in *H. hortulanus* were significantly correlated with those of *A. triostegus,* the highest disperser (r_M = 0.54, adjusted P_{perm} = 0.0398). These findings provide moderate support for our hypothesis that species of similar dispersal potential would have similar spatial patterns of genetic differentiation (H_o2).

Measures reflecting genetic diversity and population demography (Fig. 4) tended to be higher in locations that were closest to the junction of the Indian and Pacific Oceans (ASH, TIM, MOT, LIZ). However, MANOVA revealed there was no effect of location on the raw, or scaled and centered, genetic response measures (Pillai's Trace = 1.67, $F_{6,36}$ = 1.15, P = 0.2825) but there was a significant effect of species (Pillai's Trace = 2.17, $F_{3,18}$ = 6.56, P < 0.0001). Most of the variables were suitably correlated for MANOVA (i.e. r = 0.20 - 0.60, Meyers et al. 2006), however $H_{diversity}$ was excluded from the MANOVA analysis as it was highly correlated with π and $H_{effective}$ (r > 0.75) and did not have a normal distribution following transformation. Some response measures were also negatively correlated with each other. This is expected given the way that the different genetic measures are derived, but can interfere with the performance of a MANOVA. Therefore, we additionally performed two-factor univariate analyses of variance. The univariate relationships confirmed that species had an effect on all the genetic measures, but also indicated that values for Tajima's D varied consistently across species for some locations ($F_{6,18}$ = 2.77, P = 0.0435). A Tukey's test for Honestly Significant Differences revealed that this result was driven by two significant pairwise relationships among locations; KAV had significantly lower values for Tajima's D than both LIZ and MOT (after Tukey's test: LIZ-KAV, diff = 2.076, P = 0.0140; MOT-KAV = 1.992, P = 0.0200). Other than this result, there was little quantitative support for our prediction that values of genetic diversity and demographic measures would decrease away from the junction of the Indian and Pacific Oceans (H_o5).

Seascape predictors of genetic differentiation

Many of the final MRDM models retained only one predictor matrix, and in some cases none of the predictor matrices were found to be associated with the patterns of genetic differentiation (Table 2). All correlations among predictor matrices were of an acceptable level (r < 0.72) and variance inflation factors (VIF) were inspected to ensure that the predictor variables did not exhibit multicollinearity that would affect model selection. The maximum VIF observed in the analyses was 1.78, well below the recommended guideline of < 5 (Rogerson 2001). Significant final models are presented below, species by species.

Table 2. Summary of the final models for the multiple regression of distance matrices analyses for each measure of genetic differentiation (F_{ST}, Φ_{ST}, D_{est}) for each species. The top panel presents the models for the seven populations that were co-sampled for all species; the bottom panel is for all sampled populations within each species (referred to as the broad-range in text). Models were chosen via the stepwise procedure, with a significance threshold of a Bonferroni corrected p–value ($\alpha = 0.05$). Beta coefficients are presented on a standardized scale. Factors not included in the final models are denoted by a dash (i.e. –). NA represents where a factor is not applicable to the dataset and was excluded, or when the final model contained only one predictor matrix and an adjusted p–value was not applicable.

	Dascyllus trimaculatus			***Pomacentrus coelestis***			***Halichoeres hortulanus***			***Acanthurus triostegus***		
Co-sampled locations												
Factors	F_{ST}	Φ_{ST}	D_{est}	F_{ST}	Φ_{ST}	D_{est}	F_{ST}	Φ_{ST}	D_{est}	F_{ST}	Φ_{ST}	D_{est}
ii) Lydekker / Weber's line	-	-	-	-	-	0.270***	-	-	-	-	0.426**	0.411***
iii) Torres Strait	-	-	-	-	-	0.753***	-	-	-	0.588***	-	0.485***
iv) Oceanographic distance	-	-	-	-	0.914***	-	-	0.855***	-	-	-	-
v) Dispersal distance	-	-	-	-	-	-	-	-	0.318*	-	-	-
F	-	-	-	-	202.60	91.42	-	108.30	4.51	21.13	8.87	27.40
R^2	-	-	-	-	0.835	0.824	-	0.730	0.101	0.346	0.182	0.584
P	-	-	-	-	NA	<0.0001	-	NA	NA	NA	NA	<0.0001
All sampled locations												
Factors	F_{ST}	Φ_{ST}	D_{est}	F_{ST}	Φ_{ST}	D_{est}	F_{ST}	Φ_{ST}	D_{est}	F_{ST}	Φ_{ST}	D_{est}
i) Phylogeographic structure	0.645***	0.855***	0.544***	NA	NA	NA	NA	NA	NA	-	0.995***	0.950***
ii) Lydekker / Weber's line	-	-	-	-	-	0.187**	-	-	-	-	0.037*	0.129***
iii) Torres Strait	-	-	-	-	0.762***	0.778***	-	0.285***	-	0.504***	-	-
iv) Oceanographic distance	-	0.157***	-	-	-	-	0.734***	0.750***	0.458**	-	-	-
F	45.56	407.00	26.92	-	105.20	92.61	50.14	165.7	11.4	25.87	2419.00	331.70
R^2	0.416	0.928	0.296	-	0.581	0.712	0.538	0.888	0.210	0.254	0.985	0.898
P	NA	<0.0001	NA	-	NA	<0.0001	NA	<0.0001	NA	NA	<0.0001	<0.0001

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

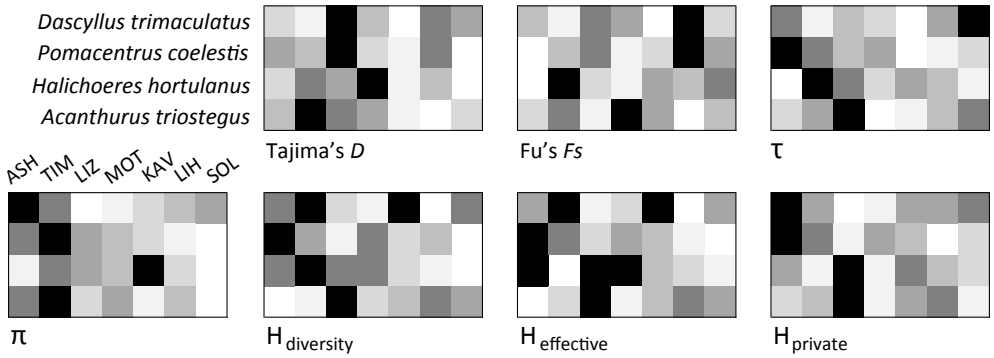


Figure 4. Matrices of species × location genetic diversity and demographic measures for the seven locations where all species were co-sampled. For each species and genetic measure combination, locations were ranked based on the value they had for that genetic measure. Shading represents the rank of that location's value within the species dataset: black is the highest value (e.g. location with the highest *Hd*); white is the lowest (e.g. location with the lowest *Hd*). Equal ranks share the same shade. Despite the highest ranked values seeming to be concentrated toward westward sites for most of the genetic measures, only values for Tajima's *D* were significantly differentiated across sites according to a multivariate analysis of variance (results discussed in text).

For *A. triostegus*, the effect of the Torres Strait (iii) on patterns of genetic differentiation defined by F_{ST} was significant across both the broad- and the co-sampled range (Table 2). Across the broad-range, phylogeographic structure (i) was significantly associated with the patterns of genetic differentiation defined by both Φ_{ST} and D_{est}, as expected. Disjunction associated with the Lydekker/Weber's line (ii) also had a significant effect on Φ_{ST} and D_{est}, over both the broad- and co-sampled range.

For *P. coelestis,* the effect of the Lydekker/Weber's line (ii) and the Torres Strait (iii) were apparent across both geographic scales when patterns of genetic differentiation were characterized using D_{est} (Table 2). When focusing on Φ_{ST}, only the Torres Strait (iii) was significant over the broad-range, and only oceanographic distance (iv) over the co-sampled range.

For *D. trimaculatus*, none of the predictor matrices were associated with the patterns of genetic differentiation across the co-sampled range. As expected, phylogeographic structure (i) was associated with patterns of genetic differentiation across the broad-range (for F_{ST}, Φ_{ST}, and D_{est}). For Φ_{ST}, oceanographic distance (iv) was also a predictor along with phylogeographic structure.

For *H. hortulanus*, oceanographic distance (v) was the best predictor across all of the final models, except for D_{est} across the co-sampled range where dispersal distance (v, Appendix 2) proved the strongest predictor. Over the broad-range of *H. hortulanus* the Torres Strait (iii) was also associated with the patterns of genetic differentiation defined by Φ_{ST}.

Discussion

Our study suggests that inference based on the observation of patterns of spatial genetic structure alone, can overlook the common role of some seascape features and processes in shaping the spatial genetic patterns of species. Although species' dispersal potential (based on egg type and PLD) did not reliably predict spatial genetic structure, nor which species share common spatial patterns of structure, we found congruent spatial patterns based on genetic differentiation among species of similar dispersal potential. Furthermore, our analytical framework focused on the processes underlying spatial genetic patterns identified seascape features that have a common influence across species despite their disparate spatial genetic patterns.

Congruent and incongruent genetic patterns for four coral reef fishes

We found judicious support for the idea that species of low dispersal potential have greater population genetic structure (H_o1). The species likely to have the lowest dispersal potential (based on egg type and PLD), *Dascyllus trimaculatus*, had the greatest number of significant clusters (k = 5) identified across the co-sampled range; similarly *Pomacentrus coelestis* the second lowest disperser had the greatest amount of genetic variance attributed to clusters (F_{CT} = 0.04). However, simultaneously considering all four species, there was no consistent association between early life

history dispersal traits and the number of significant clusters or the amount of genetic variance explained. Furthermore, the spatial genetic patterns of clustering sufficiently differed among all species, so that there were no significant correlations (using the congruence among distance matrices method, CADM, Legendre and Lapointe 2004, $W = 0.24$, adjusted $P_{perm} > 0.05$). These results, in combination with our extended inference detailed below, highlight that relying on spatial patterns of genetic structuring alone may overlook physical processes that have a common role in determining patterns of genetic differentiation for co-distributed species (also suggested by Pelc et al. 2009).

When we shift our focus from spatial patterns of genetic structure, to spatial patterns of genetic differentiation (based on F_{ST}, Φ_{ST} and D_{est}), we find instances of concordance across species of similar dispersal potential (H_o2). Patterns of spatial differentiation were significantly correlated for two species pairs (adjusted $P_{perm} < 0.05$; based on CADM): 1) *H. hortulanus* and *P. coelestis*, our two mid-range dispersers with the same average PLD (for Φ_{ST}); and 2) *H. hortulanus* and *A. triostegus*, our two high dispersal species that have pelagic eggs (for D_{est}). For our first pair, the final models derived using multiple regression of distance matrices (MRDM, Legendre et al. 1994) also provided support for our hypothesis that similar seascape features and processes shape patterns of genetic differentiation in species of similar dispersal potential (H_o3). For both *H. hortulanus* and *P. coelestis*, patterns in Φ_{ST} were highly associated with oceanographic distance (Table 2). In contrast, for *A. triostegus* and *H. hortulanus*, we were unable to derive any seascape features or processes responsible for the correlation in their patterns of genetic differentiation based on D_{est}. There are two likely reasons for this result. First, it is possible that whereas their ranked patterns within the congruence among distance matrices (CADM) analysis were correlated, the actual values for *A. triostegus* and *H. hortulanus* at equal ranks were quite disparate. Thus, in the more quantitative analysis afforded by the MRDM analyses we could identify the Torres Strait and Lydekker/Weber's line as the main drivers of the genetic patterns for *A. triostegus* patterns and the biophysically derived dispersal distance for *H. hortulanus* (Table 2). Alternatively, the similarity in the spatial patterns of genetic differentiation for *A. triostegus* and *H. hortulanus* may be driven by a seascape feature or process, other than those included in our models.

Patterns of genetic differentiation for our high dispersal species were not consistently associated with contemporary processes (i.e. larval dispersal distance and oceanographic distance); nor were the patterns of genetic differentiation of our low dispersal species consistently associated with historical processes (i.e. the Lydekker/Weber's line, Torres Strait; H_o4). Although the best fit MRDM models for *H. hortulanus*, our second highest disperser, consistently contained oceanographic distance and dispersal distance on one occasion, models for *A. triostegus* were most affected by the Torres Strait and the Lydekker/Weber's line (Table 2). These same historic seascape features were predicted to be most important in driving patterns of genetic differentiation in *P. coelestis* (based on D_{est}), one of our lowest dispersers.

Lastly, although within-species ranked values of genetic diversity seemingly illustrate patterns across geography (Fig. 4), multivariate analysis (MANOVA) suggest there were no differences among statistical estimates of genetic diversity across the co-sampled study region ($P > 0.05$). The single exception was Tajima's *D*. Tajima's *D* behaved as we predicted: significantly lower values were found at a location most distant from the junction of the Indian and Pacific Oceans (H_o5). The low values for Tajima's *D* suggest that all four reef fish species have a relatively recent and isolated population history in Kavieng. Few studies have included Kavieng in their genetic surveys. Although Barber et al. (2006) found a divergent linage that was shared with the northern part of mainland Papua New Guinea in three stomatopod species, only one of these species had a signature of population expansion.

The role of the ocean in driving genetic differentiation

The existence of boundaries to dispersal in a fluid environment is non-intuitive, yet our study joins the growing number of marine focused studies that underline the roles of straits, inundated continental shelves, and oceanographic distance in driving patterns of genetic differentiation in reef fishes (and even pelagic species, see Ackiss et al. 2013; reviewed in Riginos and Liggins 2013 – Chapter Two). Here, we identified several cases of allopatric divergence, whereby peripheral populations are highly divergent from the rest of the species sampled range. Specifically, we further detailed the presence of a distinct lineage of *A. triostegus* in Hawaii and distinct lineages of *D. trimaculatus* in the Chagos Archipelago and the Cook Islands. Such patterns are indicative of historical colonization, with subsequently reduced contemporary connectivity. In the case of the Cook Islands, where we see the occurrence of two different lineages, two temporally independent colonization events are likely (other instances of sympatric lineages have previously been reported in *D. trimaculatus*, Bernardi et al. 2001, 2003, Leray et al. 2010). Similar patterns of extreme peripheral divergence have been observed in Indo-Pacific snappers in the Eastern Pacific (Gaither et al. 2009), and five coral reef fishes in Fiji (Drew et al. 2008). In the study of Drew et al. (2008) Fijian populations were all genetically distinct from those in within the IAA. Thus, despite the pervasive focus on the IAA as a source and reservoir of genetic diversity (Bellwood et al. 2012), such peripheral locations also warrant consideration for the unique evolutionary arenas that they provide for arriving lineages.

In our study there was a trade-off in focusing on the co-sampled geographic area of our study species, versus the wider geographic sampling available for each species. Greater analytical rigor, and the ability to directly compare species were afforded by a focus on the co-sampled locations (discussed in Dawson 2014, Dawson et al. in press). However, the relevance of these inferred patterns for each species was better understood within the context of the wider geographic range (as suggested by Rocha et al. 2007, Chapter Three, Bowen et al. 2014). Our inferences of what seascape features and processes determine patterns of genetic differentiation for *A. triostegus* were largely unchanged, other than the obvious inclusion of phylogeographic structure in the final models over the broad-range. However, a shift was observed in the final models of *P. coelestis*, *H. hortulanus* and *D. trimaculatus* related to study scale.

The change in the relative importance of seascape features or processes over geographic scale may be indicative of different patterns of genetic differentiation across a species range (Slatkin 1993). For *P. coelestis*, where the most significant predictor of genetic differentiation based on Φ_{ST} changed from oceanographic distance over the co-sampled range, to the Torres Strait over the broad-range, a distinct lack of the an IBD relationship in the species southern range periphery is likely the cause (including MOO, SLT, LDH; Chapter Five). A similar shift in the role of the Torres Strait in driving patterns of genetic differentiation is evident in the final models for Φ_{ST} in *H. hortulanus*. Whereas there is a clear indication that distance (oceanographic and/or dispersal) is the most important driver of genetic differentiation patterns in *H. hortulanus* (across all metrics and geographic scales), the Torres Strait is also recognized to have had a role in the genetic differentiation of populations across the broad-range based on Φ_{ST}. In contrast, whereas none of our spatial predictor matrices are associated with the patterns of genetic differentiation across the co-sampled range of *D. trimaculatus*, oceanographic distance is associated with the patterns defined by Φ_{ST} over the broad-range. Thus, while a pattern of "chaotic-genetic patchiness" (Johnson and Black 1982) may typify the genetic differentiation of *D. trimaculatus* over smaller scales, at large range-wide scales genetic patterns have some spatial relevance.

In many cases the final models changed according to which genetic differentiation measure was used, and in doing so, provided a nuanced view of the processes driving the genetic patterns. For example, the effect of oceanographic distance was not observed when patterns of genetic differentiation were based on D_{est} or F_{ST} (except for *H. hortulanus*, Table 2). This may indicate that in some cases the turnover of haplotypes is too high among populations over the vast oceanographic distances for F_{ST} or D_{est} to be informative. In contrast, Φ_{ST} can be informative, even when no haplotypes are shared among populations (i.e. turnover = 1), and thus may be especially informative in wide-ranging species such as Indo-Pacific reef fishes (Briggs and Bowen 2012).

Patterns as described by Φ_{ST} more closely resemble changes in nucleotide diversity and thus reflect the processes of mutation accumulation and drift. In instances of contemporary migration however, F_{ST} can capture variance in haplotype frequencies without the "noise" introduced by considering the genealogical relationships among haplotypes (discussed in Bird et al. 2011). Based on these expectations, the final models for *A. triostegus* suggest that a change in nucleotide diversity is associated with the Lydekker/Weber's line (evident in models based on Φ_{ST}); whereas differentiation associated with the Torres Strait is indicative of more contemporary migration limitation (indicated by F_{ST}). Furthermore, given that D_{est} is more robust in situations of high genetic diversity that are known to compromise the performance of fixation indices (Φ_{ST} and others in the F_{ST} family, Hedrick 2005, Jost 2008) – in the case of *P. coelestis*, where the effect of the Lydekker/Weber's line is only recognized in the models for D_{est} – it is likely that the high levels of genetic diversity found within the TIM population, and other neighboring populations (i.e. ASH, NIN) mask the ability of these fixation indices to detect differentiation coincident with this biogeographic boundary (Fig. 4).

Unlike the Torres Strait that has historically been a barrier and continues to constrain connectivity between the Indian and Pacific Oceans over contemporary timescales, the influence of the Lydekker/Weber's line on demographic processes is less intuitive. The patterns described above (genetic differentiation identified based on Φ_{ST} in *A. triostegus*, and D_{est} in *P. coelestis* only) and the observed patterns in genetic diversity (Table 1, Fig. 4, although not statistically significant) suggest that rather than being a boundary as such, the line reflects a transition, or steep gradient in diversity. Such a scenario would imply that the Sunda Shelf side of the line has been a location of longer population stability and/or lineage admixture (also suggested by Barber et al. 2006). Although well-studied at the species level, such nested transitions in diversity have rarely been characterized at the genetic level despite appropriate measures being available (see Diniz-Filho et al. 2012, Chapter Five).

Contrary to results from other recent studies that have included biophysical model predictions (e.g. White et al. 2010, Crandall et al. 2012; reviewed in Liggins et al. 2013 – Chapter Three), species-specific larval dispersal distances were not a significant predictor of mtDNA genetic differentiation patterns for our study species (also found in Crandall et al. 2014 – Appendix Two). The use of more variable, multi-locus markers may have provided greater sensitivity and power to detect an effect of contemporary ocean currents on genetic patterns. In the case of our study, it is likely the tumultuous evolutionary history of the lineages across the IAA has preempted the influence of contemporary seascape features, and in particular the combined effect of present-day ocean currents and species dispersal ability. It has been noted that the effect of historical factors can have a lasting effect on patterns of genetic structure, long past the persistence of the physical boundary (coined the "ghost of dispersal past", Benzie 1999). The mechanisms that allow such historical patterns to persist include those that are passive such as low migration (Hellberg et al. 2002) and incomplete lineage sorting, and those that are deterministic. These include the pre-emption of space by certain lineages (Waters et al. 2013) that may in turn have a competitive edge over other lineages that arrive. This competitive edge can be due to the adaptation of lineages to the local environment (Choat 2006, Marshall et al. 2010), or simply due to the higher fitness of locally derived larvae due to better provisioning in higher quality environments, or their lower cost dispersal experience in the pelagic environment (Hamilton 2008, Shima and Swearer 2010).

Other predictors of spatial genetic patterns in the sea

Ultimately testing the role of early life history traits in determining which seascape features and processes are important in shaping spatial genetic patterns, requires the analysis of many more species than included here. With the addition of more species, suites of responses to oceanographic features or spatial patterns across species might emerge (as seen in Pelc. et al. 2009, and Selkoe et al. 2014, respectively). However, a focus solely on egg type and PLD may be overly simplistic. For instance, the dispersal of pelagic larvae is also known to be influenced by species-specific traits such as their sense of smell, homing behavior, and swimming ability (to name a few; Leis et al. 2007). Alternatively, spatial genetic patterns may be unrelated to early life history dispersal traits (as

suggested by the studies of (Lacson 1992, Shulman and Bermingham 1995, Bird et al. 2007, Galarza et al. 2009, for example).

Comparative spatial genetic patterns of benthic marine organisms have been studied in relation to other species-specific traits such as habitat specificity, diet and ecology. Once again, whereas some studies find support for a role of distinct species-related traits (diet specialization, Lawton et al. 2011; species ecology or habitat specialization; Ayre et al. 2009, Rocha et al. 2002, Reid et al. 2006) others do not (species ecology, Crandall et al. 2008). Thus, it is likely that the primarily stochastic processes of colonization, population size fluctuation, and extinction, also have a large role in determining spatial genetic patterns in benthic marine organisms, and cannot be easily predicted across species (also suggested by Selkoe et al. 2014, but see Hart and Marko 2010 and Dawson et al. in press, for analytical suggestions).

Conclusions

Our study contributes several species replicates aimed at understanding the importance of historical and contemporary processes in shaping genetic patterns across the juncture of the Indian and Pacific Oceans, an approach advocated by others (Avise 2000, Dawson 2014, Bowen et al. 2014). Moreover, our study promotes an analytical framework focused on identifying seascape features and processes that have a common influence across species, rather than focusing on concordant spatial genetic patterns *per se*. Certainly the aim of many comparative genetic studies is to identify spatial patterns, not the underlying processes; the spatial delineation of regions of high genetic turnover and high genetic diversity versus low genetic diversity is very important for the prioritization of regions for conservation purposes (Moritz and Faith 1998, Rocha et al. 2007). However, an understanding of the processes that have lead to such patterns is also important.

Acknowledgements

All fish sampling was undertaken with the authority of The University of Queensland Animal Ethics Committee (Approval Number: SIB/817/08/ARC). Sampling in Timor-Leste was supported by the Coral Triangle Support Partnership and the Ministério da Aquicultura e Pescas, Direcção Nacional de Pescas e Aquicultura (authorized by A Fernandes, L Fontes, J Freitas; guia de marssa: 502/DNPA/VIII/10 and 452/DNPA/VII/11). Export of samples was authorized by the Departmento de Quarentena das Pescas (export permit: 162/FQ006/EXP./DNQB/VII/2011). Sampling in the Solomon Islands was via the Australian Government's Pacific Strategy Assistance Program and with the assistance of the Roviana Conservation Foundation (Solomon Islands Government Ministry of Education and Human Resource Development and Ministry of Fisheries and Marine Resources research permit to S Albert). Sampling in Papua New Guinea was in coordination with the National

Research Institute, the Department of Foreign Affairs and Immigration (Research Visa: 10350008304) and the Department of Environment and Conservation (Permit to Export Wildlife: 011318). Authority to sample at Ashmore Reef was provided by the Australian Government Department of Sustainability, Environment, Water, Population and Communities (Access to Biological Resources in a Commonwealth Area for Non-Commercial Purposes permit number: AU-COM2010068) and with logistic support from Australian Customs and border control. Sampling in the Coral Sea was supported by the Marine Division of the Australian Government Department of Sustainability, Environment, Water, Population and Communities (Access to Biological Resources in a Commonwealth Area for Non-Commercial Purposes permit number: AU-COM2008042). Authority to sample at Wigram Island was provided by the Northern Territory Government Department of Resources (Special Permit Number: 2007-2008/S17/2696). Sampling at Ningaloo Reef was under the authority of the Western Australia Department of Environment and Conservation (License to take Fauna for Scientific Purposes: SF007126, SF006619; Authority to enter calm land/or waters: CE002227, CE002627). We are grateful to DJ Booth and W Figueira for providing collection from New South Wales. We are grateful to B Bowen for providing samples from Oahu and the Cook Islands. Sampling in Tonga was under the authority of the Ministry of Agriculture and Food, Forests, and Fisheries (permit issued to J Drew). Sampling in Tuvalu was under the authority of the Ministry of Natural Resources and Environment (permit issued to J Drew). Sampling in Fiji was under the authority of the Government of Fiji with a Ministry of Primary Industries Fisheries Department permit: C564/2009. For sampling on the Chagos Archipelago, we thank M Gaither, the California Academy of Sciences, The Darwin Foundation, The Chagos Conservation Trust, C Sheppard, J Turner, and D Wagner. We are grateful to the staff of the Australian Museum Lizard Island Research Station and Heron Island Research Station for their facilities and support (Great Barrier Reef Marine Park Authority and Queensland Parks and Wildlife Marine Parks Permit: G08/28114.1, G09/31678.1, G10/33597.1, G11/34640.1; Queensland Government Department of Primary Industries General Fisheries Permit: 118636, 150981; Australian Quarantine Inspection Service Permit to Import Quarantine Material: IP10017966). We especially thank JD Aguirre, J Aini (Ailan Awareness), S Albert, K Davis, M Jimuru, J Keyse, J Kinch (National Fisheries College, Papua New Guinea), W Lovell (Freeflow Dive, Dili), I McLeod, A Mirams, S Penny, R Pinto (and staff of the Coral Triangle Support Partnership), A Smith (Tiki2 Adventure Tours), T Sinclair-Taylor, A Turner, P Waldie, MX Weber, Stephen, Lavud, and Takenda for logistical support and field assistance. Funding for this work was provided by the Australian Research Council (DP0878306, to CR) and an Explorer's Club Exploration Fund (to LL). LL was supported by an Australian Postgraduate Award from the Australian Government and a Queensland Government Smart Futures PhD Scholarship. Many of the ideas discussed here grew out of work funded by the Sea World Research and Rescue Foundation (SWR/1/2012, to CR and LL), and a Paddy Pallin Foundation and The Foundation for National Parks and Wildlife Science Grant (to LL).

Appendix 1. Pairwise genetic differentiation (Φ_{ST}, F_{ST}, D_{est}) among study locations for all species. Location codes are listed in Table 1. Pairwise F_{ST} (lower triangle) and corresponding p–values (upper triangle); pairwise Φ_{ST} (lower triangle, Tamura-Nei corrected) and corresponding p–values (upper triangle); and pairwise D_{est} (lower triangle). All p–values are following Benjamini and Hochberg correction.

Dascyllus trimaculatus

F_{ST}	CHA	NIN	ASH	TIM	LIZ	MOT	KAV	LIH	SOL	FIJ	TGA	COK
CHA	-	0.021	0.024	0.034	0.021	0.021	0.032	0.072	0.024	0.287	0.024	0.021
NIN	0.048	-	0.027	0.077	0.032	0.026	0.077	0.092	0.027	0.368	0.053	0.027
ASH	0.027	0.036	-	0.287	0.027	0.027	0.287	0.107	0.092	0.441	0.070	0.024
TIM	0.020	0.028	0.007	-	0.043	0.147	1.000	0.101	0.287	1.000	0.077	0.027
LIZ	0.039	0.048	0.026	0.019	-	0.027	0.040	0.088	0.027	0.384	0.053	0.024
MOT	0.036	0.045	0.024	0.017	0.036	-	0.147	0.077	0.027	0.397	0.048	0.024
KAV	0.020	0.028	0.007	0.000	0.019	0.017	-	0.101	0.287	1.000	0.077	0.027
LIH	0.051	0.061	0.037	0.029	0.051	0.048	0.029	-	0.101	0.469	0.120	0.070
SOL	0.027	0.036	0.015	0.007	0.026	0.024	0.007	0.037	-	0.441	0.070	0.024
FIJ	0.023	0.033	0.008	0.000	0.022	0.019	0.000	0.035	0.008	-	0.428	0.167
TGA	0.029	0.038	0.017	0.009	0.029	0.026	0.009	0.040	0.017	0.011	-	0.026
COK	0.097	0.108	0.085	0.077	0.098	0.095	0.077	0.121	0.085	0.093	0.088	-

Φ_{ST}	CHA	NIN	ASH	TIM	LIZ	MOT	KAV	LIH	SOL	FIJ	TGA	COK
CHA	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NIN	0.743	-	0.461	0.354	0.496	0.590	0.318	0.550	0.816	0.872	0.295	0.000
ASH	0.738	0.013	-	0.822	0.970	0.636	0.539	0.872	0.763	0.816	0.307	0.000
TIM	0.746	0.025	-0.016	-	0.816	0.872	0.872	0.931	0.592	0.875	0.082	0.000
LIZ	0.749	0.012	-0.036	-0.017	-	0.794	0.577	0.816	0.645	0.751	0.295	0.000
MOT	0.752	0.001	-0.003	-0.019	-0.016	-	0.822	0.745	0.636	0.906	0.154	0.000
KAV	0.747	0.027	0.005	-0.015	0.005	-0.015	-	0.872	0.417	0.970	0.023	0.000
LIH	0.722	0.016	-0.040	-0.048	-0.033	-0.018	-0.027	-	0.577	0.875	0.295	0.010
SOL	0.743	-0.017	-0.012	0.000	-0.006	-0.007	0.018	0.005	-	0.970	0.534	0.000
FIJ	0.723	-0.058	-0.039	-0.043	-0.029	-0.063	-0.063	-0.050	-0.085	-	0.511	0.080
TGA	0.733	0.043	0.034	0.082	0.045	0.070	0.127	0.075	0.010	0.041	-	0.000
COK	0.654	0.465	0.468	0.488	0.492	0.501	0.499	0.416	0.453	0.384	0.416	-

D_{est}	CHA	NIN	ASH	TIM	LIZ	MOT	KAV	LIH	SOL	FIJ	TGA	COK
CHA	-											
NIN	1.000	-										
ASH	1.000	0.125	-									
TIM	1.000	0.269	-1.495	-								
LIZ	1.000	-0.120	-0.637	-0.535	-							
MOT	1.000	-0.106	-0.225	-0.406	-0.400	-						
KAV	1.000	0.269	-0.996	0.000	-0.096	-0.172	-					
LIH	1.000	-0.622	-0.003	0.336	-0.293	-0.275	0.004	-				
SOL	1.000	0.125	-0.412	-0.996	-0.339	-0.685	-0.497	-0.003	-			
FIJ	1.000	-0.733	-1.960	0.000	-1.600	-2.475	0.000	-0.960	-7.880	-		
TGA	1.000	0.417	-0.395	-1.192	0.067	-0.431	0.123	0.197	-0.395	-1.600	-	
COK	1.000	1.000	1.000	1.000	0.706	0.434	1.000	1.000	0.409	-0.013	0.675	-

Appendix 1. continued

Pomacentrus coelestis

F_{ST}	NIN	ASH	TIM	WIG	LIZ	MOT	KAV	LIH	HER	MOO	SLT	SOL	LDH
NIN	-	0.027	0.017	0.001	0.004	0.026	0.006	0.012	0.011	0.006	0.010	0.002	0.004
ASH	0.010	-	0.004	0.000	0.002	0.011	0.002	0.011	0.002	0.002	0.004	0.000	0.000
TIM	0.019	0.025	-	0.001	0.004	0.018	0.004	0.017	0.006	0.008	0.010	0.002	0.002
WIG	0.054	0.060	0.071	-	0.003	0.004	0.000	0.008	0.002	0.000	0.003	0.000	0.001
LIZ	0.030	0.036	0.047	0.083	-	0.014	0.004	0.018	0.004	0.005	0.007	0.002	0.003
MOT	0.016	0.022	0.032	0.069	0.044	-	0.011	0.024	0.016	0.011	0.013	0.004	0.008
KAV	0.027	0.033	0.043	0.080	0.055	0.040	-	0.022	0.004	0.004	0.007	0.002	0.002
LIH	0.019	0.026	0.036	0.074	0.048	0.032	0.044	-	0.010	0.018	0.025	0.003	0.008
HER	0.028	0.034	0.044	0.080	0.055	0.041	0.052	0.045	-	0.003	0.006	0.000	0.002
MOO	0.022	0.028	0.038	0.074	0.050	0.035	0.046	0.039	0.047	-	0.006	0.001	0.002
SLT	0.034	0.040	0.050	0.087	0.062	0.047	0.059	0.052	0.059	0.053	-	0.003	0.005
SOL	0.086	0.091	0.104	0.142	0.117	0.102	0.113	0.109	0.112	0.107	0.121	-	0.001
LDH	0.037	0.043	0.053	0.090	0.065	0.051	0.062	0.055	0.062	0.056	0.069	0.123	-

Φ_{ST}	NIN	ASH	TIM	WIG	LIZ	MOT	KAV	LIH	HER	MOO	SLT	SOL	LDH
NIN	-	0.016	0.032	0.000	0.068	0.154	0.003	0.000	0.000	0.023	0.009	0.000	0.008
ASH	0.067	-	0.748	0.003	0.004	0.007	0.000	0.000	0.000	0.001	0.000	0.000	0.001
TIM	0.059	-0.010	-	0.003	0.003	0.005	0.001	0.000	0.000	0.001	0.000	0.000	0.001
WIG	0.174	0.129	0.112	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LIZ	0.051	0.145	0.146	0.321	-	0.905	0.674	0.173	0.270	0.953	0.748	0.186	0.905
MOT	0.030	0.134	0.137	0.342	-0.028	-	0.646	0.037	0.162	0.953	0.703	0.154	0.715
KAV	0.110	0.235	0.226	0.445	-0.007	-0.005	-	0.196	0.846	0.846	0.826	0.748	0.905
LIH	0.226	0.312	0.304	0.543	0.045	0.104	0.033	-	0.239	0.199	0.082	0.098	0.279
HER	0.159	0.264	0.276	0.496	0.019	0.032	-0.015	0.027	-	0.534	0.697	0.846	0.731
MOO	0.063	0.165	0.169	0.349	-0.031	-0.030	-0.017	0.032	0.001	-	0.798	0.321	0.905
SLT	0.103	0.217	0.223	0.442	-0.015	-0.009	-0.015	0.062	-0.005	-0.015	-	0.847	0.826
SOL	0.178	0.290	0.296	0.548	0.035	0.040	-0.010	0.069	-0.015	0.012	-0.018	-	0.479
LDH	0.098	0.198	0.201	0.410	-0.030	-0.011	-0.025	0.024	-0.010	-0.024	-0.019	0.004	-

D_{est}	NIN	ASH	TIM	WIG	LIZ	MOT	KAV	LIH	HER	MOO	SLT	SOL	LDH
NIN	-												
ASH	0.869	-											
TIM	0.886	-0.161	-										
WIG	0.859	0.746	1.000	-									
LIZ	1.000	0.576	1.000	1.000	-								
MOT	0.826	0.438	1.000	1.000	-0.150	-							
KAV	1.000	0.562	1.000	1.000	-0.072	-0.370	-						
LIH	1.000	0.452	0.866	1.000	0.043	-0.496	-0.578	-					
HER	0.926	0.497	0.940	1.000	-0.127	-0.310	-0.300	-0.323	-				
MOO	1.000	0.479	0.927	1.000	-0.194	-0.613	-0.341	-0.607	-0.350	-			
SLT	1.000	0.559	1.000	1.000	-0.366	-0.311	-0.146	-0.013	-0.206	-0.269	-		
SOL	0.971	0.710	1.000	1.000	0.063	0.101	0.090	0.255	0.036	0.144	-0.035	-	
LDH	1.000	0.563	0.945	1.000	-0.226	-0.250	-0.271	-0.391	-0.205	-0.244	-0.209	0.014	-

Appendix 1. continued

Halichoeres hortulanus

F_{ST}	CHA	ASH	TIM	LIZ	MOT	KAV	LIH	SOL	FIJ	TGA
CHA	-	0.029	0.174	0.027	0.027	0.025	0.024	0.024	0.174	0.024
ASH	0.052	-	0.628	0.353	0.353	0.174	0.174	0.128	0.628	0.128
TIM	0.055	0.005	-	0.628	0.628	0.507	0.628	0.469	1.000	0.469
LIZ	0.052	0.010	0.005	-	0.353	0.174	0.174	0.128	0.628	0.128
MOT	0.052	0.010	0.005	0.010	-	0.174	0.174	0.128	0.628	0.128
KAV	0.056	0.013	0.010	0.013	0.013	-	0.128	0.053	0.507	0.098
LIH	0.060	0.017	0.014	0.017	0.017	0.021	-	0.053	0.628	0.098
SOL	0.066	0.024	0.022	0.024	0.024	0.027	0.032	-	0.469	0.040
FIJ	0.055	0.005	0.000	0.005	0.005	0.010	0.014	0.022	-	0.469
TGA	0.064	0.021	0.019	0.021	0.021	0.025	0.029	0.036	0.019	-

Φ_{ST}	CHA	ASH	TIM	LIZ	MOT	KAV	LIH	SOL	FIJ	TGA
CHA	-	0.032	0.079	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ASH	0.105	-	0.953	0.096	0.274	0.033	0.027	0.001	0.026	0.006
TIM	0.164	-0.087	-	0.347	0.546	0.347	0.192	0.061	0.214	0.053
LIZ	0.433	0.119	0.050	-	0.849	0.953	0.918	0.397	0.532	0.446
MOT	0.351	0.046	-0.008	-0.026	-	0.476	0.446	0.098	0.214	0.154
KAV	0.463	0.162	0.063	-0.036	0.006	-	0.985	0.925	0.918	0.796
LIH	0.513	0.200	0.123	-0.029	0.012	-0.044	-	0.773	0.953	0.953
SOL	0.591	0.293	0.215	0.017	0.088	-0.028	-0.013	-	0.925	0.576
FIJ	0.681	0.338	0.259	0.001	0.108	-0.051	-0.076	-0.051	-	0.985
TGA	0.590	0.287	0.231	0.010	0.072	-0.015	-0.035	0.000	-0.089	-

D_{est}	CHA	ASH	TIM	LIZ	MOT	KAV	LIH	SOL	FIJ	TGA
CHA	-									
ASH	0.009	-								
TIM	-0.107	-6.800	-							
LIZ	0.640	0.067	-4.200	-						
MOT	0.549	-1.333	-4.200	-0.867	-					
KAV	0.960	0.682	-1.780	0.046	-1.226	-				
LIH	0.669	0.035	-0.853	-0.929	-0.929	0.438	-			
SOL	0.965	0.813	-0.950	0.253	-0.307	-0.675	0.339	-		
FIJ	1.000	1.000	0.000	-6.800	-4.200	-3.170	-1.780	-1.600	-	
TGA	1.000	0.417	1.000	0.417	-0.748	-0.094	0.196	0.183	-1.780	-

Appendix 1. continued

Acanthurus triostegus

F_{ST}	NIN	ASH	TIM	LIZ	MOT	KAV	LIH	SOL	TUV	FIJ	TGA	COK	HAW
NIN	-	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.013	0.000	0.002
ASH	0.227	-	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000
TIM	0.195	0.262	-	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.001
LIZ	0.125	0.190	0.157	-	0.003	0.000	0.001	0.000	0.000	0.001	0.022	0.000	0.005
MOT	0.175	0.251	0.212	0.132	-	0.002	0.003	0.001	0.001	0.007	0.045	0.001	0.003
KAV	0.172	0.238	0.205	0.133	0.185	-	0.000	0.000	0.000	0.001	0.012	0.000	0.002
LIH	0.139	0.205	0.171	0.100	0.148	0.148	-	0.000	0.000	0.002	0.032	0.000	0.007
SOL	0.149	0.214	0.181	0.110	0.158	0.157	0.124	-	0.000	0.001	0.022	0.000	0.003
TUV	0.169	0.234	0.202	0.131	0.182	0.178	0.145	0.154	-	0.001	0.021	0.000	0.000
FIJ	0.147	0.216	0.180	0.106	0.156	0.155	0.121	0.130	0.152	-	0.033	0.001	0.005
TGA	0.113	0.189	0.150	0.069	0.119	0.122	0.085	0.095	0.119	0.090	-	0.015	0.055
COK	0.165	0.224	0.194	0.129	0.175	0.173	0.142	0.151	0.170	0.149	0.117	-	0.003
HAW	0.112	0.181	0.146	0.071	0.117	0.120	0.086	0.095	0.118	0.091	0.051	0.116	-

Φ_{ST}	NIN	ASH	TIM	LIZ	MOT	KAV	LIH	SOL	TUV	FIJ	TGA	COK	HAW
NIN	-	0.698	0.762	0.658	0.881	0.877	0.820	0.998	0.824	0.698	0.698	0.998	0.000
ASH	-0.002	-	0.094	0.434	0.799	0.810	0.658	0.808	0.390	0.771	0.505	0.799	0.000
TIM	-0.006	0.121	-	0.658	0.651	0.658	0.658	0.737	0.877	0.303	0.698	0.505	0.000
LIZ	0.021	0.072	0.026	-	0.693	0.877	0.303	0.698	0.832	0.698	0.998	0.390	0.000
MOT	-0.053	-0.029	0.059	0.027	-	0.820	0.698	0.881	0.698	0.881	0.698	0.913	0.000
KAV	-0.030	-0.008	0.023	-0.034	-0.036	-	0.698	0.939	0.824	0.881	0.877	0.824	0.000
LIH	-0.018	0.016	0.020	0.077	0.019	0.016	-	0.820	0.698	0.368	0.505	0.698	0.000
SOL	-0.057	-0.012	-0.009	0.005	-0.064	-0.043	-0.023	-	0.877	0.808	0.771	0.998	0.000
TUV	-0.021	0.053	-0.038	-0.032	0.016	-0.030	0.010	-0.028	-	0.698	0.964	0.658	0.000
FIJ	0.001	-0.014	0.091	-0.008	-0.043	-0.041	0.061	-0.015	0.016	-	0.771	0.698	0.000
TGA	0.013	0.088	0.004	-0.094	0.040	-0.053	0.059	-0.006	-0.067	-0.019	-	0.571	0.000
COK	-0.035	-0.012	0.036	0.047	-0.061	-0.016	0.000	-0.042	0.014	0.001	0.045	-	0.000
HAW	0.856	0.894	0.836	0.809	0.861	0.842	0.872	0.853	0.821	0.852	0.818	0.861	-

D_{es}	NIN	ASH	TIM	LIZ	MOT	KAV	LIH	SOL	TUV	FIJ	TGA	COK	HAW
NIN	-												
ASH	-0.084	-											
TIM	-0.078	0.033	-										
LIZ	-0.217	0.020	0.067	-									
MOT	-0.171	-0.010	0.185	-0.332	-								
KAV	-0.162	-0.083	0.023	-0.200	-0.129	-							
LIH	-0.093	-0.042	0.093	-0.156	-0.011	-0.084	-						
SOL	-0.247	-0.037	-0.081	-0.339	-0.304	-0.188	-0.149	-					
TUV	-0.170	-0.082	-0.115	-0.086	0.070	-0.124	-0.094	-0.135	-				
FIJ	-0.075	0.030	-0.052	-0.084	0.178	-0.099	-0.004	-0.116	-0.163	-			
TGA	-0.319	-0.132	-0.006	-0.694	-0.371	-0.301	-0.619	-0.445	-0.251	-0.466	-		
COK	-0.137	-0.008	0.061	-0.196	-0.263	-0.104	-0.013	-0.202	-0.028	0.031	-0.297	-	
HAW	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-

Appendix 2. 'Dispersal distances' derived from a biophysical model as described in text (v). The distances are from the locations indicated in the rows, to the locations in the columns. 'NA' indicates there were no simulated dispersal events based on the model parameters.

Dascyllus trimaculatus

	ASH	HER	KAV	LDH	LIH	LIZ	MOT	NIN	SOL	TIM	WIG
ASH	0.000	46.562	32.224	61.440	47.804	36.788	30.522	16.965	36.934	12.768	29.257
HER	28.766	0.000	29.355	16.489	17.080	11.484	15.628	36.089	22.040	23.476	14.709
KAV	33.662	27.497	0.000	42.375	25.086	18.108	14.766	41.107	10.445	28.494	19.727
LDH	NA	NA	NA	0.000	NA	NA	NA	NA	NA	NA	NA
LIH	23.230	5.964	23.820	20.842	0.000	5.948	10.093	30.553	16.504	17.940	9.173
LIZ	17.282	9.774	17.871	24.652	16.033	0.000	4.144	24.605	10.556	11.992	3.225
MOT	21.696	18.924	16.014	33.803	19.569	9.150	0.000	29.019	8.698	16.406	7.639
NIN	34.715	81.277	66.939	96.156	82.519	71.503	65.238	0.000	71.649	47.483	63.599
SOL	30.005	23.209	13.330	38.088	20.798	13.821	11.757	37.327	0.000	24.715	15.948
TIM	6.083	36.249	28.842	51.128	37.492	26.476	20.210	23.047	26.621	0.000	18.945
WIG	16.317	31.186	28.874	46.065	32.428	21.412	15.147	23.639	21.558	11.027	0.000

Pomacentrus coelestis

	ASH	HER	KAV	LDH	LIH	LIZ	MOT	NIN	SOL	TIM	WIG
ASH	0.000	65.736	47.069	79.806	71.834	56.613	51.995	63.697	46.908	15.280	56.630
HER	56.584	0.000	82.551	17.131	22.412	32.630	46.765	119.678	73.958	52.617	36.303
KAV	38.718	27.953	0.000	42.023	32.958	18.831	14.212	101.812	8.462	34.751	18.847
LDH	NA	NA	NA	0.000	NA	NA	NA	NA	NA	NA	NA
LIH	34.172	7.191	60.139	21.261	0.000	10.218	24.354	97.266	51.546	30.205	13.891
LIZ	23.954	9.123	49.921	23.192	22.344	0.000	14.135	87.048	41.328	19.987	3.673
MOT	25.909	17.195	42.276	31.264	30.416	8.072	0.000	89.003	33.683	21.942	7.261
NIN	35.402	99.881	81.213	113.951	105.979	90.758	86.140	0.000	81.053	49.424	90.774
SOL	38.392	28.066	14.787	42.135	26.816	18.943	13.787	101.486	0.000	34.425	18.756
TIM	5.635	61.356	42.688	75.425	67.454	52.233	47.615	68.896	42.528	0.000	52.249
WIG	29.742	80.767	62.100	94.837	86.865	71.644	67.026	92.836	61.939	25.775	0.000

Halichoeres hortulanus

	ASH	HER	KAV	LDH	LIH	LIZ	MOT	NIN	SOL	TIM	WIG
ASH	0.000	44.016	31.321	57.734	43.481	34.232	28.060	14.031	34.036	10.141	26.789
HER	24.425	0.000	28.932	15.412	16.758	11.445	15.583	34.011	21.559	21.030	14.505
KAV	29.064	23.853	0.000	37.570	23.108	17.611	14.690	38.650	8.393	25.669	19.144
LDH	NA	NA	NA	0.000	NA	NA	NA	NA	NA	NA	NA
LIH	18.803	5.370	23.310	19.087	0.000	5.823	9.960	28.389	15.937	15.407	8.883
LIZ	12.980	9.784	17.487	23.501	15.897	0.000	4.137	22.565	10.113	9.584	3.060
MOT	17.105	18.531	15.714	32.248	17.785	8.943	0.000	26.691	8.340	13.709	7.185
NIN	27.889	71.906	59.211	85.623	71.370	62.122	55.949	0.000	61.925	38.030	54.133
SOL	24.943	19.731	11.888	33.449	18.986	13.489	11.760	34.528	0.000	21.547	15.023
TIM	5.581	34.872	27.630	48.589	34.336	25.088	18.916	19.612	24.892	0.000	17.645
WIG	12.595	31.063	28.456	44.780	30.527	21.279	15.106	22.180	21.083	9.199	0.000

Acanthurus triostegus

	ASH	HER	KAV	LDH	LIH	LIZ	MOT	NIN	SOL	TIM	WIG
ASH	0.000	44.874	29.355	54.053	42.938	38.897	37.245	41.193	34.001	13.244	40.549
HER	41.544	0.000	48.433	11.994	19.233	27.254	40.277	82.615	44.353	39.035	29.897
KAV	24.755	18.899	0.000	28.078	16.643	12.922	11.270	65.826	7.414	22.246	14.574
LDH	51.855	29.756	58.744	0.000	29.544	37.565	50.588	92.926	54.665	49.346	40.208
LIH	22.311	5.250	29.200	14.257	0.000	8.020	21.044	63.382	25.120	19.802	10.664
LIZ	14.290	5.977	21.180	15.156	15.480	0.000	13.024	55.361	17.100	11.782	2.644
MOT	15.240	11.880	19.667	21.060	21.383	5.904	0.000	56.311	15.587	12.732	6.072
NIN	20.679	65.553	50.034	74.733	63.618	59.576	57.924	0.000	54.681	33.923	61.228
SOL	24.719	14.479	8.335	23.486	9.229	12.886	11.234	65.790	0.000	22.211	14.538
TIM	4.214	45.261	29.742	54.440	43.325	39.284	37.632	45.285	34.388	0.000	39.992
WIG	20.434	61.974	46.456	71.154	60.039	55.997	54.345	61.505	51.102	17.925	0.000

CHAPTER FIVE. Genetic patterns across the latitudinal range of a coral reef fish reveal historical effects and contrasting patterns of turnover and nestedness at the range peripheries

In preparation for Journal of Biogeography

Libby Liggins[1], David J. Booth[2], Will F. Figueira[3], Linda Tonk[4], David A. Harris[4], Tyrone Ridgway[5], Eric A. Treml[6] and Cynthia Riginos[1]

[1] School of Biological Sciences, The University of Queensland

[2]Department of Environmental Sciences, University of Technology Sydney

[3]The School of Biological Sciences, The University of Sydney

[4]ARC Centre of Excellence for Coral Reef Studies and School of Biological Sciences, The University of Queensland

[5]Australian Institute of Marine Science

[6]Department of Zoology, The University of Melbourne

Abstract

Aim Few studies have examined core-periphery genetic patterns in tropical marine taxa. We aimed to extend theory and decouple the roles of historical and contemporary demographic processes in forming latitude-wide genetic patterns of the neon damselfish (*Pomacentrus coelestis*).

Location The Indian and Pacific Oceans, from 36°N along the coastline of Japan to 37°S along the coastline of south east Australia, including the Indo-Australian-Archipelago (IAA).

Methods We investigated changes in population genetic diversity and differentiation (based on mtDNA control region) across the latitudinal range of the neon damselfish using analyses of covariance and regression frameworks. We extended typical core-periphery analyses to also include genealogical analyses to identify cryptic lineages and methods more commonly used in community ecology, including: β-diversity (β_{SOR}) and its turnover (β_{SIM}) and richness (β_{SNE}) components; and analyses of nestedness (based on overlap and decreasing fill).

Results The existence of two divergent clades (Pacific and Micronesian) of the neon damselfish in the IAA influenced levels of genetic diversity: locations with both clades had elevated levels of genetic diversity, and locations with only the Micronesian clade had significantly lower genetic diversity than those of the Pacific clade. Focusing only on the widespread Pacific clade, nucleotide diversity was significantly higher in the core than in the northern and southern peripheries. However, extended analyses revealed differences in the genetic patterns of these peripheral regions. Whereas the turnover of haplotypes (pairwise-β_{sim}) increased over distance in the north, a higher proportion of the haplotypic compositional differences among populations (β_{SOR}) were due to richness differences (β_{SNE})

in the southern periphery. Moreover, nestedness analyses revealed that the haplotype compositions of populations in the south were nested according to latitude.

Main conclusions By extending the typical characterizations of genetic patterns across a species range we were able to identify the effects of lineage admixture on measures of genetic diversity and contrasting demographic histories toward the northern and southern peripheries of the neon damselfish's range. Our partitioning of β-diversity and nestedness analyses revealed that the northern and southern peripheries of the species have different histories - the genetic patterns toward the northern periphery (increasing turnover, β_{sim}, over geographic distance) are indicative of historical colonization with little contemporary migration; whereas the genetic composition of the populations toward the southern periphery of the Pacific clade's range are consistent with contemporary connectivity and immigration from lower latitudes.

Introduction

Range-wide genetic patterns provide a window into a species' history, the contemporary demography of populations, and can ultimately help inform conservation priorities. Despite the importance of range-wide genetic information, predictive hypotheses of genetic patterns across a species' entire range are rare. The core-periphery hypothesis (CPH; also known as the central-marginal hypothesis, da Cunha et al. 1950, Brussard 1984, Eckert et al. 2008) can be considered a null hypothesis for range-wide genetic patterns, although the model is very simple. The CPH considers two factors that influence the balance of migration, mutation, and drift across a species range to predict range-wide genetic patterns. First, it assumes that the close geographic arrangement of populations in the core of a species range will lead populations to exchange migrants frequently, reducing compositional differences among core populations. Second, the CPH predicts that greater standing genetic variation will be maintained in the core of a species range (Lammi et al. 1999) based on the assumption that population sizes will be larger in the core than in the periphery (i.e. the abundant-center hypothesis: Antonovics 1976, Hengeveld and Haeck 1982). Thus, the CPH predicts two genetic patterns across a species range: core populations will have higher genetic diversity than peripheral populations; and peripheral populations will be more genetically differentiated from each other relative to comparisons among core populations (Fig. 1a).

Although several studies have described range-wide genetic patterns in agreement with the CPH expectations (reviewed in Eckert et al. 2008), the historical and contemporary demographic processes underlying those genetic patterns likely vary. A number of demographic processes affect patterns of genetic diversity and differentiation and are not considered in the CPH. For example, genetic diversity can be elevated by admixture, such as when divergent lineages become sympatric following periods of demographic isolation (Fig. 1a; e.g. Petit et al. 2003). Furthermore, genetic

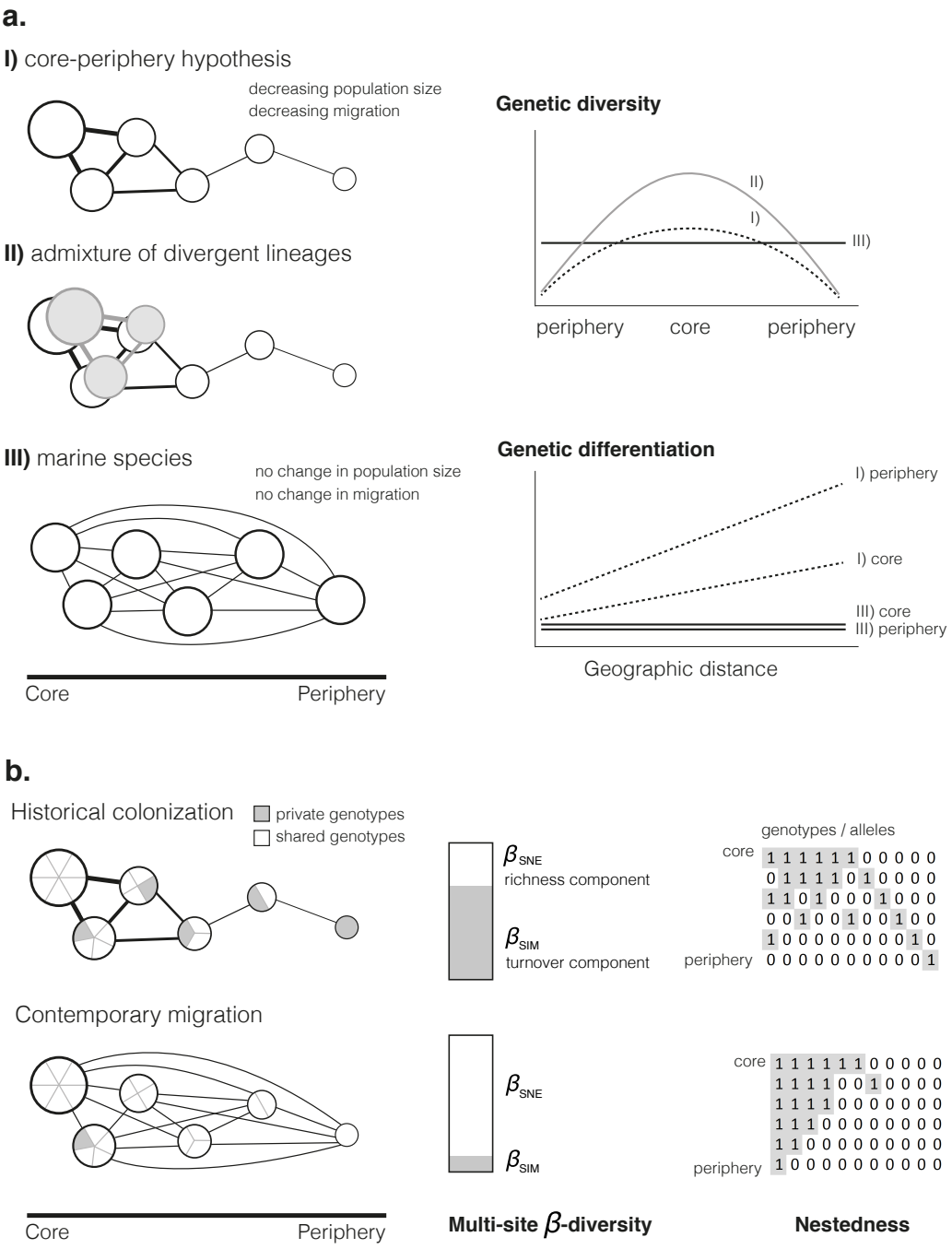


Figure 1. Hypothetical scenarios underlying the range-wide genetic patterns of a tropical marine species. a. Expected patterns of genetic diversity across the range and genetic differentiation over geographic distance based on: I) the core-periphery hypothesis; II) admixture of divergent lineages in the range core; and III) high levels of migration and large population sizes that may be typical of a marine species. Circles represent relative population sizes and weighted lines represent migration rates. b. The contribution of the turnover (β_{SIM}) and richness components (β_{SNE}) to total multi-site β_{SOR}-diversity (bar height) and expected patterns of nestedness toward the range periphery under a scenario of historical colonization and contemporary isolation, versus asymmetric contemporary migration into the range periphery. Circle partitions denote genotypic composition (private or shared) and relative diversity (number of partitions).

diversity is generally higher in older populations and decreases predictably according to colonization timing (Hewitt 1996, Hardie and Hutchings 2010). Lastly, populations toward the range periphery are less likely to be in migration-drift equilibria and therefore may have highly variable patterns of genetic diversity and differentiation. Classical tests of the CPH relying only patterns in genetic diversity and differentiation cannot capture the contribution of such underlying population histories.

Methods that extend the usual characterizations of CPH genetic patterns may help disclose the demography underlying range-wide genetic patterns. For example, genealogical analyses prior to range-wide analysis can identify lineages that are now sympatric (Eckert et al. 2008). Furthermore, coalescent approaches (Hey 1991, Beerli and Felsenstein 1999) can distinguish historical versus contemporary migration and asymmetric migration common in colonization and non-equilibrium scenarios (e.g. Wakeley 1999). Alternatively, concepts and methods used in community ecology to describe analogous spatial patterns of species diversity can be applied to genetic data (Diniz-Filho and Bini 2011, Liggins et al. 2013 – Chapter Three). In particular, two methods could be useful for characterizing population histories toward a species range periphery: the additive partitioning of β-diversity (Baselga 2010, 2012) and analyses of nestedness (Darlington 1957, Daubenmire 1975).

Traditionally, β-diversity has been used as a measure of the variability in species composition among sites or communities (Whittaker 1960) but it can also measure the variability in the genetic composition among populations (Diniz-Filho et al. 2012). Sorenson's dissimilarity, β_{SOR} (Sørensen 1948) among populations can be partitioned into components due to the replacement of genotypes among populations (the turnover component, β_{SIM}; Simpson's dissimilarity, Simpson 1943) and a component describing differences in genotype richness among populations (the nestedness-resultant component, herein referred to as the richness component, β_{SNE}, Baselga 2010, 2012). Thus, the partitioning of β-diversity can be used to reveal the contribution of private (the turnover component, β_{SIM}) versus shared genotypes (the richness component, β_{SNE}) to diversity changes and the differentiation among populations. Such partitioning of β-diversity could be useful beyond F-statistics and genetic diversity measures that are typically used in spatial genetic analyses. For example, when pairwise F_{ST} is equal to 1 between populations, and/or genetic diversity is consistently high across populations, there are still several scenarios for how these genetic patterns could be conferred by private versus shared genetic diversity (β_{SIM} and β_{SNE}, respectively). Moreover, the relative contribution of private versus shared genotypes is meaningful in a population genetic context. We would predict that a greater contribution of private genotypes (β_{SIM}) to β-diversity (β_{SOR}) among populations would be indicative of mutation and drift (Slatkin 1985), whereas high rates of genotype sharing (β_{SNE}) would be indicative of migration. Thus, toward a species range periphery we may anticipate a greater contribution of turnover (β_{SIM}) among populations along an historical colonization pathway and a greater contribution of richness differences (β_{SNE}) among populations that have contemporary migration, or are not yet in migration-drift equilibrium (Fig. 1b).

The concept of nestedness, or the way in which communities form subsets of other communities (Wright and Reeves 1992, Almeida-Neto et al. 2008), can also be used to describe the nature of genetic diversity changes among populations. Where there is a spatial gradient of interest, such as a core-periphery transect, the nested structure of genotype compositions along that gradient can be statistically compared to null expectations (Ulrich et al. 2009). In a population genetic context, populations that share the same genotypes are likely to have migration between them. Furthermore, populations that have only a reduced subset of the shared genotypes (i.e. nested populations) are likely to represent demographic 'sinks'. Thus, a significant pattern of nestedness in genotypic composition toward the range periphery would not be expected in a scenario of historical colonization because the influence of mutation and drift would predominate. Rather, we would expect to find that the genotypic compositions of populations would be significantly nested toward the range periphery when there is contemporary immigration from the core into peripheral populations (Fig. 1b).

There have been comparatively few range-wide investigations of population genetic patterns and thus few tests of the CPH in either tropical or marine systems (reviewed in Eckert et al. 2008, but see Palma-Silva et al. 2009, Miller et al. 2010, Liggins et al. 2014 – Chapter Six). Tropical systems may be especially useful for testing the CPH as there is the opportunity to address patterns in two latitudinally peripheral regions, presumably limited by similar environmental conditions. Marine systems, on the other hand, introduce several challenges for the limited parameters of the CPH. Many marine species have high levels of dispersal (Hellberg 2009) and large population sizes throughout their ranges (Sagarin and Gaines 2002) potentially compromising the expectations of the CPH for reduced migration and smaller population sizes toward the range periphery (Tuya et al. 2008, Liggins et al. 2014 – Chapter Six; Fig. 1a). Furthermore, the dispersal of many taxa (particularly those with pelagic larvae) is highly subject to ocean currents that can disconnect proximal populations and connect geographically distant populations (Riginos and Liggins 2013 – Chapter One). Thus, range position may be less relevant in determining range-wide genetic patterns of some marine species as compared to more sedentary terrestrial taxa.

Nonetheless, the population histories of tropical marine organisms might lead to genetic patterns that inadvertently support the expectations of the CPH. The range core of many tropical marine species sits around the Indo-Australian-Archipelago where the Indian and Pacific Oceans meet (IAA, Briggs and Bowen 2012). This region, and specifically the Coral Triangle, is posited to be the "Centre of Origin" for several tropical marine species, and the "Centre of Overlap" and "Centre of Accumulation" for others (Bowen et al. 2013, Fig. 2). Cryptic lineages are common in marine populations (Knowlton 1993, Rocha and Bowen 2008, Huelsken et al. 2013 - Appendix Three) and cases of lineage sympatry have been described within the IAA (reviewed in Carpenter et al. 2011). Furthermore, many reef-associated fishes that originated in the Coral Triangle, have since colonized and expanded their ranges into the Pacific and Indian Oceans (Cowman and Bellwood 2012). Thus,

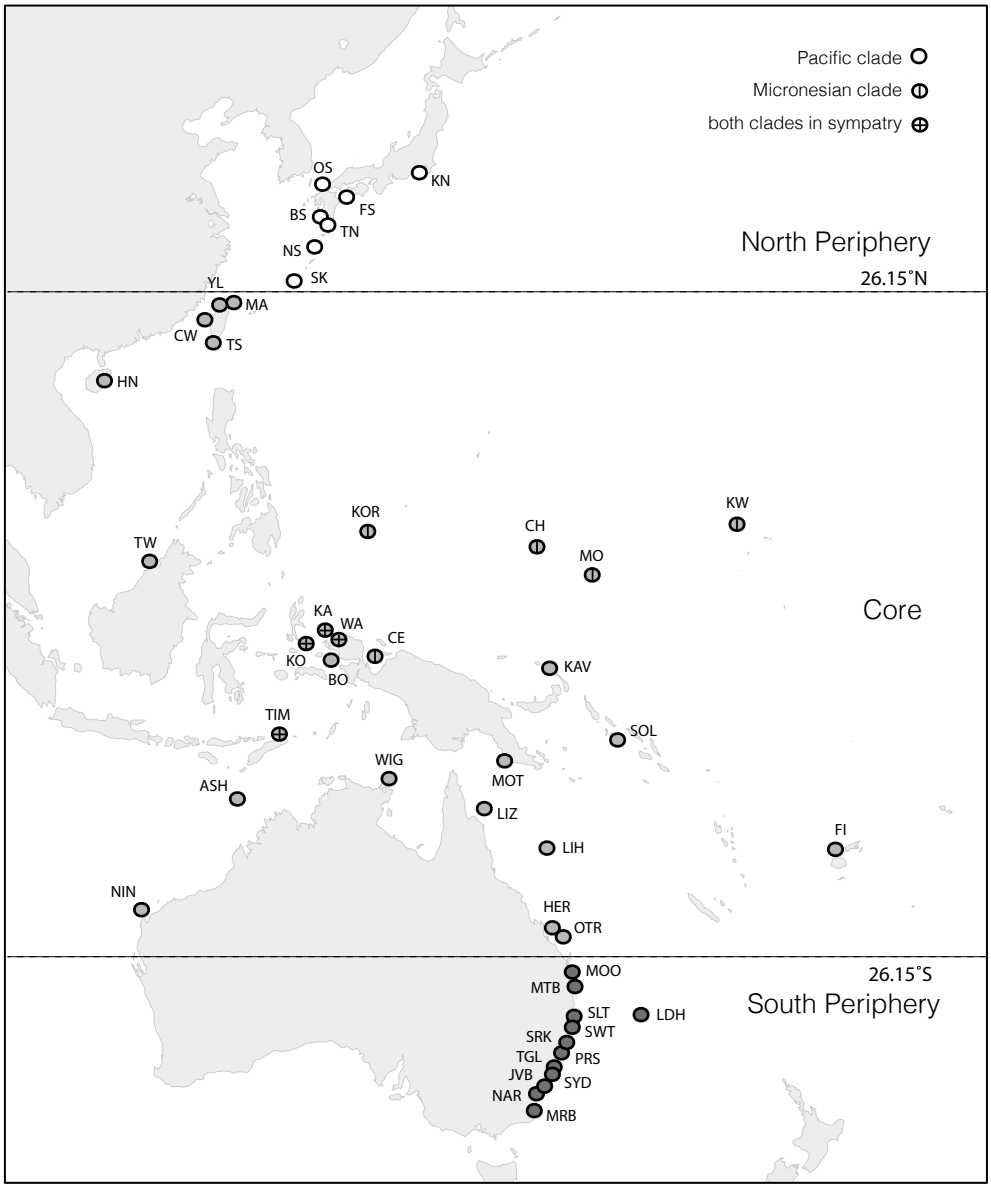


Figure 2. Map of the study area including the latitudinal core-periphery range limits of the neon damselfish. Populations from which genetic diversity measures (π, *Hd*) were attained are shown by the points. Attributes of populations are designated as follows: range position- white = north periphery, light grey = core, dark grey = south periphery; clade affinities- no bars = Pacific clade, vertical bar = Micronesian clade, crossed = Micronesian and Pacific clades in sympatry.

there is cause to expect that genetic diversity will be elevated in the range core of many tropical reef-fish species due to the sympatry of divergent lineages, and that genetic diversity will decline toward their range peripheries according to colonization timing.

In this study we investigate the latitude-wide genetic patterns of a common coral reef fish, the neon damselfish species complex (*Pomacentrus coelestis*, Jordan and Starks 1902). A recent study of the neon damselfish reported the existence of two genetically divergent, yet morphologically cryptic lineages (Liu et al. 2012). The authors describe a Micronesian clade (re-described as *Pomacentrus micronesicus* in Liu et al. 2013) that is found at low latitudes, but co-occurs in parts of the Coral Triangle with another widespread clade (*P. coelestis*; herein referred to as the Pacific clade *sensu* Liu et al. 2012, Fig. 2). In this study we first address the impacts of population history on a test of the CPH, using the extreme case whereby two divergent clades co-occur in the core of a species range, but not the periphery. We expect that genetic diversity will be elevated in the core of the species range due to the sympatry of divergent lineages, providing a pattern in support of the CPH, but resulting from an entirely different process (e.g. lineage admixture, not core abundance and high connectivity, Fig. 1a.). We then focus on the widespread Pacific clade of the neon damselfish. Across the latitudinal range of the Pacific clade, we expect either: genetic patterns that support the CPH; or no change in genetic diversity or differentiation as might be expected for a relatively high dispersal marine species (pelagic larval duration ~19.5 days, Thresher et al. 1989) that maintains large population sizes (Fig. 1a.).

Finding patterns of genetic diversity and genetic differentiation in support of, or contrary to, the CPH alone may be inadequate to understand the historical and contemporary processes acting toward the range periphery of the Pacific clade. There are two demographic scenarios likely, either: historical colonization toward the range periphery; or contemporary migration into the range periphery driven by western boundary currents (i.e. the Kuroshio current in the north, Liang et al. 2008; and the East Australian Current in the south, Suthers et al. 2011; Fig. 1b.). In both of these scenarios we might anticipate patterns that concur with those described by the CPH – a decrease in genetic diversity toward the range periphery, and an increase in genetic differentiation (although patterns could also deviate unpredictably in the case of non-equilibrium populations). To distinguish between our alternate hypotheses we extend the usual characterizations of genetic patterns across a species range to include partitioned β-diversity and nestedness analyses. Under an historical colonization scenario, we expect that turnover (β_{SIM}) will be high in the peripheral regions indicating contemporary demographic isolation. Conversely, under a scenario of contemporary migration, we expect that the richness component (β_{SNE}) will be high (indicating shared genotypes) and the genotypic composition of populations will be nested toward the range periphery (Fig. 1b.).

Methods

Fieldwork, laboratory methods, and data acquisition

Neon damselfish were collected from 21 locations throughout the species range using pole spears and hand nets while on SCUBA or snorkel. Total genomic DNA was extracted from collected tissue using a salt extraction method (modified from Aljanabi and Martinez 1997). The mitochondrial control region (CR) and cytochrome-oxidase subunit one (CO1) were amplified using CR-A and CR-E (Lee et al. 1995) and the universal fish primers (Fish F1 and Fish R1, Ward et al. 1995), respectively, following the protocols detailed in Mirams et al. (2011 – Appendix One). Amplicons were purified using Exonuclease I and Antarctic Phosphatase following the Exo-SAP protocol (New England Biolabs) and sequenced by Macrogen (Korea) via capillary electrophoresis. Sequences were manually checked and edited in CodonCode Aligner v3.7.1.2 (CodonCode Corporation) and aligned using Se-Al v2.0a11 (Rambaut 1996). The aligned sequences were translated into amino acid sequences using the vertebrate mitochondrial code to ensure they were not of nuclear origin and the primer sequence and regions of insignificant overlap at either end of the sequences were deleted in Se-Al. All *de novo* sequences are available on GenBank (KJ779106-KJ779397, KJ779872-KJ779960).

Our *de novo* sequences for the neon damselfish species complex were combined with other sequences available on GenBank (Liu et al. 2008: EF420785-EF420855, Liu et al. 2010: EU366845-EU366889, Mirams et al. 2011: JF717975-JF718155, Liu et al. 2012: JF314773-JF314842, JQ418300-JQ418311). Where the sequence data from GenBank had insufficient relating information (i.e. no geo-reference, no haplotype frequencies), data were excluded from some analyses and in some cases published summary statistics were taken from the original publication (detailed in Supp. Methods and Appendix 1.).

Identification of clades and core-periphery range limits

Unrooted neighbor-joining gene trees were constructed in Geneious Pro 5.6.3 (Biomatters, www.geneious.com; bootstrap values estimated by 1,000 replicates) using both the CR and CO1 datasets (540 sequences, 336 bp and 151 sequences, 569bp, respectively) to identify the clade affinities of individuals and any locations of clade sympatry (hereafter also treated as, and referred to as 'populations').

We categorically designated populations as being in the "core" or "periphery" of the species range according to the method of Channell and Lomolino (2001) where the core and periphery are of equal area and the area defining the core-boundary sits equidistant from the species range edge in each dimension. We used a rectangle to represent the range extremes rather than any other polygon, as our focus is on the latitudinal extremes (range limits were taken from Allen 1991 and extended where necessary to include populations sampled herein). The latitudinal core-boundaries designated three regions: north periphery (NP), core, and south periphery (SP).

To view the genetic structure and genealogical relationships among regions (NP, core, SP) a median-joining network was constructed for individuals of the Pacific clade for which we had geo-referenced CR sequences using Network 4.6.1.0 and Network Publisher 2.0 (fluxus-engineering.com, Bandelt et al. 1999). To reduce complexity, non-parsimonious links were deleted using the maximum-parsimony calculation option (Polzin and Daneshmand 2003).

Core-periphery patterns of genetic diversity

According to the expectations of the CPH, we expected genetic diversity to be higher in the core of the neon damselfish's range, and we expected this trend to be exacerbated by the sympatry of divergent clades. To test these hypotheses we combined summary statistics from our *de novo* dataset with those of previous studies. We calculated measures of nucleotide diversity π (Nei 1978) and haplotype diversity *Hd*, for each of our sampled populations, and populations studied in Mirams et al. 2011, in the same manner as Liu et al. (2008, 2010, 2012) using Arlequin 3.5 (Excoffier and Lischer, 2010; see Supp. Methods).

We used two methods to investigate the influence of range position on genetic diversity. First, to examine patterns in genetic diversity across the latitudinal range we used second-order polynomial regression. Hence, a significant, negative curvature coefficient would indicate a peak in genetic diversity in the core of the species range, and a significant linear coefficient would reveal any north to south trends in genetic diversity (base package, R v2.15.3, R Core Team). Second, to examine whether the relationship between distance from the core and genetic diversity for the northern and southern hemispheres differed, we conducted analyses of covariance (ANCOVA). In ANCOVA analyses, genetic diversity was the response, hemisphere was a fixed factor with two levels (north and south), and absolute latitude was the covariate. All analyses were conducted using both the combined Micronesian and Pacific clade dataset and the Pacific clade only dataset to understand the influence of divergent lineages and sympatry. Model residuals were visualized in all cases to check that the data met the assumptions of normality and homogeneity of variance.

To further understand any influence of clade affinity and clade sympatry on levels of genetic diversity we used a standard univariate analysis of variance (ANOVA). In ANOVA analyses, populations were classified according to their clade affinities based on our genealogical analyses and affinities reported in Liu et al. (2012). Clade was included in the model as a fixed factor with three levels: Micronesian (Mic), Pacific (Pac), and sympatric (Sym). Tukey's test for Honestly Significant Differences was used to uncover pairwise differences among locations in cases where we found a significant main effect of clade.

Core-periphery patterns of genetic differentiation

To understand how patterns of genetic differentiation changed across the range of the Pacific clade we used isolation-by-distance (IBD) analyses. An IBD pattern (Wright 1943) can indicate the geographic scale of approximate migration-drift equilibrium (Hutchison and Templeton 1999).

According to the CPH expectations for patterns of genetic differentiation, we expected to find a greater intercept value and a greater positive relationship between genetic differentiation and geographic distance in the species range periphery (Fig. 1b.). We generated pairwise Φ_{ST} values for our populations and those studied in Mirams et al. (2011) in the same manner as Liu et al. (2008, 2010; in Arlequin using uncorrected pairwise differences, 1,000 permutations). We compared the IBD relationship between Φ_{ST} and Euclidean geographic distances for the NP, SP and two sub-regions of the core (i.e. only populations within 1,580km of each other, comparable to the smaller pairwise geographic distances of the peripheral regions). Reduced major axis regression was used to derive intercepts and measures of fit (R^2); Mantel tests were used to assess correlation (r_M) and significance (p–values based on 10,000 permutations; Isolation By Distance Web Service, Jensen et al. 2005).

Coalescence analyses

Our historical colonization scenario would be supported by a historical signal of migration from the core to the periphery and contemporary isolation. In contrast high levels of migration, possibly without any indication of direction, would be indicative of our contemporary migration scenario. To estimate colonization history and directional migration in the southern and northern hemisphere of the Pacific clade's range we used Isolation with Migration (IMa2 2.0, Hey 2010). We selected four representative populations along a similar transect to the north (KN, FS, NS, and SK) and the south (MRB, SRK, MTB, and HER) of the clade's range. We ran analyses for neighboring pairs, with the intention of comparing posterior distributions and values for north-south migration, south-north migration ($2Nm$) and divergence times (t) among all pairs. Despite several preliminary runs using up to 10 chains, different chain-mixing regimes (linear, geometric, and heating options), and a broad range of priors, we were unable to attain satisfactory effective sample sizes or reliable parameter estimates for the dataset. For both the migration parameter (m) and the population size parameter (q), posterior probabilities were very flat and did not return to 0 at the upper bound, suggesting our parameter estimates could be very sensitive to our selected priors (as detailed in the program documentation). For these reasons we elected not to continue with coalescent analyses.

Genetic patterns toward the northern and southern periphery

Measures of β-diversity (Sorenson's β_{SOR}) and its turnover (β_{SIM}) and richness (β_{SNE}) components were calculated across all Pacific clade populations for which we had geo-referenced sequence data. A binary matrix of haplotype presence and absence for each region was constructed, then multi-site β_{SOR} and its β_{SIM} and β_{SNE} components were calculated for the NP, core, and SP using the beta.multi function of the package 'betapart v1.2' in R (Baselga and Orme 2012). According to the CPH, we expected to find higher values of β_{SOR} in the range periphery of the Pacific clade as compared to the core (corresponding to peripheral populations being more genetically dissimilar). In the range periphery, we expected that higher values of the turnover component β_{SIM} would be indicative of historical colonization and contemporary isolation. Alternatively, finding larger values of $\beta_{SNE,}$ in the

range periphery would indicate a greater influence of contemporary migration in this part of the species range.

Pairwise β-diversity measures across all populations were also calculated using the beta.pair function of the package 'betapart' in R. These distance matrices (pairwise-β_{sor}, pairwise-β_{sne} and pairwise-β_{sim}) were each individually analyzed in an IBD framework for the maximum transect possible in the north (TS - KN, excluding CW, YL, MA; total distance 2387km) and a similar transect in the south (LIH - MRB; total distance 2269 km). For comparison, we also conducted IBD analyses over the same transects using standard genetic differentiation measures: F_{ST}, Φ_{ST} (in Arlequin), and D_{est} (Jost 2008, in Genodive v. 2.0b23, Meirmans and van Tienderen 2004). Under a scenario of historical colonization toward the range periphery we expected to find a positive relationship between pairwise-β_{sor} (as well as F_{ST}, Φ_{ST}, and D_{est}) and geographic distance, a stronger positive relationship between pairwise-β_{sim} and geographic distance, and a negative or nil relationship between pairwise-β_{sne} and geographic distance (i.e. richness differences between populations decrease over distance or have no relationship to distance). Under a scenario of contemporary migration we did not expect to find any relationships significantly different from 0 between geographic distance and any of the pairwise-β measures (or F_{ST}, Φ_{ST}, and D_{est}, Slatkin 1993).

A targeted analysis of nestedness (NODF, nestedness based on overlap and decreasing fill, as in Almeida-Neto et al. 2008) using the nestednodf function of the package 'vegan v2.0-7' (Oksanen et al. 2007) was used to address the nested structure of population haplotype composition in relation to latitude for the north and the south of the Pacific clade's range (transects as described above). The rows (representing populations) of the haplotype presence/absence matrices were organized according to latitude and columns (representing haplotype identity) were organized using a custom, standardized procedure to optimize their arrangement for 'decreasing fill'. This enabled us to attain the maximum value for $NODF_{rows}$ (nestedness according to latitude) in both the northern and southern transects, ranging from 0 (non-nested) to 100 (perfectly nested). The organization of rows was subsequently randomized 1,000 times, each time followed by an optimization of their arrangement for 'decreasing fill', to test the significance of the nested relationship of haplotype composition with latitude. Under a scenario of historical colonization and contemporary isolation we did not expect to find nestedness according to latitude, however for a scenario of asymmetric contemporary migration into the periphery we expected that haplotypes would be nested predictably according to latitude (i.e. the haplotype diversity found within higher latitude populations would be nested within lower latitude populations).

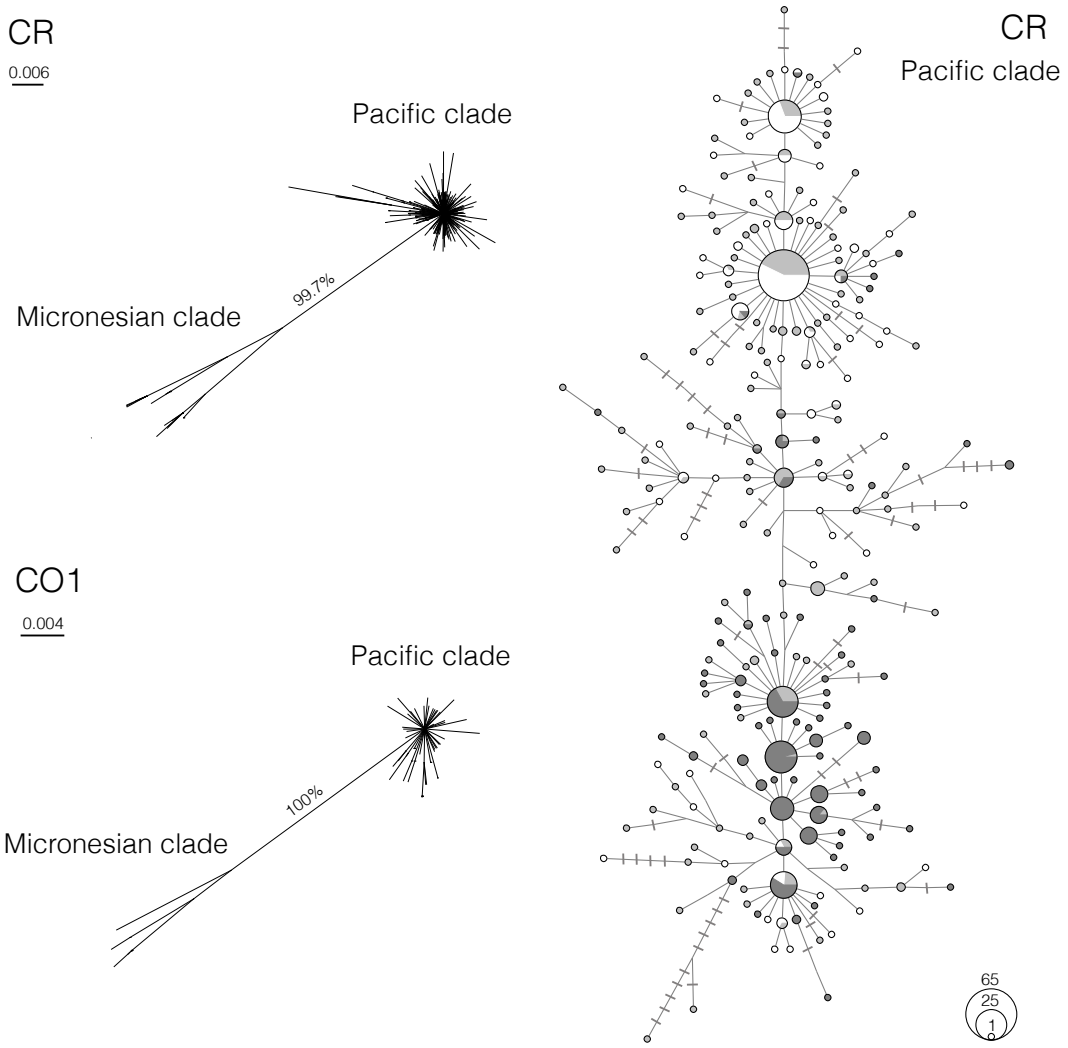


Figure 3. Left: Neighbor-joining consensus gene trees based on control region (CR) and cytochrome oxidase subunit one (CO1) sequences showing the relationship between the Micronesian and Pacific clades of the neon damselfish. Bootstrap support for the relationship between the Micronesian and Pacific clade following 1,000 replicates is indicated above the connecting branch. Right: Maximum parsimony median-joining network showing unique CR haplotypes and their frequencies (see size key, bottom right) across the range of the Pacific clade. Haplotypes are shaded according to range position: white = north periphery, light gray = core, dark gray = south periphery. Edges between haplotypes represent one mutational step, with additional steps indicated by cross bars.

Results

Identification of clades and core-periphery range limits

For both the CR and CO1 dataset, our consensus neighbor-joining gene trees portrayed two distinct and divergent clades across the sampled range of the neon damselfish species complex, with high bootstrap support (CO1: 100% and CR: 99.7%; Fig. 3). Our *de novo* dataset contained individuals from both the Micronesian clade (21 individuals, TIM only) and the Pacific clade (333 individuals, including all *de novo* sampled populations) identified in Liu et al. (2012). Thus, our complete dataset included 5 populations consisting of the Micronesian clade only, 4 locations comprising both clades, and 37 populations including the Pacific clade only (Table 1, Fig. 2).

The polygon representing the core of the species range was bounded by 100.5°W, 160.5°E, 26.15°N, 26.15°S designating 27 populations into the core, 7 populations into the NP, and 12 populations into the SP (Fig. 2).

The median-joining haplotype network recovered 240 unique haplotypes within the Pacific clade. The core had 141 haplotypes, the NP had 61 haplotypes, and the SP had 66 haplotypes. The maximum parsimony analysis of the constructed median-joining network recovered many equally parsimonious network configurations; however there were minimal structural differences among the networks as most differences related to the position of single haplotypes that were separated by only 1-2 base pairs (one random network shown in Fig. 3). There were several shared haplotypes among the core, NP and SP of the species range.

Core-periphery patterns of genetic diversity

As we had anticipated, ANOVA revealed that populations comprising only one clade had significantly lower π than the locations with sympatric clades (Table 2; after Tukey's test: Mic-Sym, diff = -0.029, $P < 0.001$; Pac-Sym= -0.021, $P < 0.001$). Surprisingly the Micronesian clade also had significantly lower π than the Pacific clade (diff = -0.008, $P = 0.008$). Results for *Hd* were slightly different (Table 2); populations of the Micronesian clade had significantly lower *Hd* than locations of sympatry (diff = -0.222, $P = 0.029$) and Pacific clade populations (diff = -0.341, $P < 0.001$). However, populations of the Pacific clade did not have significantly different levels of *Hd* than locations with sympatric clades (diff = 0.119, $P = 0.178$).

As expected under the CPH expectations, the second-order polynomial regression revealed a peak in π at low latitudes (Fig. 4). The peak was highest in the overall dataset where locations of sympatry generally had high values (except WA, Fig. 4), however this was not significant, probably owing to very low π of Micronesian clade populations also located at low latitudes (Table 2; note that all *t*–values and the significance of relationships were evaluated using orthogonal second-order polynomial regression analyses, the presented coefficients are on the raw data scale to assist with interpretation). The ANCOVA analyses also indicated that in the dataset comprising both clades π was higher at low latitudes ($F_{1,43} = 3.229$, $P = 0.079$).

Table 1. Details of all included neon damselfish populations including location, designated code, and categories for analysis: north periphery (NP); Core; south periphery (SP); Pacific clade (Pac); Micronesian clade (Mic); both clades (Sym).

Code	Location	n	Latitude, longitude	Range position	Clade
KN	Kominato, Japan	20	35.04°N, 140.07°E	NP	Pac
OS	Okinoshima Is., Japan	29	34.09°N, 130.04°E	NP	Pac
FS	Funakoshi, Japan	15	33.02°N, 132.16°E	NP	Pac
BS	Bohnotsu, Japan	23	31.09°N, 131.09°E	NP	Pac
TN	Tanegashima Is., Japan	26	30.43°N, 130.59°E	NP	Pac
NS	Nakanoshima, Japan	22	29.30°N, 129.31°E	NP	Pac
SK	Sesoko Is., Japan	16	26.23°N, 127.31°E	NP	Pac
MA	Maoao, Taiwan	25	25.21°N, 121.73°E	Core	Pac
YL	Yehliu, Taiwan	26	25.10°N, 122.07°E	Core	Pac
CW	Chinwan Outer Bay, Taiwan	23	23.47°N, 119.62°E	Core	Pac
TS	Tiaoshi, Taiwan	19	21.57°N, 120.46°E	Core	Pac
HN	Hainan Is., China	16	18.64°N, 110.67°E	Core	Pac
KW	Kwajalein Atoll, Marshall Is.	10	9.21°N, 167.47°E	Core	Mic
KOR	Koror, Palau	14	8.17°N, 134.66°E	Core	Mic
CH	Chuuk Atoll	11	7.30°N, 151.53°E	Core	Mic
MO	Mortlock Is., Chuuk State	7	5.50°N, 153.83°E	Core	Mic
TW	Two Fat Thom, Brunei	8	5.25°N, 115.07°E	Core	Pac
KA	Kawe Is., Indonesia	7	0.17°N, 130.05°E	Core	Sym
WA	Waigeo Is., Indonesia	6	0.42°S, 131.30°E	Core	Sym
KO	Kofiau Is., Indonesia	10	1.14°S, 129.93°E	Core	Sym
BO	Boo Kecil, Indonesia	3	2.12°S, 130.88°E	Core	Pac
CE	Cenderawasih Bay, Indonesia	13	2.44°S, 135.32°E	Core	Mic
KAV	Kavieng, Papua New Guinea	17	2.57°S, 150.78°E	Core	Pac
SOL	New Georgia, Solomon Islands	16	8.17°S, 157.41°E	Core	Pac
TIM_{Sym}	North Coast, Timor-Leste	40	8.30°S, 126.42°E	Core	Sym
TIM_{Pac}	North Coast, Timor-Leste	19	8.30°S, 126.42°E	Core	Pac
MOT	Motupore Is., Papua New Guinea	15	9.44°S, 147.24°E	Core	Pac
WIG	Wigram Is., Australia	16	11.15°S, 136.59°E	Core	Pac
ASH	Ashmore Reef, Australia	32	12.24°S, 123.13°E	Core	Pac
LIZ	Lizard Is., Australia	22	14.69°S, 145.49°E	Core	Pac
FI	Fiji	3	16.73°S, 177.67°E	Core	Pac
LIH	Lihou Reef, Australia	11	17.62°S, 151.53°E	Core	Pac
NIN	Ningaloo Reef, Australia	24	21.73°S, 113.98°E	Core	Pac
OTR	One Tree Is., Australia	12	23.42°S, 151.90°E	Core	Pac
HER	Heron Is., Australia	20	23.49°S, 151.93°E	Core	Pac
MOO	Mooloolaba, Australia	19	26.64°S, 153.12°E	SP	Pac
MTB	Moreton Bay, Australia	23	27.19°S, 153.25°E	SP	Pac
SLT	Solitary Is., Australia	14	30.30°S, 153.16°E	SP	Pac
SWT	South West Rocks, Australia	5	30.88°S, 153.06°E	SP	Pac
LDH	Lord Howe Is., Australia	18	31.52°S, 159.07°E	SP	Pac
SRK	Seal Rocks, Australia	12	32.42°S, 152.57°E	SP	Pac
PRS	Port Stephens, Australia	3	32.68°S, 152.19°E	SP	Pac

Table 1. continued

TGL	Terrigal, Australia	5	33.43°S, 151.47°E	SP	Pac
SYD	Sydney, Australia	10	33.88°S, 151.30°E	SP	Pac
JVB	Jervis Bay, Australia	6	35.11°S, 150.78°E	SP	Pac
NAR	Narooma, Australia	4	36.20°S, 150.14°E	SP	Pac
MRB	Merimbula, Australia	10	36.88°S, 149.95°E	SP	Pac

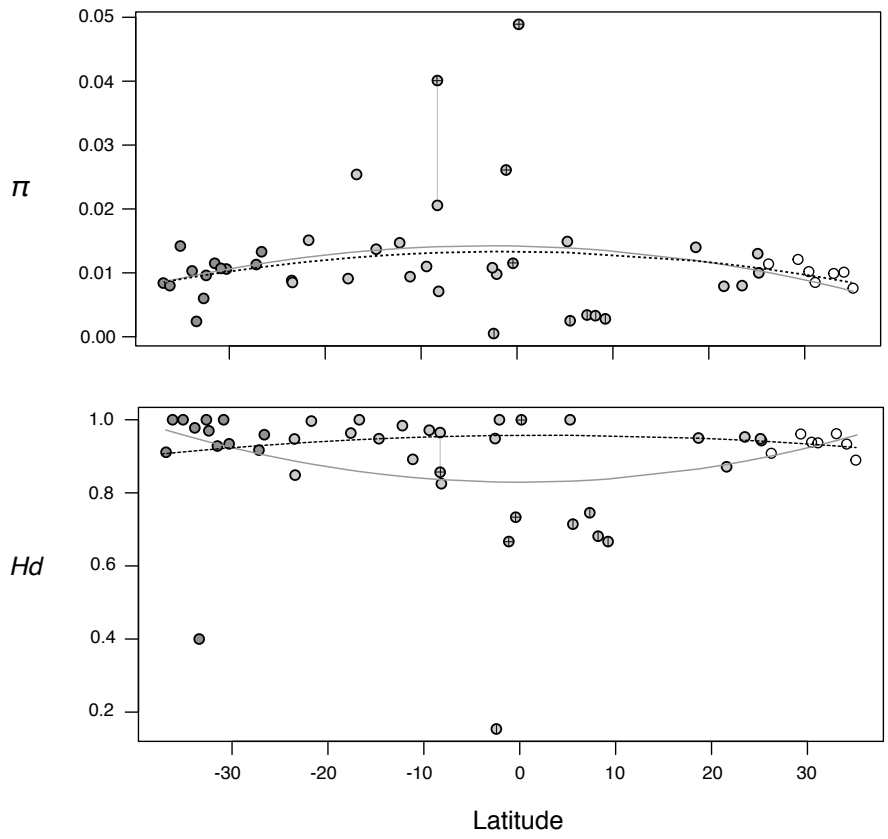


Figure 4. Relationship between latitude and genetic diversity measures (π, *Hd*) for populations of both the Micronesian and Pacific clades (gray line) and the Pacific clade only (dashed line). Lines indicate predicted values from the second order polynomial regression. Clade affinities and range position of populations are designated as in Fig. 2. Thin vertical gray line connects two representations of Timor-Leste, one comprising the Pacific clade only, the other including the Pacific clade and Micronesian clade.

For the Pacific clade, although the peak in π at low latitudes was lower than the curve representing both clades (Fig. 4), the polynomial regression term for curvature was significant ($t_{2,35}$ = -2.561, P = 0.014, Table 2). The ANCOVA analyses also indicated π had a negative relationship with absolute latitude as expected under the CPH ($F_{1,35}$ = 4.806, P = 0.035).

In contrast to the CPH expectations for patterns in genetic diversity, the curve derived from the polynomial regression of *Hd* across the range of the neon damselfish reached its lowest point at low latitudes for the overall dataset ($t_{2,43}$ = 1.868, P = 0.069, Table 2). The dip of the curve for the overall dataset appeared to be due to consistently low *Hd* for populations consisting of the Micronesian clade only, and low *Hd* for a few locations of sympatry (KO and WA, Fig. 4). The ANCOVA also revealed that *Hd* increased with latitude ($F_{1,43}$ = 5.345, P = 0.026, Table 2). When the dataset was reduced to include individuals from the Pacific clade only, a slight peak in *Hd* at low latitudes could be observed but this was not significant in the polynomial regression analysis or the ANCOVA (Fig. 4, Table 2). The ANCOVA analyses and polynomial regressions detected no significant differences in the genetic diversity trends in the northern hemisphere versus the southern hemisphere, and no overall trends from north to south in genetic diversity values, for either dataset (i.e. Micronesian and Pacific clade, and Pacific clade only; Table 2).

Core-periphery patterns of genetic differentiation

According to the expectation under the CPH that peripheral populations should be more differentiated, we expected to find a stronger IBD pattern and a higher intercept in the peripheral regions of the Pacific clade's range than in the core Fig. 1a). Although the pattern of Φ_{ST} differentiation in the NP had a positive relationship to distance that was stronger than the core regions, this was not significant (r_M = 0.356, R^2 = 0.127, P = 0.182). There was no detectable IBD relationship in either of the core sub-regions (near the NP: r_M = -0.119, R^2 = 0.012, P = 0.509; near the SP: r_M = -0.075, R^2 = 0.006, P = 0.665) or in the SP (r_M = -0.050, R^2 = 0.002, P = 0.520). The value of the intercept was higher in the SP than the neighboring core sub-region (0.153 and 0.111, respectively), but this was not mirrored in the north periphery (-0.068 and 0.381, respectively).

Genetic patterns toward the northern and southern periphery

Counter to our expectations based on the CPH, overall haplotypic compositional differences among populations (β_{SOR}) were highest in the core of the species range (β_{SOR} = 0.965), followed by the SP and NP (β_{SOR} = 0.928 and 0.860, respectively; Fig. 5). Partitioning of the β_{SOR} revealed that the composition differences among populations in the core region of the species range were mostly due to turnover (β_{SIM} = 0.954), as were those for the SP, and NP (β_{SIM} = 0.883 and 0.841, respectively; Fig. 5). The greatest contribution via the richness differences (β_{SNE}) was in the SP (β_{SNE} = 0.044) followed by the NP and core (β_{SNE} = 0.019 and 0.010, respectively; Fig. 5). These findings suggest that there is greater contemporary migration among populations near the south periphery than in the north periphery.

Table 2. Table of coefficients for univariate analyses of variance (ANOVA), analyses of covariance (ANCOVA), and second-order polynomial regression (Regression). Left: relationships between clade (Pac/Mic/Sym) and genetic diversity (π, *Hd*), and range position (latitude, hemisphere) and genetic diversity, for the pooled Micronesian and Pacific clades. Right: relationships between range position and genetic diversity for the Pacific clade (Pac) only. Degrees of freedom for *F*– and *t*–ratios are indicated by subscripts. Significance of results, and significance following the exclusion of populations with small sample sizes (n < 10), is indicated as demonstrated by inset key (bottom right). All *t*–values and the significance of relationships were evaluated using orthogonal second-order polynomial regression analyses; however, coefficients are presented on the data scale for interpretation.

		Micronesian + Pacific clades				**Pacific clade**			
		π		***Hd***		**π**		***Hd***	
Analysis	**Factors**	*F* or *t*	*P*	*F* or *t*	*P*	*F* or *t*	*P*	*F* or *t*	*P*
ANOVA	clade	$34.130_{2,43}$	***[10]	$17.160_{2,43}$	***[10]	-		-	
ANCOVA	latitude $_{0\text{-}37°}$	$3.229_{1,43}$	.	$5.345_{1,43}$	*[10]	$4.806_{1,35}$	*	$0.811_{1,35}$	
	hemisphere × latitude $_{0\text{-}37°}$	$0.089_{1,42}$		$0.015_{1,42}$		$0.155_{1,34}$		$0.010_{1,34}$	
	hemisphere	$0.165_{1,43}$		$0.015_{1,43}$		$0.235_{1,35}$		$0.037_{1,35}$	
Regression	latitude $_{37\text{-}0\text{-}37°\ curvature}$	$-1.661_{2,43}$		$1.868_{2,43}$	.[10]	$-2.561_{2,35}$	*[10]	$-0.811_{2,35}$	
	latitude $_{37\text{-}0\text{-}37°\ linear}$	$-0.218_{2,43}$		$-0.307_{2,43}$		$-0.055_{2,35}$		$0.357_{2,35}$	

[10] significant excluding populations with n < 10
*** < 0.0001
** < 0.001

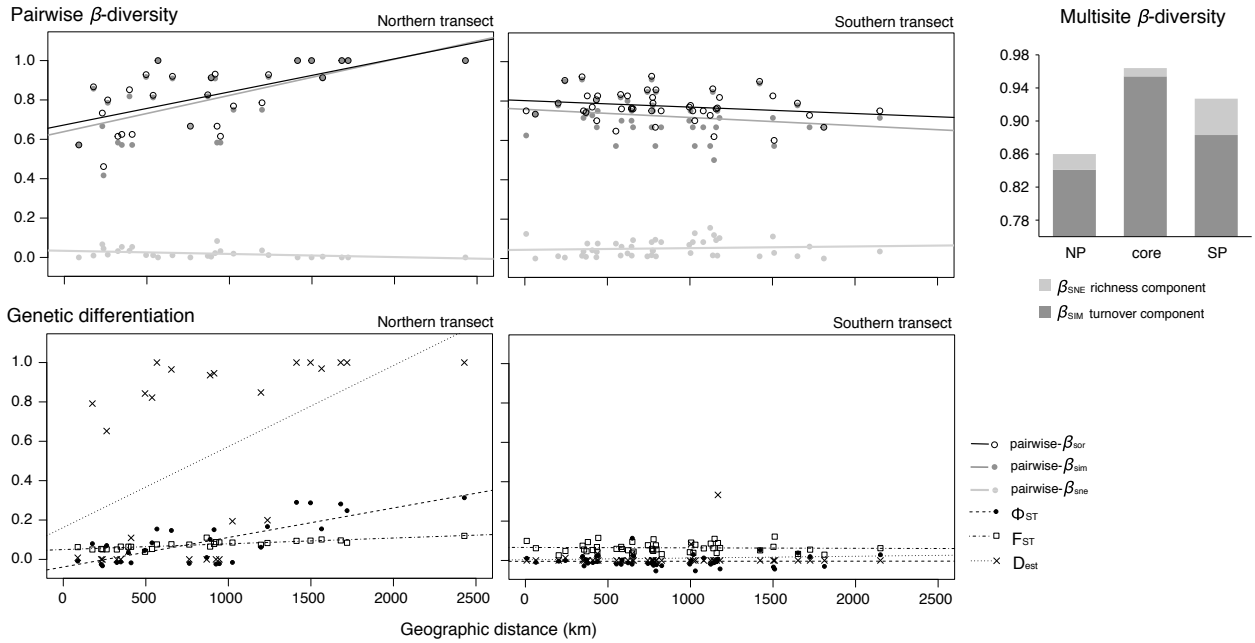


Figure 5. Left: Isolation-by-distance relationships across the northern and southern parts of the Pacific clade's range. Top panel: black circles and black line represent pairwise-β_{sor}, dark grey filled circles and dark grey line represent pairwise-β_{sim}, and light grey filled circles and light gray line represent pairwise-β_{sne}; bottom panel: black filled circles and dashed line represent Φ_{ST}; black squares and dot-dash line represent F_{ST}; and crosses and dotted line represent D_{est}. Right: the turnover (β_{SIM}, dark grey) and richness (β_{SNE}, light grey) components of total multi-site β_{SOR}-diversity (bar height) for the north periphery (NP), core, and south periphery (SP) of the Pacific clade.

Along a transect in the northern part of the range pairwise-β_{sor} had a positive and significant relationship with geographic distance (pairwise-β_{SOR} R^2 = 0.376, r_M = 0.613, P_{perm} = 0.011) as did pairwise-β_{sim} (pairwise-β_{sim} R^2 = 0.375, r_M = 0.612, P_{perm} = 0.013) and all of the genetic differentiation measures (Φ_{ST} R^2 = 0.582, r_M = 0.763, P = 0.001; F_{ST} R^2 = 0.698, r_M = 0.835, P_{perm} = 0.001; D_{est} R^2 = 0.272, r_M = 0.521, P_{perm} < 0.001), whereas the richness component decreased with distance (pairwise-β_{sne} R^2 = 0.160, r_M = -0.400, P_{perm} = 0.946, Fig. 5). Overall, this suggests that the genetic patterns in the northern part of the Pacific clade's range support a scenario of historical colonization. In the southern part of the clade's range none of the genetic differentiation measures (Φ_{ST} R^2 = 0.002, r_M = 0.045, P_{perm} = 0.334; F_{ST} R^2 = 0.009, r_M = -0.093, P_{perm} = 0.624; D_{est} R^2 = 0.025, r_M = -0.157, P_{perm} = 0.907) or pairwise-β measures had a significant relationship with geographic distance (pairwise-β_{sor} R^2 < 0.001, r_M = 0.018, P_{perm} = 0.494; pairwise-β_{sim} R^2 < 0.001, r_M = 0.013, P_{perm} = 0.541; pairwise-β_{sne} R^2 < 0.001, r_M = -0.007, P_{perm} = 0.448, Fig. 5).

The nestedness of the haplotype compositions of populations organized according to latitude was higher along the southern transect of the clade's range than the northern transect ($NODF_{rows}$ = 28.687 and 10.556, respectively). Our randomization test confirmed that the nestedness statistic of the southern transect was significantly higher than would be expected from the chance arrangement of our study populations (P_{perm} = 0.020), whereas the statistic for the northern transect was no different to what may be expected by chance (P_{perm} = 0.336). These results from the southern transect (and those above for β_{SNE} of the SP) fitted with our suggested scenario of asymmetric contemporary migration into the periphery (Fig. 1b).

Discussion

Our approach to characterizing range-wide genetic patterns in a tropical marine species offers a novel understanding of the demographic context and conservation value of peripheral populations than standard analyses used in spatial genetics. For the tropical marine neon damselfish (*Pomacentrus coelestis* species complex), we found that population history, in terms of divergent clades that were in sympatry, confounded our test of the CPH but not in the direction we had anticipated. Lower values for π and *Hd* in the Micronesian clade, and the presence of this clade only in the core of the species range, resulted in lower genetic diversity in the core. Across the range of a single clade (Pacific clade) we also found mixed support for the CPH using conventional genetic measures indicating an increase of π in the core compared with the periphery, but no pattern in *Hd* (Fig. 4) or genetic differentiation (Φ_{ST}) across the latitudinal range. The use of partitioned β-diversity and nestedness measures enabled us to better characterize the spatial arrangement of genetic diversity across the species latitudinal range. We found a strong relationship between geographic distance and pairwise-β_{sor}, its turnover component (pairwise-β_{sim}), and several genetic differentiation measures (Fig, 5) toward the northern periphery – consistent with a historical colonization scenario. In contrast, toward the southern

periphery we found that more of the haplotypic compositional differences among populations (β_{SOR}) were due to richness differences (β_{SNE}, Fig. 5) and that the haplotypic composition of populations was nested with latitude, consistent with asymmetric contemporary migration into the range periphery.

Latitude-wide genetic patterns in the neon damselfish

Historical impacts and population history can strongly influence contemporary patterns of genetic diversity for any taxon. In marine systems the sympatry of divergent lineages or clades has been found to increase genetic diversity in several North Atlantic species (Maggs et al. 2008) and Indo-Pacific reef fishes (Bay and Caley 2011, Gaither et al. 2011). Yet, few studies have considered population history in their tests of the CPH (Eckert et al. 2008, but see Pfeifer et al. 2009, Pinheiro et al. 2011). In our study, the population history of the neon damselfish in the Coral Triangle led to some "populations" to be perceived as having significantly higher genetic diversity than others due to the sympatry of divergent lineages. Surprisingly though, the inherently different levels of genetic diversity within clades had the greatest effect on range-wide genetic patterns. Genetic diversity (π, *Hd*) within the Micronesian clade was significantly lower than in the Pacific clade, obscuring the signature of higher π found in the core of the Pacific clade's range. Based on our results we caution against genetic diversity surveys and testing range-wide hypotheses using datasets of unknown genealogical history (also suggested by Eckert et al. 2008).

Several factors other than those associated with range position have a role in shaping the genetic patterns in the core of the neon damselfish's range. Although the Micronesian clade occurs in the range core, its pure populations are peripheral to the Indo-Australasian-Archipelago (IAA; CE), and largely isolated to the Micronesian Islands (KOR, CH, MO, KW). The low genetic diversity found in this clade is likely to this geographic context. Liu et al. (2012) suggest that the Micronesian Islands were colonized in a stepping-stone manner from the Coral Triangle that would have entailed serial genetic bottlenecks. Moreover, small, isolated populations tend to have low genetic diversity; this is an assumption that underlies the predicted genetic patterns for the range periphery in the CPH, but is evidently not restricted to the range periphery for the neon damselfish species complex.

Although our findings for the Pacific clade were in accordance with the CPH expectations (π peaked in the core of the species range), overall, we did not find compelling support for the hypothesis across the clade's range. We found no latitudinal pattern in *Hd* and populations in the core were more differentiated than predicted by the CPH, and for a high dispersal marine organism (high values of Φ_{ST}, and β_{SOR} and particularly the turnover component, β_{SIM}). Such high genetic differentiation is not unusual among populations in the IAA that may have been separated during periods of the Pleistocene when sea levels were lowered (Voris 2000, Gaither et al. 2011). For example, Liggins et al. in prep. – Chapter Four) found that the Torres Strait land bridge that would have been exposed for much of the Pleistocene period has impacted the patterns of genetic differentiation found across the sampled ranges of three coral reef fishes, including the neon damselfish (Pacific clade only), and two higher

dispersal species (*Halichoeres hortulanus* and *Acanthurus triostegus*). Thus, there are several geographic factors that affect genetic patterns across a seascape other than range position (discussed in Riginos and Liggins 2013 – Chapter Two).

Contrasting patterns in the northern and southern periphery

Our findings describe contrasting genetic patterns toward the latitudinal extremes of the neon damselfish. Contrasting patterns at the northern and southern range peripheries for temperate species may not be surprising given the vastly different environmental conditions organisms are likely to experience (e.g. Hoban et al. 2010, Assis et al. 2013, Hasselman et al. 2013). However for a tropical species such as the neon damselfish we expected similar genetic patterns as, presumably, individuals experience similar environmental conditions at both latitudinal range peripheries. Toward the northern periphery we found support for a scenario of historical colonization followed by relative isolation and time for populations to equilibrate and private haplotypes to arise (low π, higher differentiation, i.e. Φ_{ST}-based IBD; Table 2, Fig. 4, Fig. 5). A recent review of marine phylogeographic studies of this northwestern Pacific region suggests that most species have significant population structure (Ni et al. 2014). Ni et al. (2014) also suggested, that most of the included species had undergone a range expansion, based on network topology, neutrality statistics, and mismatch analyses. Our study extends these genetic diversity patterns; we find that the turnover component of β-diversity (β_{SIM}) can also be high, and the richness component (β_{SNE}) very low (indicative of restricted contemporary migration, Fig. 5).

At the southern range periphery, we find a pattern suggestive of asymmetric contemporary migration (as indicated by a lack of IBD, Hutchison and Templeton 1999, Table 2). Although the southern periphery may have been colonized on a similar time scale as the north, immigration is likely too high for populations to have reached migration-drift equilibrium (Hutchison and Templeton 1999). The East Australian Current that runs north to south along the coastline and is known to drive the recruitment of many organisms (barnacles, Hidas et al. 2013) including the neon damselfish (Booth et al. 2007, Figueira et al. 2009, Booth et al. 2011). Furthermore, this southeast coast of Australia is a hotspot for marine tropical vagrants and species range shifts (Booth et al. 2011, Baird et al. 2012, Feary et al. 2013). The most southerly observed breeding pair of *P. coelestis* is recorded at the Solitary Islands (SLT) beyond which occupancy can be ephemeral, as individuals succumb to particularly harsh over-winter water temperatures in some years (Figueira and Booth 2010). The high levels of genetic diversity in spite of the temporal instability of populations, suggests that immigration into the southern periphery is on-going (termed a "black-hole sink" habitat, Gomulkiewicz et al. 1999).

The pattern of high genetic diversity and low differentiation toward the range periphery that we found in the south of the Pacific clade's range may be common in marine organisms. Such patterns have been described toward the range edge of an urchin species along the same coastline of Australia (Banks et al. 2010). Having a larval dispersal phase theoretically enables benthic marine organisms to

be opportunistic in their colonization and dispersal (Thorson 1971). A predicted consequence of this life history is termed "chaotic patchiness" in their genetic patterns (Johnson and Black 1982). Overtime, a chaotic a pattern of dispersal and recruitment would gradually build the genetic diversity of a population. Such a process of diversity accumulation may be particularly pervasive toward the periphery of a species range where there may be low abundance and/or unoccupied habitat, leading to weak density-dependent selection processes.

The genetic diversity of a population that is heavily dependent on immigration (a sink) should, in theory, be a sub-sample of that found within the source population/s. The partitioning of genetic diversity allowed by the methods herein enable the characterization of such "sub-sets" of genetic diversity across a species range, that are otherwise difficult to distinguish using traditional genetic analyses. The utility of these methods has been demonstrated elsewhere. Diniz-Filho et al. (2012) described an increase in the pairwise nestedness-resultant component of genetic compositions (β_{sne}) along a longitudinal gradient following the range expansion of the Baru tree. Furthermore, Habel et al. (2013) successfully demonstrated using nestedness analysis to characterize genetic patterns following a range expansion of butterflies, reflecting a process of "allele elimination" (Reinig 1938). In our study, multi-site β-diversity revealed that despite the southern periphery having equivalent, or slightly greater, genetic diversity than the northern periphery, a higher proportion of the genetic diversity in this region was shared among populations (indicated by β_{SNE} Fig. 5). Our analysis focused on the nestedness of the population haplotype compositions relative to latitude (predicting higher nestedness with higher latitude) also confirmed that the populations in the southern periphery were nested "sub-sets" of those at lower latitudes. Our study, in concert with previous studies, demonstrates the utility of these methods to understand the organization genetic diversity along spatial gradients (Diniz-Filho and Bini 2011, Diniz-Filho et al. 2012).

Conservation implications

Genetic diversity is frequently quoted as being informative for developing biodiversity management plans and the conservation of species (Laikre et al. 2010). *Hd* and π are the most commonly reported measures of genetic diversity for DNA sequence data, and are thus the most accessible to decision-makers. However, our study indicates instances where these measures of genetic value may be misleading. The sympatry of divergent lineages within the core of the species range enhanced the perceived levels of genetic diversity in some populations and hence their conservation value. Moreover, the low genetic diversity of the Micronesian clade (π and *Hd*), relative to the widespread Pacific clade had a profound influence on the interpreted range-wide genetic patterns. These two clades have recently been described as separate species and are highly unlikely to interbreed (Liu et al. 2013). Therefore, conservation decisions based on these perceived levels of genetic diversity and range-wide patterns would be inappropriate.

There has been much debate about the conservation worth of peripheral populations (Lesica and Allendorf 1995, Beger et al. 2014). The contrasting genetic make-up of peripheral populations found in the north and the south of Pacific clade's range, indicates that not only may this question need to be addressed independently for each species (Guo 2012) but potentially also for each peripheral region within a species range. Although high levels of immigration increase genetic diversity, such genetic swamping can preclude local adaptation (Kirkpatrick and Barton 1997, Kawecki and Holt 2002). Whereas self-recruitment (as might be the case in the northern periphery) can promote the efficacy of local selection, where genotypes suited to local conditions are specifically favored to survive and reproduce (Lenormand 2002). These conditions can also lead to unique genetic variants not found elsewhere across the species range (identified as turnover in our study, Channell and Lomolino 2001, increasing "complementarity" among populations, Araujo 2002). Thus, whereas both peripheral regions of the Pacific clade's range harbor largely equivalent levels of genetic diversity, the unique genetic diversity of the north periphery may be more valued, and is indicative of demographic conditions conducive to local adaptation (Araujo 2002).

Acknowledgements

All fish sampling was undertaken with the authority of The University of Queensland Animal Ethics Committee (Approval Number: SIB/817/08/ARC). Sampling in Timor-Leste was supported by the Coral Triangle Support Partnership and the Ministério da Aquicultura e Pescas, Direcção Nacional de Pescas e Aquicultura (authorized by A Fernandes, L Fontes, J Freitas; guia de marssa: 502/DNPA/VIII/10 and 452/DNPA/VII/11). Export of samples was authorized by the Departmento de Quarentena das Pescas (export permit: 162/FQ006/EXP./DNQB/VII/2011). Sampling in the Solomon Islands was via the Australian Government's Pacific Strategy Assistance Program and with the assistance of the Roviana Conservation Foundation (Solomon Islands Government Ministry of Education and Human Resource Development and Ministry of Fisheries and Marine Resources research permit to S Albert). Sampling in Papua New Guinea was in coordination with the National Research Institute, the Department of Foreign Affairs and Immigration (Research Visa: 10350008304) and the Department of Environment and Conservation (Permit to Export Wildlife: 011318). Authority to sample at Ashmore Reef was provided by the Australian Government Department of Sustainability, Environment, Water, Population and Communities (Access to Biological Resources in a Commonwealth Area for Non-Commercial Purposes permit number: AU-COM2010068) and with logistic support from Australian Customs and border control. Sampling in the Coral Sea was supported by the Marine Division of the Australian Government Department of Sustainability, Environment, Water, Population and Communities (Access to Biological Resources in a Commonwealth Area for Non-Commercial Purposes permit number: AU-COM2008042). Authority to sample at Wigram Island was provided by the Northern Territory Government Department of Resources (Special Permit

Number: 2007-2008/S17/2696). Sampling at Ningaloo Reef was under the authority of the Western Australia Department of Environment and Conservation (License to take Fauna for Scientific Purposes: SF007126, SF006619; Authority to enter calm land/or waters: CE002227, CE002627). We are grateful to the staff of the Australian Museum Lizard Island Research Station and Heron Island Research Station for their facilities and support (Great Barrier Reef Marine Park Authority and Queensland Parks and Wildlife Marine Parks Permit: G08/28114.1, G09/31678.1, G10/33597.1, G11/34640.1; Queensland Government Department of Primary Industries General Fisheries Permit: 118636, 150981; Australian Quarantine Inspection Service Permit to Import Quarantine Material: IP10017966). We especially thank JD Aguirre, J Aini (Ailan Awareness), S Albert, K Davis, M Jimuru, J Keyse, J Kinch (National Fisheries College, Papua New Guinea), W Lovell (Freeflow Dive, Dili), I McLeod, A Mirams, S Penny, R Pinto (and staff of the Coral Triangle Support Partnership), A Smith (Tiki2 Adventure Tours), T Sinclair-Taylor, A Turner, P Waldie, Stephen, Lavud, and Takenda for logistical support and field assistance; G David for laboratory assistance; and H Possingham and JD Aguirre for providing helpful comments on the manuscript. Funding for this work was provided by the Australian Research Council (DP0878306, to CR) and an Explorer's Club Exploration Fund (to LL). LL was supported by an Australian Postgraduate Award from the Australian Government, a Queensland Government Smart Futures PhD Scholarship, and a National Evolutionary Synthesis Center Graduate Fellowship. Many of the ideas discussed here grew out of work funded by the Sea World Research and Rescue Foundation (SWR/1/2012, to CR and LL), a Paddy Pallin Foundation and The Foundation for National Parks and Wildlife Science Grant, an Ecological Society of Australia Student Research Grant, the Lerner Gray Memorial Fund of the American Museum of Natural History, and a Great Barrier Reef Marine Park Authority's Science for Management Award (to LL).

Supplementary methods

Explanation of data integration, data inclusion for analysis, and error/consistency checking

Where the same location was used across multiple studies, we included the location only once (note that summary statistics were consistent across studies, but care was taken not to have pseudo-replication, see Table 1 and Appendix 1 for included locations and the source of data). After removing duplicate locations across studies, potential overlapping locations (e.g. Guihou, Taiwan from Liu et al. 2012 was removed as it was unclear whether it overlapped with Maoao and Yehliu reported in Liu et al. 2010; Australia from Liu et al. 2012 was removed because it may overlap with some of the localities used herein), and locations with very small sample sizes, our dataset for the genetic diversity analyses included 46 populations from across the range of *Pomacentrus coelestis*. Some populations of $n < 10$ from the southern periphery were kept in the dataset as their small sample sizes were indicative of naturally low densities.

All sequences from across the studies were aligned to check for overlap, sequence length consistency, and the likely influence of missing data on across study comparison of genetic diversity (π, *Hd*). Sequences across studies had few sites with missing data and were of comparable length. There was overlap across all polymorphic sites in the combined sequence dataset except two sites at the 3' end of two unique haplotypes identified in Liu et al. (2012). The sequences of Liu et al. (2008, 2010) did not overlap with these sites, whereas our *de novo* sequences overlapped, but the polymorphism was not shared with either our Pacific or Micronesian clade individuals. Given that there was substantial overlap across all sequences, we did not make any correction of genetic diversity or neutrality statistic values for differing sequence length.

To evaluate any inconsistency in our calculation of the summary statistics and genetic differentiation, we reconstructed the sequence dataset of Liu et al. (2008; using haplotypes uploaded on GenBank and the frequency data provided in their supplementary material: 71 haplotypes, 170 individuals, 343bp) and compared our calculated values (π, *Hd*, and Φ_{ST}) with those reported in their original publication. Our recalculated π and *Hd* values for the locations studied by Liu et al. (2008) were largely unchanged from their reported values. For π: two values differed from the original value by 10 and 14%; and for *Hd*: one value differed by less than 0.1%, verifying that our method of calculation was the same and therefore these summary statistics could be compared across studies (inconsistencies between values were likely due to the use of different rounding conventions).

Appendix 1. Location data, source of data, and summary of data included in each of the analyses. Blank cells indicate where there is no data available; '-' indicate where data was omitted for the relevant analysis.

Code	Location	Latitude, longitude	Range position	Clade affinity	*n*	*H*	*Hd*	π	Tajima's *D*	Fu's F_s	IBD analyses testing CPH	IBD analyses of *β*-diversity and nestedness	Source
KN	Kominato, Japan	35.04°N, 140.07°E	NP	Pac	20	12	0.8895	0.0076	-1.3025	-6.2042	NP	northern transect	Liu et al. 2008
OS	Okinoshima Is., Japan	34.09°N, 130.04°E	NP	Pac	29	18	0.9335	0.0101	-1.8308	-10.6354	NP	northern transect	Liu et al. 2008
FS	Funakoshi, Japan	33.02°N, 132.16°E	NP	Pac	15	12	0.9619	0.0099	-1.4282	-7.0533	NP	northern transect	Liu et al. 2008
BS	Bohnotsu, Japan	31.09°N, 131.09°E	NP	Pac	23	14	0.9368	0.0085	-1.0606	-7.6660	NP	northern transect	Liu et al. 2008
TN	Tanegashima Is., Japan	30.43°N, 130.59°E	NP	Pac	26	14	0.9385	0.0102	-1.3984	-5.3999	NP	northern transect	Liu et al. 2008
NS	Nakanoshima, Japan	29.30°N, 129.31°E	NP	Pac	22	16	0.9610	0.0121	-1.0890	-8.7088	NP	northern transect	Liu et al. 2008
SK	Sesoko Is., Japan	26.23°N, 127.31°E	NP	Pac	16	11	0.9083	0.0114	-1.3855	-4.0303	NP	northern transect	Liu et al. 2008
MA	Maoao, Taiwan	25.21°N, 121.73°E	Core	Pac	25	17	0.9430	0.0100	-1.1020	-10.6770	N core		Liu et al. 2010
YL	Yehliu, Taiwan	25.10°N, 122.07°E	Core	Pac	26	19	0.9480	0.0130	-1.2530	-11.2940	N core		Liu et al. 2010
CW	Chinwan Outer Bay, Taiwan	23.47°N, 119.62°E	Core	Pac	23	15	0.9530	0.0080	-0.2630	-9.3650	N core		Liu et al. 2010
TS	Tiaoshi, Taiwan	21.57°N, 120.46°E	Core	Pac	19	12	0.8713	0.0079	-1.8208	-6.3494	N core	northern transect	Liu et al. 2008
HN	Hainan Is., China	18.64°N, 110.67°E	Core	Pac	16	13	0.9500	0.0140	-1.2820	-6.1890	N core		Liu et al. 2010
KW	Kwajalein Atoll, Marshall Is.	9.21°N, 167.47°E	Core	Mic	10	5	0.6667	0.0028					Liu et al. 2012
KOR	Koror, Palau	8.17°N, 134.66°E	Core	Mic	14	6	0.6813	0.0033					Liu et al. 2012
CH	Chuuk Atoll	7.30°N, 151.53°E	Core	Mic	11	4	0.7455	0.0034					Liu et al. 2012
MO	Mortlock Is., Chuuk State	5.50°N, 153.83°E	Core	Mic	7	3	0.7143	0.0025					Liu et al. 2012
TW	Two Fat Thom, Brunei	5.25°N, 115.07°E	Core	Pac	8	8	1.0000	0.0149					Liu et al. 2012
KA	Kawe Is., Indonesia	0.17°N, 130.05°E	Core	Sym	7	7	1.0000	0.0489					Liu et al. 2012
WA	Waigeo Is., Indonesia	0.42°S, 131.30°E	Core	Sym	6	3	0.7333	0.0115					Liu et al. 2012
KO	Kofiau Is., Indonesia	1.14°S, 129.93°E	Core	Sym	10	5	0.6667	0.0261					Liu et al. 2012

Appendix 1. continued

Code	Location	Latitude, longitude	Range position	Clade affinity	*n*	*H*	*Hd*	π	Tajima's *D*	Fu's F_s	IBD analyses testing CPH	IBD analyses of *β*-diversity and nestedness	Source
BO	Boo Kecil, Indonesia	2.12°S, 130.88°E	Core	Pac	3	3	1.0000	0.0098					Liu et al. 2012
CE	Cenderawasih Bay, Indonesia	2.44°S, 135.32°E	Core	Mic	13	2	0.1538	0.0005					Liu et al. 2012
KAV	Kavieng, Papua New Guinea	2.57°S, 150.78°E	Core	Pac	17	13	0.9485	0.0108	-1.5473	-7.0750	-	-	this study
SOL	New Georgia, Solomon Islands	8.17°S, 157.41°E	Core	Pac	16	10	0.8250	0.0071	-1.9916	-4.8478	-	-	this study
TIM_{Sym}	North Coast, Timor L'este	8.30°S, 126.42°E	Core	Sym	40	22	0.8564	0.0401	-	-	-	-	this study
TIM_{Pac}	North Coast, Timor L'este	8.30°S, 126.42°E	Core	Pac	19	16	0.9649	0.0206	-1.2464	-6.8450	-	-	this study
MOT	Motupore Is., Papua New Guinea	9.44°S, 147.24°E	Core	Pac	15	13	0.9714	0.0110	-1.3640	-8.5335	-	-	this study
WIG	Wigram Is., Australia	11.15°S, 136.59°E	Core	Pac	16	10	0.8917	0.0094	-1.8136	-3.5870	-	-	this study
ASH	Ashmore Reef, Australia	12.24°S, 123.13°E	Core	Pac	32	27	0.9839	0.0147	-1.0390	-23.3781	-	-	this study
LIZ	Lizard Is., Australia	14.69°S, 145.49°E	Core	Pac	22	16	0.9481	0.0137	-0.7652	-7.7530	S core	-	Mirams et al. 2011, this study
FI	Fiji	16.73°S, 177.67°E	Core	Pac	3	3	1.0000	0.0254					Liu et al. 2012
LIH	Lihou Reef, Australia	17.62°S, 151.53°E	Core	Pac	11	9	0.9636	0.0091	-1.0115	-4.5789	S core	southern transect	Mirams et al. 2011
NIN	Ningaloo Reef, Australia	21.73°S, 113.98°E	Core	Pac	24	23	0.9964	0.0151	-1.4257	-21.8343	-	-	Mirams et al. 2011, this study
OTR	One Tree Is., Australia	23.42°S, 151.90°E	Core	Pac	12	8	0.8485	0.0085	-1.1895	-2.7362	S core	southern transect	this study
HER	Heron Is., Australia	23.49°S, 151.93°E	Core	Pac	20	16	0.9474	0.0088	-1.7637	-12.8862	S core	southern transect	Mirams et al. 2011
MOO	Mooloolaba, Australia	26.64°S, 153.12°E	SP	Pac	19	15	0.9591	0.0133	-1.4973	-7.7519	SP	southern transect	this study
MTB	Moreton Bay, Australia	27.19°S, 153.25°E	SP	Pac	23	15	0.9170	0.0113	-0.8821	-7.0937	SP	southern transect	this study
SLT	Solitary Is., Australia	30.30°S, 153.16°E	SP	Pac	14	11	0.9341	0.0106	-1.7389	-5.5284	SP	southern transect	this study
SWT	South West Rocks, Australia	30.88°S, 153.06°E	SP	Pac	5	5	1.0000	0.0107	-0.1901	-1.9011	SP	southern transect	this study
LDH	Lord Howe Is., Australia	31.52°S, 159.07°E	SP	Pac	18	13	0.9281	0.0115	-1.1129	-6.1042	SP	southern transect	this study
SRK	Seal Rocks, Australia	32.42°S, 152.57°E	SP	Pac	12	10	0.9697	0.0096	-1.0752	-5.5172	SP	southern transect	this study

Appendix 1. continued

Code	Location	Latitude, longitude	Range position	Clade affinity	*n*	*H*	*Hd*	π	Tajima's *D*	Fu's F_s	IBD analyses testing CPH	IBD analyses of *β*-diversity and nestedness	Source
PRS	Port Stephens, Australia	32.68°S, 152.19°E	SP	Pac	3	3	1.0000	0.0060	0.0000	-0.6932	SP	southern transect	this study
TGL	Terrigal, Australia	33.43°S, 151.47°E	SP	Pac	5	2	0.4000	0.0024	-0.9726	1.0404	SP	southern transect	this study
SYD	Sydney, Australia	33.88°S, 151.30°E	SP	Pac	10	9	0.9778	0.0103	-0.7159	-4.9334	SP	southern transect	this study
JVB	Jervis Bay, Australia	35.11°S, 150.78°E	SP	Pac	6	6	1.0000	0.0142	-0.1057	-2.2702	SP	southern transect	this study
NAR	Narooma, Australia	36.20°S, 150.14°E	SP	Pac	4	4	1.0000	0.0080	-0.2125	-1.4142	SP	southern transect	this study
MRB	Merimbula, Australia	36.88°S, 149.95°E	SP	Pac	10	7	0.9111	0.0084	-1.0449	-2.2221	SP	southern transect	this study

Appendix 2. a. Pairwise Φ_{ST} values and p–values for the Pacific clade. **b.** Pairwise F_{ST} values and p–values for the northern and southern transects of the Pacific clade. **c.** Pairwise D_{est} values for the northern and southern transects of the Pacific clade. P–values are adjusted for multiple pairwise comparisons using the method of Benjamini and Hochberg (1995; "BH" or its alias "fdr") implemented in the R base statistical package.

a. Pairwise Φ_{ST} (lower triangle) and p–values (upper triangle) for the Pacific clade - calculated *de novo.*

	NIN	ASH	TIM	ENG	KAV	MOT	SOL	LIZ	LIH	HER	OTR	TS	MOO	MTB	SLT	SWT	LDH	SRK	PRS	TGL	NAR
NIN	-	0.012	0.018	0.000	0.004	0.144	0.000	0.049	0.000	0.000	0.133	0.000	0.019	0.002	0.007	0.007	0.004	0.011	0.154	0.027	0.042
ASH	0.066	-	0.651	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.027	0.009	0.022
TIM	0.057	-0.006	-	0.002	0.000	0.006	0.000	0.002	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.004	0.000	0.002	0.070	0.009	0.038
ENG	0.170	0.128	0.110	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000
KAV	0.107	0.233	0.217	0.442	-	0.623	0.555	0.340	0.198	0.876	0.651	0.000	0.822	0.301	0.817	0.374	0.947	0.515	0.983	0.999	0.951
MOT	0.028	0.132	0.131	0.339	-0.004	-	0.103	0.787	0.029	0.149	0.876	0.000	0.975	0.313	0.681	0.110	0.706	0.355	0.771	0.640	0.561
SOL	0.173	0.285	0.278	0.531	-0.001	0.044	-	0.055	0.071	0.660	0.235	0.000	0.203	0.064	0.695	0.406	0.366	0.400	0.957	0.999	0.628
LIZ	0.045	0.131	0.132	0.287	0.011	-0.018	0.061	-	0.099	0.098	0.618	0.000	0.890	0.660	0.406	0.246	0.765	0.432	0.787	0.560	0.731
LIH	0.217	0.307	0.285	0.533	0.032	0.101	0.070	0.060	-	0.274	0.021	0.000	0.184	0.355	0.080	0.434	0.340	0.189	0.985	0.704	0.958
HER	0.153	0.259	0.262	0.488	-0.015	0.030	-0.004	0.036	0.022	-	0.371	0.000	0.496	0.194	0.706	0.534	0.761	0.825	0.994	0.999	0.990
OTR	0.037	0.179	0.170	0.397	-0.011	-0.027	0.029	-0.007	0.113	0.011	-	0.000	0.752	0.260	0.900	0.210	0.643	0.563	0.612	0.636	0.352
TS	0.386	0.173	0.161	0.431	0.587	0.517	0.650	0.476	0.654	0.611	0.589	-	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.004
MOO	0.064	0.162	0.159	0.341	-0.015	-0.027	0.020	-0.020	0.030	0.002	-0.014	0.502	-	0.714	0.724	0.399	0.922	0.840	0.971	0.913	0.949
MTB	0.114	0.216	0.218	0.404	0.013	0.017	0.052	-0.008	0.013	0.023	0.025	0.560	-0.009	-	0.240	0.433	0.790	0.480	0.897	0.628	0.900
SLT	0.096	0.210	0.205	0.427	-0.014	-0.009	-0.007	0.008	0.053	-0.007	-0.025	0.576	-0.011	0.021	-	0.523	0.840	0.528	0.990	0.999	0.745
SWT	0.170	0.286	0.250	0.532	0.022	0.083	0.017	0.050	0.015	0.010	0.058	0.655	0.017	0.011	0.005	-	0.467	0.350	0.978	0.787	0.822
LDH	0.094	0.194	0.191	0.402	-0.025	-0.011	0.012	-0.015	0.019	-0.012	-0.010	0.549	-0.021	-0.015	-0.020	0.008	-	0.717	0.986	0.957	0.957
SRK	0.109	0.204	0.203	0.440	-0.002	0.013	0.014	0.004	0.038	-0.019	-0.001	0.582	-0.019	0.003	0.000	0.030	-0.012	-	0.900	0.840	0.927
PRS	0.098	0.230	0.181	0.522	-0.120	-0.036	-0.096	-0.062	-0.158	-0.141	-0.025	0.661	-0.107	-0.097	-0.130	-0.175	-0.133	-0.098	-	0.739	0.996
TGL	0.130	0.259	0.228	0.553	-0.083	-0.005	-0.097	-0.003	-0.025	-0.097	-0.015	0.684	-0.051	-0.022	-0.086	-0.019	-0.067	-0.050	-0.169	-	0.616
NAR	0.111	0.228	0.190	0.500	-0.071	-0.010	-0.012	-0.035	-0.095	-0.081	0.021	0.647	-0.074	-0.069	-0.030	-0.053	-0.076	-0.068	-0.180	-0.053	-
SYD	0.122	0.243	0.217	0.471	-0.016	0.014	0.018	0.006	-0.030	-0.006	0.007	0.614	-0.026	-0.017	-0.006	-0.008	-0.016	-0.004	-0.137	-0.063	-0.094
JVB	-0.002	0.102	0.052	0.196	0.151	0.077	0.265	0.060	0.279	0.230	0.085	0.513	0.086	0.165	0.155	0.237	0.138	0.166	0.181	0.262	0.180
MRB	0.077	0.197	0.184	0.441	-0.048	-0.034	-0.018	-0.024	0.029	-0.033	-0.044	0.603	-0.043	-0.019	-0.053	-0.010	-0.053	-0.018	-0.150	-0.089	-0.060
BS	0.181	0.080	0.049	0.268	0.414	0.301	0.488	0.296	0.518	0.463	0.397	0.287	0.326	0.395	0.409	0.516	0.377	0.444	0.505	0.525	0.482
FS	0.182	0.065	0.033	0.221	0.410	0.300	0.490	0.286	0.507	0.462	0.394	0.248	0.319	0.389	0.400	0.491	0.372	0.432	0.485	0.520	0.466
KN	0.188	0.080	0.046	0.254	0.429	0.316	0.515	0.296	0.532	0.479	0.418	0.313	0.333	0.400	0.426	0.543	0.386	0.466	0.536	0.560	0.510
NS	0.135	0.010	-0.002	0.098	0.380	0.270	0.452	0.251	0.459	0.419	0.334	0.167	0.289	0.354	0.364	0.451	0.339	0.369	0.417	0.446	0.401
OS	0.209	0.104	0.069	0.253	0.412	0.308	0.476	0.302	0.498	0.458	0.391	0.281	0.335	0.396	0.403	0.491	0.376	0.439	0.472	0.491	0.455
SK	0.268	0.072	0.059	0.264	0.481	0.390	0.550	0.363	0.546	0.514	0.463	0.009	0.390	0.458	0.463	0.535	0.440	0.468	0.518	0.554	0.506
TN	0.187	0.091	0.058	0.230	0.395	0.287	0.462	0.280	0.485	0.442	0.368	0.290	0.315	0.377	0.383	0.471	0.358	0.421	0.455	0.476	0.441

Appendix 2. a. continued

	SYD	JVB	MRB	BS	FS	KN	NS	OS	SK	TN	BS
NIN	0.012	0.557	0.039	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.039
ASH	0.000	0.065	0.002	0.009	0.029	0.012	0.332	0.000	0.033	0.002	0.002
TIM	0.002	0.183	0.002	0.033	0.107	0.072	0.563	0.004	0.053	0.007	0.002
ENG	0.000	0.002	0.000	0.000	0.002	0.000	0.014	0.000	0.000	0.000	0.000
KAV	0.746	0.014	0.999	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999
MOT	0.393	0.125	0.916	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.916
SOL	0.340	0.002	0.765	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.765
LIZ	0.459	0.152	0.768	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.768
LIH	0.840	0.006	0.272	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.272
HER	0.632	0.007	0.979	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.979
OTR	0.427	0.160	0.957	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.957
TS	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.371	0.000	0.000
MOO	0.897	0.078	0.999	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999
MTB	0.724	0.024	0.757	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.757
SLT	0.628	0.012	0.999	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999
SWT	0.618	0.032	0.641	0.000	0.002	0.000	0.000	0.000	0.000	0.002	0.641
LDH	0.714	0.039	0.999	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999
SRK	0.557	0.029	0.714	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.714
PRS	0.999	0.134	0.983	0.000	0.002	0.004	0.000	0.004	0.000	0.007	0.983
TGL	0.927	0.027	0.974	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.974
NAR	0.978	0.116	0.856	0.000	0.000	0.000	0.000	0.000	0.002	0.002	0.856
SYD	-	0.036	0.876	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.876
JVB	0.160	-	0.064	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.064
MRB	-0.028	0.151	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-
BS	0.448	0.310	0.413	-	0.982	0.929	0.019	0.889	0.000	0.742	0.413
FS	0.438	0.285	0.407	-0.033	-	0.876	0.089	0.992	0.012	0.850	0.407
KN	0.466	0.321	0.433	-0.022	-0.020	-	0.049	0.999	0.000	0.882	0.433
NS	0.394	0.187	0.360	0.070	0.046	0.062	-	0.004	0.145	0.006	0.360
OS	0.437	0.308	0.398	-0.013	-0.025	-0.024	0.084	-	0.000	0.974	0.398
SK	0.497	0.352	0.475	0.147	0.102	0.155	0.035	0.151	-	0.000	0.475
TN	0.423	0.285	0.378	-0.009	-0.016	-0.015	0.080	-0.017	0.154	-	0.378

Pairwise Φ_{ST} (lower triangle) and p–values (upper triangle) - as reported in Liu et al. (2010) and included in isolation-by-distance analyses herein.

	YL	MA	CW	HN
YL	-	P > 0.05	P > 0.05	P < 0.01
MA	0.003	-	P > 0.05	P < 0.01
CW	0.007	-0.013	-	P < 0.01
HN	0.087	0.186	0.226	-

Appendix 2. b. Pairwise F_{ST} (lower triangle) and p–values (upper triangle) for the northern transect of the Pacific clade.

	BS	FS	KN	NS	OS	SK	TN	TS
BS	-	0.010	0.000	0.001	0.000	0.000	0.000	0.000
FS	0.051	-	0.008	0.007	0.013	0.009	0.004	0.005
KN	0.087	0.075	-	0.000	0.000	0.001	0.000	0.000
NS	0.051	0.039	0.074	-	0.001	0.001	0.000	0.001
OS	0.065	0.053	0.088	0.053	-	0.000	0.000	0.000
SK	0.077	0.065	0.101	0.065	0.078	-	0.000	0.001
TN	0.062	0.050	0.085	0.050	0.064	0.076	-	0.000
TS	0.095	0.084	0.120	0.083	0.096	0.111	0.094	-

Pairwise F_{ST} (lower triangle) and p–values (upper triangle) for the southern transect of the Pacific clade.

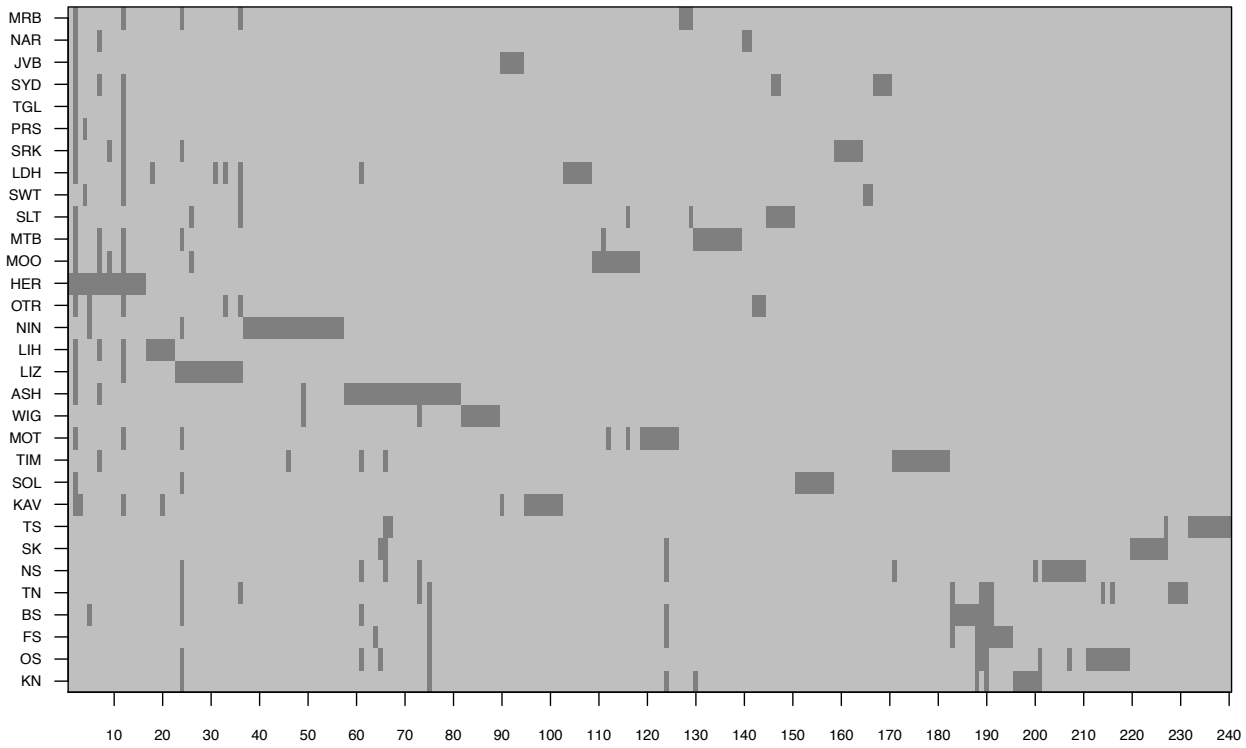
	LIH	HER	OTR	MOO	MTB	SLT	SWT	LDH	SRK	PRS	TGL	NAR	SYD	JVB	MRB
LIH	-	0.024	0.020	0.035	0.020	0.050	0.420	0.020	0.076	0.456	0.009	0.428	0.179	0.365	0.024
HER	0.045	-	0.007	0.008	0.004	0.013	0.420	0.008	0.018	0.505	0.006	0.446	0.089	0.390	0.008
OTR	0.095	0.099	-	0.008	0.008	0.008	0.228	0.008	0.021	0.390	0.008	0.354	0.042	0.187	0.016
MOO	0.039	0.047	0.093	-	0.005	0.013	0.351	0.007	0.041	0.515	0.008	0.397	0.070	0.320	0.018
MTB	0.061	0.068	0.115	0.062	-	0.008	0.226	0.000	0.018	0.378	0.007	0.320	0.031	0.156	0.013
SLT	0.052	0.059	0.108	0.053	0.075	-	0.390	0.013	0.035	0.476	0.008	0.424	0.080	0.367	0.019
SWT	0.020	0.031	0.087	0.024	0.049	0.038	-	0.310	0.425	1.000	0.089	1.000	0.564	1.000	0.259
LDH	0.055	0.062	0.110	0.056	0.078	0.069	0.042	-	0.023	0.390	0.004	0.330	0.059	0.266	0.013
SRK	0.033	0.042	0.091	0.036	0.058	0.048	0.017	0.052	-	0.473	0.008	0.445	0.175	0.425	0.020
PRS	0.023	0.035	0.099	0.027	0.055	0.043	0.000	0.047	0.019	-	0.134	1.000	0.659	1.000	0.405
TGL	0.262	0.250	0.323	0.244	0.263	0.268	0.300	0.263	0.255	0.366	-	0.090	0.013	0.034	0.008
NAR	0.021	0.032	0.092	0.025	0.052	0.040	0.000	0.044	0.018	0.000	0.326	-	0.590	1.000	0.266
SYD	0.029	0.038	0.088	0.032	0.055	0.045	0.012	0.049	0.026	0.014	0.259	0.013	-	0.550	0.092
JVB	0.020	0.030	0.084	0.023	0.048	0.037	0.000	0.041	0.017	0.000	0.283	0.000	0.012	-	0.226
MRB	0.062	0.069	0.121	0.063	0.086	0.077	0.050	0.080	0.059	0.057	0.296	0.052	0.056	0.048	-

Appendix 2. c. Pairwise D_{est} values (lower triangle) for the northern transect of the Pacific clade.

	BS	FS	KN	NS	OS	SK	TN	TS
BS	-	-	-	-	-	-	-	-
FS	0.000	-	-	-	-	-	-	-
KN	0.000	0.000	-	-	-	-	-	-
NS	0.652	0.843	0.848	-	-	-	-	-
OS	0.000	0.000	0.000	0.822	-	-	-	-
SK	0.965	0.936	0.969	0.040	0.945	-	-	-
TN	0.008	0.000	0.194	0.791	0.109	1.000	-	-
TS	1.000	1.000	1.000	0.199	1.000	0.000	1.000	-

Pairwise D_{est} values (lower triangle) for the southern transect of the Pacific clade.

	LIH	HER	OTR	MOO	MTB	SLT	SWT	LDH	SRK	PRS	TGL	NAR	SYD	JVB	MRB
LIH	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HER	0.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OTR	0.035	0.000	-	-	-	-	-	-	-	-	-	-	-	-	-
MOO	0.000	0.000	0.000	-	-	-	-	-	-	-	-	-	-	-	-
MTB	0.000	0.000	0.000	0.000	-	-	-	-	-	-	-	-	-	-	-
SLT	0.000	0.000	0.000	0.000	0.001	-	-	-	-	-	-	-	-	-	-
SWT	0.000	0.300	0.588	0.051	0.230	0.597	-	-	-	-	-	-	-	-	-
LDH	0.000	0.000	0.000	0.000	0.000	0.000	0.144	-	-	-	-	-	-	-	-
SRK	0.000	0.000	0.083	0.000	0.000	0.013	0.000	0.000	-	-	-	-	-	-	-
PRS	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-	-	-	-	-	-
TGL	0.000	0.307	0.029	0.362	0.235	0.270	0.867	0.222	0.437	0.000	-	-	-	-	-
NAR	0.000	0.000	0.000	0.000	0.000	0.000	2.000	0.000	0.000	0.000	0.354	-	-	-	-
SYD	0.000	0.000	0.333	0.000	0.000	0.035	0.000	0.000	0.000	0.000	0.592	0.000	-	-	-
JVB	0.000	0.000	0.125	0.000	0.020	0.000	0.000	0.000	0.000	0.000	0.546	0.000	0.000	-	-
MRB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.150	0.000	0.000	0.000	-



Appendix 3. Table of haplotype presence and absence (columns) at each study location (rows; organized according to latitude) for the Pacifc clade.

CHAPTER SIX. Evaluating edge-of-range genetic patterns for tropical echinoderms, *Acanthaster planci* and *Tripneustes gratilla*, of the Kermadec Islands, southwest Pacific

Published in Bulletin of Marine Science 90: 379–397. 2014

Libby Liggins, Lachlan Gleeson and Cynthia Riginos

School of Biological Sciences, The University of Queensland

Abstract

Edge-of-range populations are often typified by patterns of low genetic diversity and high genetic differentiation relative to populations within the core of a species range. The "core-periphery hypothesis," also known as the "central-marginal hypothesis," predicts that these genetic patterns at the edge-of-range are a consequence of reduced population size and connectivity toward a species range periphery. It is unclear, however, how these expectations relate to high dispersal marine species that can conceivably maintain high abundance and high connectivity at their range edge. In the present study, we characterize the genetic patterns of two tropical echinoderm populations in the Kermadec Islands, the edge of their southwest Pacific range, and compare these genetic patterns to those from populations throughout their east Indian and Pacific ranges. We find that the populations of both *Acanthaster planci* (Linnaeus, 1758) and *Tripneustes gratilla* (Linnaeus, 1758) are represented by a single haplotype at the Kermadec Islands (based on mitochondrial cytochrome oxidase c subunit I). Such low genetic diversity concurs with the expectations of the "core-periphery hypothesis." Furthermore, the haplotypic composition of both populations suggests they have been founded by a small number of colonists with little subsequent immigration. Thus, local reproduction and self-recruitment appear to maintain these populations despite the ecologically marginal conditions of the Kermadec Islands for these tropical species. Understanding rates of self-recruitment versus reliance on connectivity with populations outside of the Kermadec Islands has implications for the persistence of these populations and range stability of these echinoderm species.

Introduction

Population attributes are expected to differ according to range position. Differences in abundance, reproduction, and dispersal (connectivity) can often be observed across a species range, and over evolutionary timescales these population attributes are expected to lead to predictable neutral genetic patterns across their range. For example, the "core-periphery hypothesis" (hereafter CPH) predicts

that populations at the periphery of a species range will have lower genetic diversity and be more genetically differentiated as a consequence of smaller effective population sizes (as predicted by the "abundant-center hypothesis") and lower connectivity than core populations (Eckert et al. 2008). Although support for the CPH has been equivocal in the empirical literature (reviewed in Eckert et al. 2008), most tests have focused on terrestrial organisms and thus this theory's relevance in marine systems remains unclear.

Some marine genetic studies support the applicability of the CPH (e.g., urchins, Palumbi et al. 1997; coral, Ayre and Hughes 2004; algae, Faugeron et al. 2004); however, several do not (detailed below). There are many reasons why range-wide genetic patterns of marine organisms may differ from terrestrial organisms (Liggins et al. 2013 – Chapter Three). Most notably, marine species often have high dispersal potential influencing their ability to colonize, expand their ranges, and maintain connectivity among established populations (Kinlan and Gaines 2003, Mora et al. 2011). Consequently, population abundance remains high toward the range edge of many marine species (e.g., flatfish, Leggett and Frank 1997; mole crab, Defeo and Cardoso 2004; intertidal limpet, Gilman 2005) and genetic diversity as well as connectivity can be maintained in peripheral populations by immigration, contradicting the expectations of the CPH (e.g., barnacle, Dawson et al. 2010; coral, Nakajima et al. 2012; coral reef fishes, Bay and Caley 2011).

For any population to establish and persist, it must rely on some combination of immigration and self-recruitment (Table 1). Peripheral populations are often ecologically marginal (Lesica and Allendorf 1995) and thus by definition have limited reproduction and are heavily reliant on immigration (Barton 2001). For high dispersal marine species, there is the very real possibility that some peripheral populations are sustained by immigration and that self-recruitment is insufficient for long term population persistence. In the extreme, high levels of immigration into peripheral populations may cause levels of genetic diversity to be similar to those of core populations, contradicting the expected pattern of decreased diversity for peripheral populations under the CPH (see "Migration load" scenario in Table 1). Furthermore, frequent immigration and low levels of reproduction and self-recruitment could lead peripheral populations to be less genetically differentiated than core populations also contradicting the CPH (see "Metapopulation" scenario in Table 1). Thus, the influence of immigration and local reproductive success, often determined by ecological circumstances, can promote patterns of genetic diversity and differentiation that do not conform to the expectations for a peripheral population according to the CPH.

Here we focus on the Kermadec Islands, a remote chain of eleven volcanic islands (29°15′s–31°21′s and 177°55′W–178°48′W), between Tonga and New Zealand in the southwest Pacific. The islands are geologically young, having originated only 1.8–3 Ma and provide marginal habitat for tropical marine species (Watt 1975). Despite falling within New Zealand's largest marine reserve

Table 1. Predicted genetic patterns of peripheral *Acanthaster planci* and *Tripneustes gratilla* populations at the Kermadec Islands under extreme competing scenarios. Intermediate scenarios and predictions are possible (i.e., multiple colonization events, low levels of ongoing immigration).

		IMMIGRATION	
		Rare	**Frequent**
SELF-RECRUITMENT	**Minimal**	*Ephemeral population – neither immigration or self-recruitment are enough to sustain a population* (not applicable to the current investigation)	*Metapopulation – peripheral population relies on immigration for persistence* Genetic diversity – low compared to core populations Genetic differentiation – may differ from core populations Genetic novelty – none
	High	*Colonization – peripheral population is founded by a small group of colonists and is self-sustaining* Genetic diversity – low compared to core populations Genetic differentiation – likely to differ from core populations Genetic novelty – possible especially if colonization is old	*Migration load – peripheral population is maintained by high gene flow between immigrants and local individuals* Genetic diversity – similar to core populations Genetic differentiation – similar to core populations Genetic novelty – none

(Gardner et al. 2006) the transitional coral-algal community of the islands is relatively understudied. Genetic patterns and affinities of the islands' marine biota have mostly been addressed above the species level (e.g., algae, Heesch et al. 2009; coral symbionts, Wicks et al. 2010b), or for subtropical (neritid snail, Spencer et al. 2007) and endemic species (limpets, Wood and Gardner 2007). Recent research has highlighted the diversity and abundance of the tropical marine fauna (Richards and Liggins in press. - Appendix Four), yet there has been little population genetic investigation of resident tropical taxa (but see Vogler et al. 2013, discussed below). Here, we investigate how this island group's characteristics (peripheral and ecologically marginal) influence genetic patterns in two tropical echinoderms: the crown-of-thorns starfish, *Acanthaster planci* (Linnaeus 1758); and the collector urchin, *Tripneustes gratilla* (Linnaeus 1758).

Acanthaster planci and *T. gratilla* are common in tropical, shallow reef habitat throughout the Indian and Pacific Oceans. *Acanthaster planci* occurs as far south as Lord Howe Island (31°32′S) and the Kermadec Islands (Macauley Island, 30°14′S; Francis et al. 1987). In contrast, *T. gratilla* has been recorded in small numbers in the northeast of mainland New Zealand (C Duffy, New Zealand Department of Conservation, pers. comm.). Both species broadcast spawn and have high dispersal potential. The pelagic larvae of *T. gratilla* are known to disperse for at least 18 d (Mortensen 1937), and those of *A. planci* typically survive up to 28 d in the pelagic environment (Yamaguchi 1973). Since the first published records of *A. planci* (McKnight 1978) and *T. gratilla* [(Farquhar 1897); formerly *Tripneustes variegatus* (Leske 1778)] at the Kermadec Islands, both species have been repeatedly recorded at the island group (*A. planci*: Schiel et al. 1986, Francis et al. 1987, Cole et al. 1992, Brook 1999, Gardner et al. 2006; *T. gratilla*: McKnight 1968, Schiel et al. 1986, Cole et al. 1992, Gardner et al. 2006), including observations of new recruits (Appendix Four).

Acanthaster planci and *T. gratilla* are particularly abundant around the islets of Raoul Island (the northernmost island of the Kermadec Islands); the density of *A. planci* was recently estimated to be >80 individuals per hectare (Richards and Liggins in press – Appendix Four) and *T. gratilla* was present in an estimated similar abundance (L Liggins unpubl. data). Such densities are typical of both *A. planci* (reviewed in Birkeland and Lucas 1990) and *T. gratilla* (reviewed in Lawrence and Agatsuma 2001) in the core of their ranges. The high dispersal potential of these species and their abundance within a peripheral population suggest the CPH may not apply to the Kermadec populations of *T. gratilla* and *A. planci*. Furthermore, the marginal environment of the Kermadec Islands may cause these populations of tropical echinoderms to have poor reproduction and self-recruitment, acting as a demographic "sink," which would also disrupt typical expectations for an edge-of-range population under the CPH.

Here we examine genetic patterns of *A. planci* and *T. gratilla* from the Kermadec Islands and compare these observations against the predictions of the CPH. In the context of the Kermadec Islands, the relative influence of immigration (i.e., dispersal to, and survival in, the Kermadec population) and

self-recruitment is expected to leave predictable patterns of genetic diversity, genetic differentiation, and genetic novelty (Table 1). A scenario of "colonization" whereby immigration is rare and self-recruitment is high would result in a population with the expected patterns of the CPH. However, low self-recruitment ("Metapopulation" scenario) or high immigration ("Migration load" scenario, Table 1) could cause genetic patterns to deviate from CPH expectations.

Methods

The study region included the known range of the Pacific clade of *A. planci* (Vogler et al. 2008; herein referred to as *A. planci*) and the corresponding range of *T. gratilla* so that the genetic patterns of the Kermadec Islands could be put into a broader east Indian-Pacific Ocean context (Fig. 1). Tube feet (*A. planci*) and gonad tissue (*T. gratilla*) were hand collected while on snorkel or scuba from three (LZ, KE, SO) and six (LZ, KE, KV, MO, PG, SO) locations (see Fig. 1 for location details), respectively, to complement existing sequence datasets found on GenBank (Appendix 1). The Kermadec samples were taken from reef surrounding islets to the northeast of Raoul Island: Dayrell Island, west of Meyer Island (*A. planci*), Egeria rock, and east of Meyer Island (*T. gratilla*). Total genomic DNA was extracted from collected tissue using a salt extraction method (modified from Aljanabi and Martinez 1997). The mitochondrial cytochrome oxidase c subunit I (COI) was amplified using COTS_COI_F4734 and COTS_COI_R5433 in *A. planci* (Vogler et al. 2008), and COIp or COIf and COIa in *T. gratilla* (Lessios et al. 2003). Amplicons were purified using Exonuclease I and Antarctic Phosphatase following the Exo-SAP protocol (New England Biolabs) and sequenced by Macrogen (Korea) via capillary electrophoresis.

Sequences were manually checked and edited using CodonCode Aligner v3.7.1.2 (CodonCode corporation) and aligned with existing COI data sets using Se-AL v2.0a11 (Rambaut 1996). The aligned sequences were translated into amino acid sequences using the invertebrate mitochondrial code to ensure they were not of nuclear origin and a BLAST search against sequences on the GenBank database confirmed their species origin. The primer sequence and regions of insignificant overlap at either end of the sequences were deleted in Se-AL, so that all sequences within each species dataset were of a common length.

To describe patterns of genetic diversity, differentiation, and novelty of the Kermadec populations in relation to other populations across the studied ranges, analysis of DNA polymorphism (polymorphic sites θ, nucleotide diversity π, number of haplotypes H, and diversity of haplotypes *Hd*), Tajima's *D* test (Tajima 1989, Tajima 1996), pairwise Φ_{ST} (Tamura-Nei corrected), and pairwise F_{ST} measures were carried out using Arlequin v3.5.1.2 (Excoffier and Lischer 2010; all with 10,000 permutations). P–values for all pairwise measures were adjusted for multiple comparisons using the method of Benjamini and Hochberg (Benjamini and Hochberg 1995; "BH" or its alias "fdr") implemented in the R statistical package v2.15.2 (R Core Team 2012).

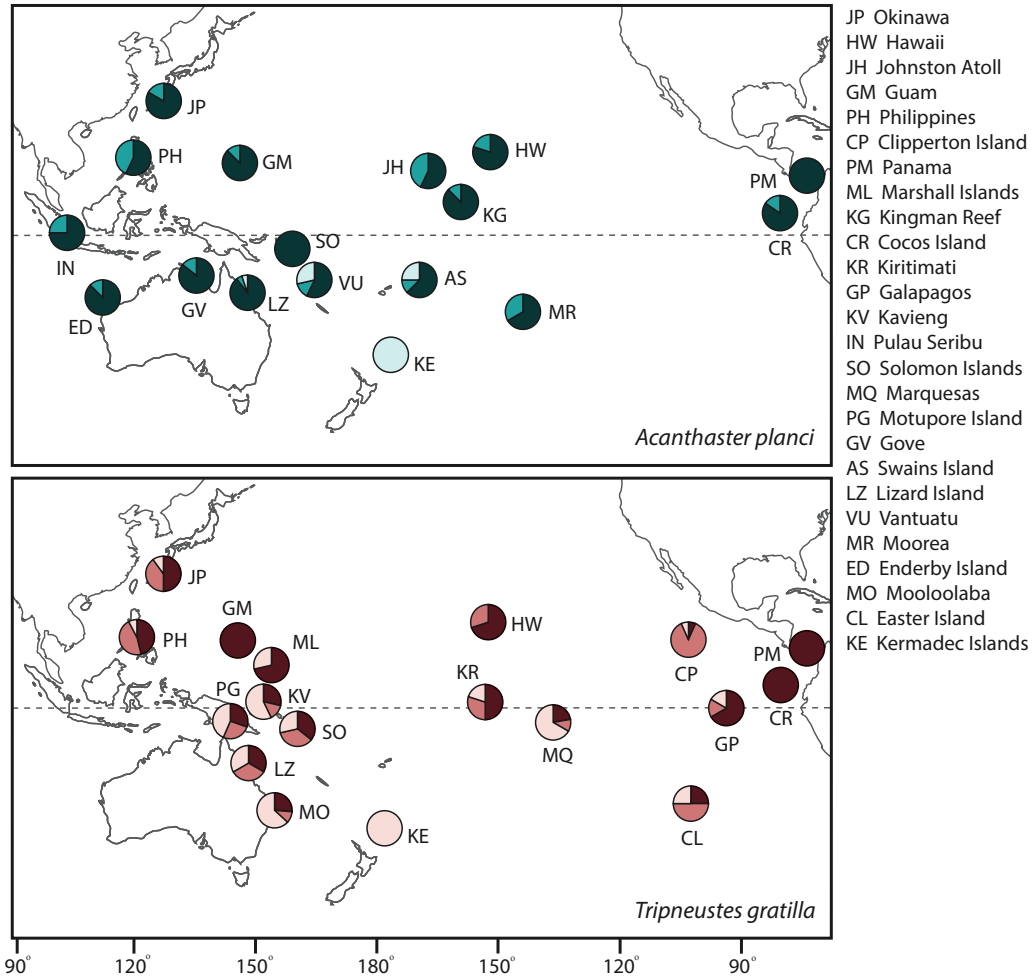


Figure 1. Map of the study locations and their haplotypic composition for *Acanthaster planci* (top) and *Tripneustes gratilla* (bottom). Pies indicate the proportion of individuals that have a haplotype that is shared among locations but not found in the Kermadec Islands (darkest tone), proportion of individuals that have a haplotype unique to their population (medium tone), and the proportion of individuals that share the haplotype found in the Kermadec Islands (lightest tone).

To understand the distribution of genetic diversity and regions of genetic disjunction we conducted several analyses of molecular variance (AMOVAs, using Tamura-Nei corrected Φ_{ST}, 10,000 permutations). Hierarchical analyses were conducted using *a priori* groupings designated according to known biogeographic disjunctions (i.e., east Pacific/central Pacific, the Eastern Pacific barrier, Ekman 1953; central Pacific/western Pacific + Australia, Drew et al. 2008). Median joining haplotype networks were constructed using Network 4.6.1.0 and Network Publisher 1.3 (fluxus-engineering.com, Bandelt et al. 1999). To compare how genetic differentiation (untransformed pairwise Tamura-Nei corrected Φ_{ST} and untransformed pairwise F_{ST}) related to geographic distance (untransformed Euclidean distance) and to address how comparisons including the Kermadec populations may deviate from the trend established across other parts of the study region, an isolation-by-distance (IBD, Wright 1943) trend was calculated for each species including all locations, and also excluding the Kermadec population. IBD trends were calculated using reduced major axis regression (Sokal and Rohlf 1981) and tested using Mantel's permutation test (10,000 matrix randomizations, Mantel 1967), executed by the Isolation by Distance Web service (Jensen et al. 2005).

We used simulations to determine whether the haplotypic compositions of the Kermadec populations were likely due to recent immigration events. We conducted 10,000 random draws (of 29 and 7 individuals for *A. planci* and *T. gratilla*, respectively) from a pool consisting of all individuals (and their haplotypes) from across the sampled range of each species, but excluding the Kermadec populations. Each random draw was a simulated immigration event to the Kermadec Islands. The random draws created a null distribution for the expected number of haplotypes in the Kermadec population (given our sample size) and the expected identity of those haplotypes via immigration. The probability of recreating the observed number of haplotypes and haplotype identity of the observed Kermadec populations via immigration alone could then be determined. The simulations were repeated for both species using two weighting systems: unweighted—each individual was weighted equally regardless of population of origin; and weighted—each individual was weighted according to their population of origin based on the predicted IBD trend for all locations excluding the Kermadec population (i.e., the predicted genetic differentiation value for the Euclidean distance between the sample location and the Kermadec Islands as a proportion of the maximum predicted genetic differentiation value for any location, subtracted from 1; all 0 values were adjusted to 0.001).

Results

The final data sets included 151 (622 bp, 17 locations) and 187 (557 bp, 18 locations) COI sequences of *A. planci* and *T. gratilla*, respectively (Table 2, Fig. 1). The *A. planci* data set comprised 43 new sequences from three locations (two previously unstudied: SO, KE) and complementary sequences from the east Indian and Pacific Oceans downloaded from GenBank (Vogler et al. 2008, Appendix 1). Sequences from GenBank (Lessios et al. 2003; Appendix 1) were also used to complement our 77 new

Table 2. Summary of included data and genetic diversity statistics for each location studied for *Acanthaster planci* and *Tripneustes gratilla*: number of sequences (n), polymorphic sites (θ), number of haplotypes (H), haplotype diversity [*Hd* (SD)], nucleotide diversity [π (SD)], Tajima's *D* statistic and significance (*P*, no correction). Source (Src) of the CO1 data: a = Vogler et al. (2008), b = present study, c = Lessios et al. (2003).

Acanthaster planci											
Code	Location	n	Latitude	Longitude	θ	H	*Hd* ± SD	π ± SD	Tajima's *D*	*P*	Source
JP	Okinawa, Japan	6	36.18	138.25	2	3	0.73 ± 0.16	0.0014 ± 0.0013	-0.05	0.45	a
HW	Hawaii, United States of America	5	19.92	-155.60	3	4	0.90 ± 0.16	0.0019 ± 0.0017	-1.05	0.15	a
JH	Johnston Atoll	7	16.73	-169.54	3	4	0.81 ± 0.13	0.0021 ± 0.0017	0.40	0.66	a
GM	Guam	8	13.44	144.79	5	5	0.86 ± 0.11	0.0030 ± 0.0022	-0.17	0.45	a
PH	Philippines	7	13.04	121.71	6	6	0.95 ± 0.10	0.0031 ± 0.0023	-1.13	0.16	a
PM	Panama	2	8.53	-80.78	0	1	0.00 ± 0.00	0.0000 ± 0.0000	na	na	a
KG	Kingman Reef	8	6.45	-162.40	12	6	0.93 ± 0.08	0.0070 ± 0.0044	-0.29	0.40	a
CR	Cocos Island, Costa Rica	13	5.52	-87.07	1	2	0.28 ± 0.14	0.0005 ± 0.0006	-0.27	0.30	a
IN	Pulau Seribu, Indonesia	8	-5.79	105.71	4	4	0.64 ± 0.18	0.0016 ± 0.0014	-1.53	0.05	a
SO	Solomon Islands	3	-8.24	157.37	2	2	0.67 ± 0.32	0.0021 ± 0.0022	0.00	0.93	b
GV	Gove, Australia	7	-12.35	136.79	2	3	0.52 ± 0.21	0.0009 ± 0.0010	-1.24	0.12	a
AS	Swains Island, American Samoa	8	-14.27	-170.70	10	6	0.93 ± 0.08	0.0067 ± 0.0042	0.37	0.67	a
LZ	Lizard Island, Australia	19	-14.67	145.46	7	4	0.63 ± 0.07	0.0025 ± 0.0018	-0.71	0.27	a,b
VU	Vanuatu	7	-15.38	166.96	9	5	0.91 ± 0.10	0.0063 ± 0.0041	0.33	0.64	a
MR	Moorea, French Polynesia	6	-17.52	-149.84	15	5	0.93 ± 0.12	0.0104 ± 0.0066	-0.10	0.48	a
ED	Enderby Island, Australia	8	-20.61	116.53	1	2	0.25 ± 0.18	0.0004 ± 0.0006	-1.05	0.21	a
KE	Kermadec Islands, New Zealand	29	-29.27	-177.92	0	1	0.00 ± 0.00	0.0000 ± 0.0000	na	na	b

Table 2. continued

Tripneustes gratilla											
Code	Location	n	Latitude	Longitude	θ	H	*Hd* ± SD	π ± SD	Tajima's *D*	*P*	Source
JP	Japan	10	36.18	138.25	6	6	0.84 ± 0.10	0.0036 ± 0.0025	-0.50	0.34	c
HW	Hawaii, United States of America	10	19.92	-155.60	8	7	0.91 ± 0.08	0.0035 ± 0.0024	-1.47	0.07	c
GM	Guam	2	13.44	144.79	0	1	0.00 ± 0.00	0.0000 ± 0.0000	na	na	c
PH	Philippines	13	13.04	121.71	13	9	0.94 ± 0.05	0.0059 ± 0.0037	-1.04	0.16	c
CP	Clipperton Island, France	15	10.28	-109.22	11	8	0.85 ± 0.07	0.0055 ± 0.0034	-0.40	0.38	c
PM	Panama	4	8.53	-80.78	0	1	0.00 ± 0.00	0.0000 ± 0.0000	na	na	c
ML	Marshall Islands	7	7.13	171.18	4	3	0.67 ± 0.16	0.0034 ± 0.0025	0.80	0.81	c
CR	Cocos Island, Costa Rica	10	5.52	-87.07	2	2	0.20 ± 0.15	0.0008 ± 0.0009	-1.40	0.08	c
KR	Kiritimati, Kiribati	10	1.87	-153.36	7	7	0.91 ± 0.08	0.0044 ± 0.0030	-0.24	0.42	c
GP	Galapagos, Ecuador	6	-0.82	-91.10	7	6	1.00 ± 0.10	0.0046 ± 0.0033	-1.01	0.20	c
KV	Kavieng, Papua New Guinea	14	-2.57	150.80	8	7	0.69 ± 0.14	0.0026 ± 0.0019	-1.70	0.03	b
SO	Solomon Islands	14	-8.24	157.37	13	10	0.92 ± 0.06	0.0042 ± 0.0027	-1.74	0.03	b
MQ	Marquesas, French Polynesia	9	-9.45	-139.39	8	3	0.56 ± 0.17	0.0039 ± 0.0027	-1.37	0.09	c
PG	Motupore Island, Papua New Guinea	23	-9.51	147.31	12	12	0.81 ± 0.08	0.0030 ± 0.0020	-1.77	0.02	b,c
LZ	Lizard Island, Australia	6	-14.67	145.46	5	5	0.93 ± 0.12	0.0034 ± 0.0027	-1.34	0.06	b
MO	Mooloolaba, Australia	19	-26.68	153.12	6	7	0.61 ± 0.13	0.0016 ± 0.0013	-1.54	0.05	b
CL	Easter Island, Chile	8	-27.12	-109.37	5	6	0.93 ± 0.08	0.0033 ± 0.0024	-0.42	0.38	c
KE	Kermadec Islands, New Zealand	7	-29.27	-177.92	0	1	0.00 ± 0.00	0.0000 ± 0.0000	na	na	b

sequences of *T. gratilla* from six locations (five previously unstudied: KV, SO, PG, MO, KE). All unique haplotypes represented in the new data sets were uploaded to GenBank (*A. planci*: KF012825-28, *T. gratilla*: KF012802-24).

Haplotype networks (Fig. 2) and AMOVA revealed that *A. planci* had greater genetic structuring across the study region than *T. gratilla* (*A. planci*: global Φ_{ST} = 0.5638, $P < 0.0001$; *T. gratilla*: global Φ_{ST} = 0.2849, $P < 0.0001$, Table 3). The hierarchical AMOVAs revealed the Eastern Pacific barrier as contributing the most to the genetic structure of both species (*A. planci*: Φ_{CT} = 0.2747, P = 0.0030; *T. gratilla*: Φ_{CT} = 0.1348, P = 0.0128; Table 3). However, once split further into the three *a priori* designated regions, considerably more genetic structure across *A. planci*'s range could be attributed to among-region differences (*A. planci*: Φ_{CT} = 0.3985, P = 0.0002; Table 3). Haplotype diversity was variable across study locations in both species (Table 2), ranging from 0 (*A. planci*: KE, PM; *T. gratilla*: KE, PM, GM) to 0.95 in *A. planci* (PH) and 1 in *T. gratilla* (GP). Most locations contained unique haplotypes, except KE (both species) and SO, PM in *A. planci*, and GM, ML, CR, PM in *T. gratilla* (see Table 2 and Figs. 1, 2). For both species, the sole haplotype found at the Kermadec Islands was shared (among 4/17 locations in *A. planci* and 14/18 locations in *T. gratilla*, Fig. 2). This haplotype was the most common and central haplotype across the entire study region for *T. gratilla*, and the third most widely shared haplotype for *A. planci*.

There was a significant positive IBD trend across the entire data set for both species based on pairwise Φ_{ST} (Fig. 3) and this trend was stronger and of higher significance once the Kermadec populations were excluded (*A. planci*: all locations, R^2 = 0.2100, P = 0.0010; excluding KE, R^2 = 0.3670, $P < 0.0001$; *T. gratilla*: all locations, R^2 = 0.0973, P = 0.0028; excluding KE, R^2 = 0.1530, P = 0.0006). The Kermadec population of *A. planci* was more differentiated from other populations than would be expected based on the IBD trend found throughout the rest of the sampled range, whereas there was no consistent pattern for *T. gratilla* (Fig. 3, Appendix 2). IBD patterns based on pairwise F_{ST} were weaker and of lower significance across the entire data set for both species (*A. planci*: R^2 = 0.0691, P = 0.0015; *T. gratilla*: R^2 = 0.0380, P = 0.0250; Fig. 3, Appendix 2), indicating that the geographic signature of mutation has not been eroded by high dispersal and is informative in these species (Bird et al. 2011).

For *A. planci*, the probability of the Kermadec Islands containing 29 individuals of the same haplotype using a random draw from the existing haplotypes was <0.0001 (individuals were weighted according to population origin, Fig. 4). For *T. gratilla*, the probability of drawing seven individuals with the same haplotype was 0.0008 (Fig. 4), and the probability that the haplotype would have the same identity as that found in the Kermadec Islands was also 0.0008. These results were robust to the method of weighting; in the unweighted comparisons for *A. planci* 29 individuals with the same haplotype were never drawn, whereas for *T. gratilla* the probability of drawing seven individuals with

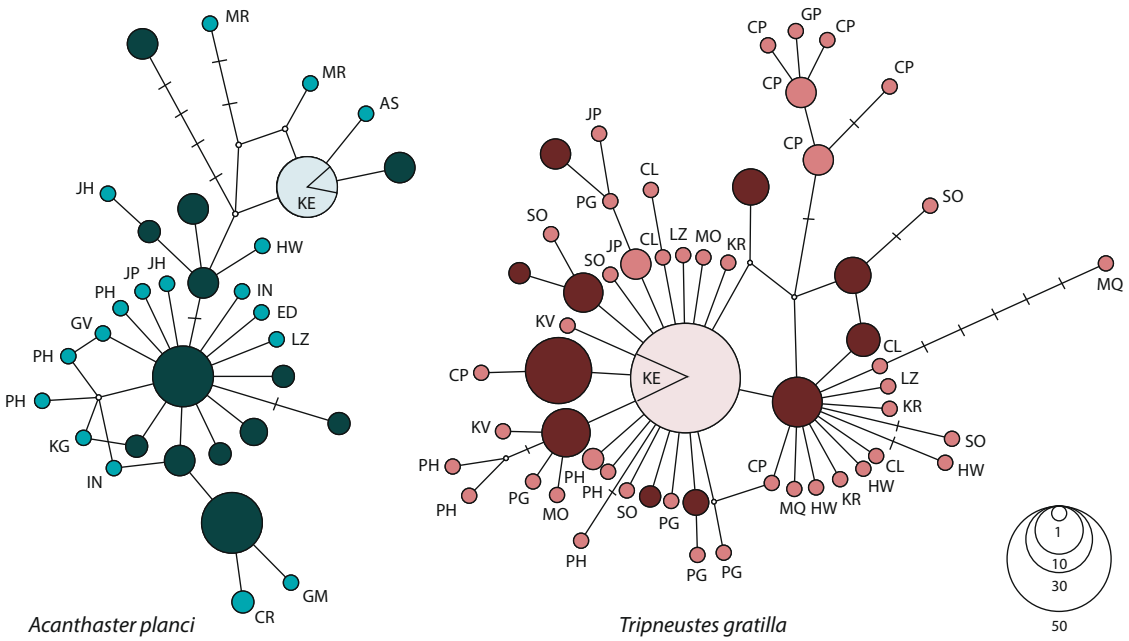


Figure 2. Median-joining haplotype networks for *Acanthaster planci* and *Tripneustes gratilla* displaying which haplotype is found in the Kermadec Islands (lightest tone), which haplotypes are shared among locations but not found in the Kermadec Islands (darkest tone), and haplotypes that are unique to one location (medium tone, also with location code indicated). Hollow circles represent hypothetical historical haplotypes or current haplotypes not sampled; edges between haplotypes or small cross-bars indicate a mutational step. The frequency of each haplotype is indicated by size (see key, bottom right).

Table 3. Analysis of molecular variances (AMOVAs, Tamura-Nei corrected Φ_{ST}) and hierarchical AMOVAs testing the effects of *a priori* designated barriers on the genetic structuring across the studied ranges of *Acanthaster planci* and *Tripneustes gratilla*. WPac = west Pacific, the Coral Triangle and Australia, including IN, ED, PH, JP, GV, GM, PG, LZ, KV, ML, MO, SO, VU; CPac = central Pacific, including KE, AS, JH, KG, KR, HW, MR, MQ; EPac = east Pacific, including CL, CP, GP, CR, PM. %var = percent variation.

Scenario	Source of variation	df	%variation	Φ	*P*
Acanthaster planci					
Sampled populations	Among populations	16	56.38	Φ_{ST} 0.5638	<0.0001
	Within populations	134	43.62		
WPac/CPac/EPac	Among regions	2	39.85	Φ_{CT} 0.3985	0.0002
	Among populations within regions	14	22.82	Φ_{SC} 0.3794	<0.0001
	Within populations	150	37.33	Φ_{ST} 0.6267	0.0000
WPac/CPac	Among regions	1	19.26	Φ_{CT} 0.1927	0.0612
	Among populations within regions	15	40.92	Φ_{SC} 0.5068	<0.0001
	Within populations	134	39.82	Φ_{ST} 0.6018	<0.0001
CPac/EPac	Among regions	1	27.47	Φ_{CT} 0.2747	0.0752
	Among populations within regions	15	38.54	Φ_{SC} 0.5314	<0.0001
	Within populations	134	33.99	Φ_{ST} 0.6601	<0.0001
Tripneustes gratilla					
Sampled populations	Among populations	17	28.49	Φ_{ST} 0.2849	<0.0001
	Within populations	169	71.51		
WPac/CPac/EPac	Among regions	2	11.04	Φ_{CT} 0.1105	0.0030
	Among populations within regions	15	20.23	Φ_{SC} 0.2274	<0.0001
	Within populations	169	68.72	Φ_{ST} 0.3128	<0.0001
WPac/CPac	Among regions	1	9.16	Φ_{CT} 0.0916	0.0065
	Among populations within regions	16	22.39	Φ_{SC} 0.2464	<0.0001
	Within populations	169	68.46	Φ_{ST} 0.3154	<0.0001
CPac/EPac	Among regions	1	13.48	Φ_{CT} 0.1348	0.0128
	Among populations within regions	16	20.98	Φ_{SC} 0.2425	<0.0001
	Within populations	169	65.54	Φ_{ST} 0.3446	<0.0001

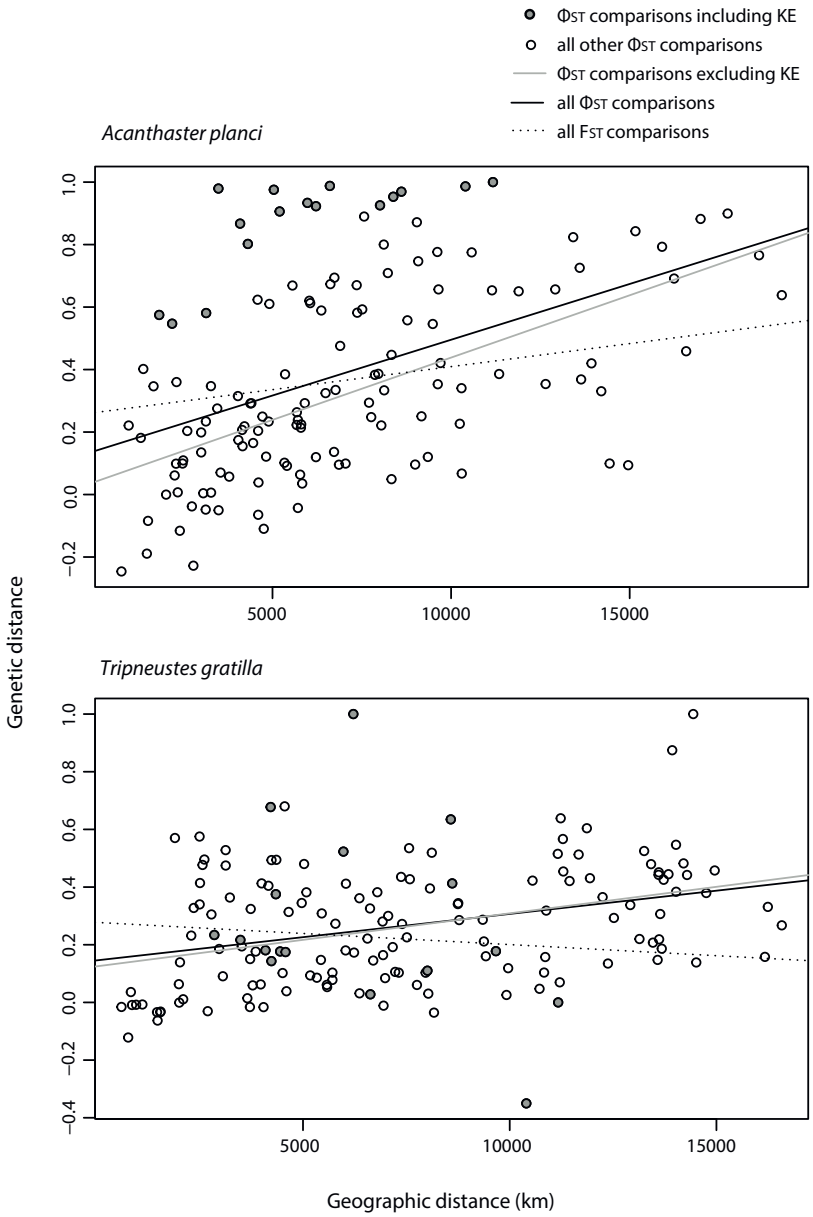


Figure 3. The relationship between genetic differentiation (untransformed pairwise Tamura-Nei corrected Φ_{ST} and untransformed pairwise F_{ST}) and Euclidean geographic distance (kilometers, km) among study locations for *Acanthaster planci* and *Tripneustes gratilla*. The gray filled circles denote pairwise Φ_{ST} relationships between the Kermadec population and other study locations. The black line represents the regression line for all Φ_{ST} comparisons (*A. planci* R^2 = 0.2100, P = 0.0010; *T. gratilla* R^2 = 0.0973, P = 0.0028), the gray line represents the regression line excluding the Kermadec population (*A. planci* R^2 = 0.3670, P < 0.0001; *T. gratilla* KE R^2 = 0.1530, P = 0.0006), and the dashed line represents the regression line for all pairwise F_{ST} comparisons (*A. planci* R^2 = 0.0691, P = 0.0015; *T. gratilla* R^2 = 0.0380, P = 0.0250).

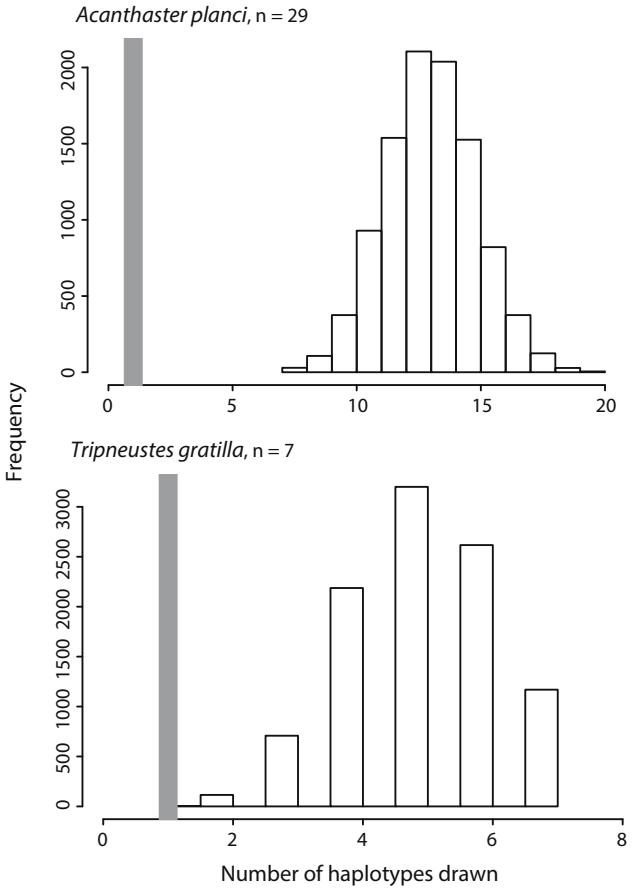


Figure 4. Frequency distribution displaying the number of haplotypes drawn from a pool consisting of all individuals (and their haplotype; weighted according to location) sampled across the study area for both species, but excluding the Kermadec populations. To simulate recruitment events to the Kermadec Islands, 29 individuals for *Acanthaster planci* and seven individuals for *Tripneustes gratilla*, were drawn 10,000 times. The gray line represents the haplotypic composition of the Kermadec populations (i.e., one haplotype).

the same haplotype was 0.0001. The haplotype drawn was always the same as that found in the Kermadec population.

Discussion

This study investigated the genetic affinities and relative genetic patterns of an edge-of-range location, the Kermadec Islands, for two common tropical echinoderms. The genetic patterns for both *A. planci* and *T. gratilla* suggest similar processes have shaped their populations around the Kermadec Islands. In both species, the Kermadec populations contained only one haplotype that was shared with other populations (Table 2, Fig. 1). Such a pattern of low diversity is consistent with a self-sustaining population founded by a limited number of colonists ("colonization" scenario, Table 1). Our resampling results (Fig. 4) indicate it is highly unlikely that the haplotypic compositions of the Kermadec populations could result from contemporary immigration events (i.e., "Metapopulation" and "Migration load" scenarios, Table 1).

The Kermadec populations in an Indo-Pacific context

Despite both echinoderm species having only one haplotype within the Kermadec population, the genetic patterns throughout the rest of their ranges varied between species. *Acanthaster planci* had considerably more genetic structuring across its range than T. *gratilla*. Recent surveys of *A. planci* based on more variable genetic markers have indicated that island populations separated by approximately 1500 km can be significantly genetically differentiated (Yasuda et al. 2009) suggesting that individuals of this species do not always meet their full dispersal potential (Timmers et al. 2012, Vogler et al. 2013). Comparatively less is known about the scale of genetic structuring and the likely influence of early life history characteristics in *T. gratilla*. However, based on a genetic survey Lessios et al. (2003) described *T. gratilla* as one large metapopulation throughout the Indo-Pacific, with non-equilibrium local variation likely due to chaotic colonization and local extinction, typical of a high dispersal and high fecundity urchin species.

The genetic affinities (i.e., the shared haplotype) of *A. planci* with American Samoa (AS), Vanuatu (VU), and Lizard Island (LZ), suggest colonization of the Kermadec Islands may have been from these locations, or from a source common to all of these regions (Fig. 1). This finding supports a phylogeographic study conducted by Vogler et al. (2013) that reveals the close genetic relationship that *A. planci* of the Kermadec Islands has with populations of the central Pacific (specifically Vanuatu and Moorea), the northern Great Barrier reef (Swains reef), and Lord Howe Island. Accordingly, descriptions of ocean currents (Schiel et al. 1986, Gardner et al. 2006) and biogeographic similarities of the Kermadec Islands marine fauna suggest connectivity with the north and west (e.g., western and central tropical Pacific for molluscs, Brook 1998; southeastern Australia for corals, Wicks et al. 2010a). In *T. gratilla*, the source of the Kermadec population is difficult to infer based on the data herein, as the haplotype found at the Kermadec Islands is also shared across much of the studied species range

(Philippines, PH 122°E to Galapagos Islands, GP 90°W, and the entire studied latitudinal range; Fig. 1). These differing genetic affinities of the two echinoderm species greatly influence the perceived differentiation of the Kermadec populations as measured by Φ_{ST}. Although the haplotype found in the Kermadec population of *A. planci* is locally shared (with As, VU, and LZ), pairwise Φ_{ST} values between the Kermadec population and other locations are consistently high because this haplotype is relatively rare. In contrast, pairwise Φ_{ST} values between the *T. gratilla* population and elsewhere are variable owing to the wide distribution and varying frequency of the haplotype shared with the Kermadec population.

The only other location represented by a single haplotype in *A. planci* was Panama (PM), which had two individuals sampled, relative to the 29 individuals from the Kermadec Islands (Table 2). All of the other locations sampled for *A. planci* contained unique haplotypes, and nearby populations of the central Pacific had several unique haplotypes (e.g., AS, MR; Table 2, Fig. 1). Similarly, for *T. gratilla* there were only two locations other than the Kermadec Islands that had one haplotype, both of which had small sample sizes (GM = 2, PM = 4; table 2). Other peripheral and isolated locations (CL, CP, and JP) sampled for *T. gratilla* had high levels of genetic novelty (4, 13, and 4 unique haplotypes respectively; Table 2 and Fig. 1) some of which are very divergent (see position of unique haplotypes in the network, Fig. 2). The lack of unique haplotypes in the Kermadec Islands could indicate that these populations of *A. planci* and *T. gratilla* are relatively young. The peripheral and isolated location of the Kermadec Islands, their short geological history (1.8–3 Ma; Watt 1975), and continuing active volcanism support this notion of recent colonization, however alternative explanations are possible (discussed below).

Evidence for self-recruitment and little immigration in an edge-of-range population

We find no genetic diversity in the Kermadec populations (Table 2), which suggests that bottlenecks have affected both *A. planci* and *T. gratilla*. Such a pattern is reminiscent of contemporary invasion scenarios (Roman and Darling 2007). Colonization of introduced species often involves a small number of initial colonists, causing the colonizing population to be much less genetically diverse than the source population. Our findings contrast with genetic patterns observed following the natural colonization of Krakatau after its 1883 eruption (Barber et al. 2002b). Despite the short timeframe, stomatopod populations had high levels of genetic diversity, possibly owing to the central location of Krakatau in the species ranges and consequent high levels of immigration (Barber et al. 2002b). In contrast, for the peripheral Kermadec populations, our procedure of taking random draws from the species' sampled ranges confirmed it is highly unlikely that the haplotypic compositions (multiple individuals sharing the same haplotype) could result from contemporary immigration events (Fig. 4). Therefore we can reject a frequent immigration scenario (Table 1).

It is conceivable that larvae of these high dispersal echinoderm species reach the Kermadec Islands, but that their successful immigration is inhibited by high-density blocking (Hewitt 1993), the

latent effects of long-distance dispersal, or selection. A pattern of "founder takes all" via high-density blocking of the existing conspecific population has been proposed for several marine colonization scenarios where the standing genetic diversity is maintained despite the arrival of potential immigrants (reviewed in Waters et al. 2013). "Legacy effects" have also been suggested to bias recruitment in larval fish whereby the survival is greater in fish that have undergone a less stressful pelagic larval phase (Shima and Swearer 2010). It is possible that locally derived larvae of the Kermadec Islands have greater fitness due to their shorter and potentially less stressful pelagic larval phase (Nosil et al. 2005). Both of these mechanisms could be couched as forms of local selection, as the locally derived larvae survive in greater proportions than larvae from elsewhere (either by "aggregate" fitness as a consequence of being related to the surviving genotypes, or individual fitness and competitive ability as a consequence of early life experience in the pelagic environment). Although posited as a neutral locus, it is also possible that the mitochondrion, or some genetic element that is linked to the COI haplotype found at the Kermadec Islands is under positive selection. As the mitochondrion is known to have function in metabolism, the haplotype found in the Kermadec Islands may offer some functional benefit in the colder water temperatures (similar examples are reviewed in Galtier et al. 2009). Unfortunately, the influence of such phenomena on the genetic composition of the Kermadec populations is undetectable using our current study design. Acquiring such knowledge would require the capture (and genotyping using multiple unlinked loci) of potential immigrants prior to any form of selection, and/or some proof of a haplotype by environment interaction.

Assuming that the mitochondrion is behaving as a neutral locus, the absence of variation in the Kermadec populations would imply that there has been insufficient time and/or opportunity since colonization for new genetic variants to arise and differentiation to occur (Lesica and Allendorf 1995; Table 1). Although previous studies indicate that *A. planci* and *T. gratilla* have remained abundant around Meyer Island of the Kermadec Islands (*A. planci*: up to 0.008 individuals per m^2 in 1995, Brook 1999; and 0.25 individuals per m^2, Gardner et al. 2006; *T. gratilla*: up to 0.7 individuals per m^2, Cole et al. 1992; and 0.75 individuals per m^2, Gardner et al. 2006), this area is small and may not support a large effective population size. If the Kermadec populations are relatively small and/or experience fluctuating abundance (common in echinoderms, Uthicke et al. 2009) the resultant effective population sizes would maintain low standing diversity in the populations.

Regardless of the timing of colonization or post-colonization mechanisms that keep genetic diversity low, our results suggest both *A. planci* and *T. gratilla* populations are self-recruiting. Thus, while we can only speculate that the peripheral nature of the Kermadec Islands has precluded recurrent immigration (and not selection processes at settlement), the marginality of the islands has certainly not impeded reproduction and self-recruitment following initial colonization. A scenario of "colonization" (Table 1) followed by self-recruitment and little immigration for *A. planci* is further supported by the findings of Vogler et al. (2013). Using the more variable and faster mutating control region of the mitochondrial locus, the authors found six unique haplotypes from a sample of seven *A.*

planci individuals from the Kermadec Islands. None of these haplotypes were shared outside of the Kermadec Islands, and are likely to be locally derived (see Appendix 3) indicating *A. planci* has had an isolated demographic history in the Kermadec Islands.

Implications for range stability of tropical echinoderms at the Kermadec Islands

Understanding the reproductive capacity and reliance on immigration with regions outside of the Kermadec Islands Marine reserve is important for predicting the persistence of these species at this locality. Moreover, attributes of peripheral populations (reproduction and connectivity) interact to determine their capacity for local adaptation and species range expansion into other locales (Sexton et al. 2009). For example, if reproduction is largely unsuccessful within the Kermadec Islands, and the populations are reliant on immigrants from elsewhere in their range, the echinoderms are unlikely to adapt to the marginal conditions of the Kermadec Islands ("Migration load" scenario, e.g., Dawson et al. 2010) and are unlikely to extend their range (García-Ramos and Kirkpatrick 1997, Kirkpatrick and Barton 1997). At the other extreme, if the Kermadec populations are reproductive and have no immigration, their range may be only temporarily limited ("Genetic impoverishment" scenario, Holt 2003). This situation can quickly change with a subsequent immigration event (Gomulkiewicz et al. 1999), and/or the generation of any genetic novelty in the population (Bataillon 2003). While patterns of neutral genetic diversity are not directly related to adaptive genetic diversity (but see Pujol and Pannell 2008) our findings suggest the Kermadec populations are subject to a scenario of "Genetic impoverishment" where local adaptation and subsequent range shifts may only be a matter of time.

Tropical vagrants (and some reproductive populations) are known to occur along the northeast coast of mainland New Zealand (molluscs, reviewed by Powell 1976; marine reptiles, reviewed by Gill 1997; fishes, reviewed by Francis et al. 1999). Most insurgent tropicals are presumed to have originated from Norfolk Island and to a lesser degree, Lord Howe Island (Francis et al. 1999). Although simulations of larval dispersal from Raoul Island to mainland New Zealand indicate that transit times would be in excess of fifty days, dispersal from the southern most island of the Kermadec archipelago (L'Esperance rock) may be as little as twenty days (Sutton et al. 2012), well within the pelagic larval duration of many benthic reef species. Moreover, fish previously considered endemic to the Kermadec Islands have been found in mainland New Zealand, providing evidence for larval transport from the north (Francis et al. 1999). With changes to ocean currents and global temperatures now considered imminent (Doney et al. 2012) the Kermadec Islands may become an important stepping-stone for tropical marine species into New Zealand.

In conclusion, the genetic patterns found in the Kermadec populations for *A. planci* and *T. gratilla* upheld the expectations of the CPH. Both populations had low genetic diversity, and the *A. planci* population was consistently more genetically differentiated from other populations throughout the sampled range (Fig. 3, Appendix 2). Furthermore, we have demonstrated that these populations appear to be self-sustaining. Such conditions would foster local adaptation at the range edge (García-

Ramos and Kirkpatrick 1997). With the onset of climate change, the Kermadec Islands may represent a "leading-edge" location for many tropical marine populations, and provide an important dispersal corridor for marine organisms into New Zealand. As such, these populations should be monitored and conserved appropriately.

Acknowledgments

Expeditions to the Kermadec Islands were made possible by the Sir Peter Blake Trust, the Commanding Officer and Ship's Company of HMNZS Canterbury, the RV Braveheart Crew, and the Auckland Museum Tamaki Paenga Hira (Department of Conservation Authority to undertake scientific study within a marine reserve to T Trnski: DOCDM-737382). Sampling in the Solomon Islands was via the Australian Government's Pacific Strategy Assistance Program and with the assistance of the Roviana Conservation Foundation (Solomon Islands Government Ministry of Education and Human Resource Development and Ministry of Fisheries and Marine resources research permit to S Albert). Sampling in Papua New Guinea was in coordination with the National Research Institute, the Department of Foreign Affairs and Immigration (Research Visa: 10350008304) and the Department of Environment and Conservation (Permit to Export Wildlife: 011318). We are grateful to the staff of the Australian Museum Lizard Island research station for their facilities and support (Great Barrier Reef Marine Park Authority and Queensland Parks and Wildlife Marine Parks Permit: G08/28114.1, G09/31678.1, G10/33597.1, G11/34640.1; Queensland Government Department of Primary Industries General Fisheries Permit: 118636, 150981; Australian Quarantine Inspection Service Permit to Import Quarantine Material: IP10017966). We especially thank JD Aguirre, J Aini (and Ailan Awareness), S Albert, A Berry, H Bostock, K Davis, C Duffy, M Jimuru, J Keyse, J Kinch (National Fisheries College, Papua New Guinea), A Mirams, A Smith (Tiki2 Adventure tours), EA Treml, T Trnski, SR Ullrich, Stephen, Lavud, Takenda and the Young Blake Expedition crew for logistical support and field assistance. We are also grateful to JD Aguirre, D Blower, ED Crandall, C Duffy, J Giles, A Mather, L Pope, T Trnski, and three anonymous reviewers for providing helpful comments on the manuscript, and G Wörheide for providing us advance access to sequence data used in Vogler et al. (2013). Funding for this work was provided by the Australian Research Council (DP0878306 to CR) and an Explorer's Club Exploration Fund (to LL). LL was supported by an Australian Postgraduate Award from the Australian Government and a Queensland Government Smart Futures PhD scholarship. Many of the ideas discussed here grew out of work funded by the Sea World Research and Rescue Foundation (SWR/1/2012, to CR and LL), a Paddy Pallin Foundation and The Foundation for National Parks and Wildlife Science Grant, an Ecological Society of Australia Student Research Grant, the Lerner Gray Memorial Fund of the American Museum of Natural History, and a Great Barrier Reef Marine Park Authority's Science for Management Award (to LL).

Appendix 1. Data source and GenBank accession numbers for *Acanthaster planci* and *Tripneustes gratilla* sequences included in this study. Source of the CO1 data: a. Vogler et al. 2008, b. this study, c. Lessios et al. 2003. Numbers in brackets following an accession number denote frequency of that haplotype for that location if greater than 1 (b. only).

Acanthaster planci			
Location	n	Source	GenBank accession numbers
Okinawa, Japan	6	a	FM174508-13
Hawaii, United States of America	5	a	FM202070-74
Johnston Atoll	7	a	FM174472-78
Guam	8	a	FM174500-07
Philippines	7	a	FM177197-203
Panama	2	a	FM202089-90
Kingman Reef	8	a	FM174479-86
Cocos Island, Costa Rica	13	a	FM202076-88
Pulau Seribu, Indonesia	8	a	FM174537-44
Solomon Islands	3	b	KF012825(2), 012826
Gove, Australia	7	a	FM174530-36
Swains Island, American Samoa	8	a	FM174487-94
Lizard Island, Australia	19	a,b	FM174514-21, KF012825(5), 012826(4), 012827, 012828
Vanuatu	7	a	FM177190-96
Moorea, French Polynesia	6	a	FM174495-99, 202075
Enderby Island, Australia	8	a	FM174522-29
Kermadec Islands, New Zealand	29	b	KF012825(29)

Appendix 1. continued

Tripneustes gratilla			
Location	n	Source	GenBank accession numbers
Japan	10	c	AY205384, 205416, 205422-25, 205429-30, 205439, 205452
Hawaii, United States of America	10	c	AY205373, 205376-78, 205405-06, 205419, 205443, 205450-51
Guam	2	c	AY205407-08
Philippines	13	c	AY205394, 205409-14, 205417-18, 205435-36, 205438, 205440
Clipperton Island, France	15	c	AY205525-32, 205546, 205552, 205554, 205556-59
Panama	4	c	AY205547-50
Marshall Islands	7	c	AY205458-59, 205461-64, 205469
Cocos Island, Costa Rica	10	c	AY205533-41, 205551
Kiritimati, Kiribati	10	c	AY205379, 205385, 205420, 205426-28, 205449, 205453-55
Galapagos, Ecuador	6	c	AY205542-45, 205553, 205555
Kavieng, Papua New Guinea	14	b	KF012802(8), 012803, 012804, 012805, 012815, 012816, 012817
Solomon Islands	14	b	KF012802(4), 012803(2), 012804, 012806, 012812, 012820, 012821, 012822, 012823, 012824
Marquesas, French Polynesia	9	c	AY205380, 205386-91, 205444-45
Motupore Island, Papua New Guinea	23	b,c	AY205395-97, 205415, 205441-42, KF012802(7), 012803, 012805, 012806, 012807, 012808, 012811, 012812, 012813, 012814
Lizard Island, Australia	6	b	KF012802(2), KF012803, KF012804, 012809, 012810
Mooloolaba, Australia	19	b	KF012802(12), KF012803(2), KF012804, 012813, 012815(14), 012818, 012819
Easter Island, Chile	8	c	AY205374-75, 205392-93, 205431-34
Kermadec Islands, New Zealand	7	b	KF012802(7)

Appendix 2. Pairwise genetic differentiation (Φ_{ST}, F_{ST}) among study locations for both species. Location codes are listed in Table 1. a. Pairwise Φ_{ST} (lower triangle, Tamura-Nei corrected) and corresponding p−values (upper triangle). b. Pairwise F_{ST} (lower triangle) and corresponding p−values (upper triangle). Significance $P < 0.05$ following Benjamini and Hochberg correction is denoted by italicized text.

Acanthaster planci

a. Φ_{ST}

	JP	HW	JH	GM	PH	PM	KG	CR	IN	SO	GV	AS	LZ	VU	MR	ED	KE
JP	-	*0.0071*	*0.0016*	0.0591	0.5093	0.0618	0.2556	*0.0005*	0.1582	0.3142	0.1582	0.0620	*0.0456*	0.0729	*0.0224*	0.1305	*0.0000*
HW	0.6734	-	0.7781	*0.0028*	*0.0048*	0.0736	*0.0028*	*0.0005*	*0.0025*	*0.0380*	*0.0034*	0.0558	*0.0000*	*0.0456*	*0.0130*	*0.0026*	*0.0000*
JH	0.6690	-0.0844	-	*0.0013*	*0.0009*	*0.0456*	*0.0005*	*0.0000*	*0.0024*	*0.0219*	*0.0025*	*0.0113*	*0.0000*	*0.0041*	*0.0025*	*0.0000*	*0.0000*
GM	0.2035	0.5891	0.6101	-	0.1077	0.4557	0.4037	*0.0026*	0.1137	0.9999	0.0970	0.0802	0.7416	*0.0362*	*0.0030*	0.1008	*0.0000*
PH	0.0041	0.5576	0.5820	0.1093	-	0.0535	0.4138	*0.0000*	0.6941	0.7753	0.6374	0.0729	0.0520	0.0850	*0.0075*	*0.0456*	*0.0000*
PM	0.7257	0.8001	0.7764	0.0993	0.4586	-	0.2551	0.9999	*0.0446*	0.4560	0.0535	0.3285	0.5850	0.1582	0.1018	*0.0449*	*0.0057*
KG	0.0955	0.3465	0.4019	0.0352	0.0492	0.0960	-	*0.0005*	0.4279	0.8799	0.1715	0.1976	0.0802	0.0882	0.1573	0.1330	*0.0000*
CR	0.8237	0.8898	0.8715	0.4199	0.6913	-0.2462	0.4471	-	*0.0000*	*0.0485*	*0.0000*	*0.0026*	*0.0028*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
IN	0.0636	0.6536	0.6566	0.1213	-0.0380	0.6382	0.0672	0.7658	-	0.8270	0.8868	0.0710	0.0591	*0.0406*	*0.0039*	0.8259	*0.0000*
SO	0.1018	0.6202	0.6235	-0.2278	-0.0648	0.3684	-0.1097	0.6565	-0.0430	-	0.3541	0.4304	0.9999	0.2551	0.1203	0.3329	*0.0013*
GV	0.0918	0.7089	0.6941	0.1987	0.0061	0.7930	0.0991	0.8428	-0.0503	0.0987	-	0.0729	0.0729	0.0758	*0.0034*	0.5028	*0.0000*
AS	0.2478	0.2068	0.2754	0.2138	0.2212	0.2265	0.0991	0.5459	0.2503	0.0703	0.2634	-	*0.0209*	0.9645	0.2150	0.0650	*0.0000*
LZ	0.2386	0.5926	0.6126	-0.0482	0.1744	0.0938	0.1196	0.3306	0.1649	-0.1892	0.2209	0.2495	-	*0.0053*	*0.0000*	*0.0449*	*0.0000*
VU	0.3246	0.2227	0.2923	0.3151	0.2924	0.3537	0.1546	0.6505	0.3347	0.1813	0.3470	-0.1160	0.3600	-	0.2761	*0.0208*	*0.0005*
MR	0.3531	0.2184	0.2915	0.3863	0.3404	0.3336	0.1348	0.6701	0.3858	0.2246	0.3831	0.0611	0.4758	0.0386	-	*0.0026*	*0.0000*
ED	0.1364	0.7751	0.7466	0.2336	0.0572	0.8995	0.1203	0.8821	-0.0003	0.2040	0.0071	0.2939	0.2339	0.3849	0.4209	-	*0.0000*
KE	0.9694	0.9335	0.9060	0.9228	0.9256	1.0000	0.8021	0.9863	0.9532	0.9794	0.9756	0.5748	0.8672	0.5469	0.5809	0.9878	-

Appendix 2. continued

Acanthaster planci

b. F_{ST}

	JP	HW	JH	GM	PH	PM	KG	CR	IN	SO	GV	AS	LZ	VU	MR	ED	KE
JP	-	0.0440	*0.0075*	*0.0043*	*0.0062*	0.0728	*0.0052*	*0.0001*	*0.0039*	0.0616	*0.0046*	*0.0065*	*0.0007*	*0.0048*	0.0450	*0.0009*	*0.0000*
HW	0.1876	-	0.0455	0.0716	0.1489	0.1025	0.0993	*0.0005*	*0.0158*	0.1445	*0.0149*	0.0992	*0.0056*	0.0952	0.1675	*0.0026*	*0.0000*
JH	0.2270	0.1490	-	*0.0087*	0.0463	0.0581	*0.0069*	*0.0000*	*0.0025*	0.0574	*0.0034*	*0.0073*	*0.0007*	*0.0192*	0.0434	*0.0005*	*0.0000*
GM	0.2005	0.1237	0.1660	-	0.0472	0.0440	*0.0246*	*0.0000*	*0.0042*	0.0530	*0.0015*	*0.0250*	*0.0011*	*0.0198*	0.0451	*0.0006*	*0.0000*
PH	0.1532	0.0719	0.1191	0.0964	-	0.0533	0.1013	*0.0004*	*0.0142*	0.0988	*0.0141*	0.1005	*0.0035*	0.1017	0.1672	*0.0008*	*0.0000*
PM	0.4454	0.3502	0.3826	0.3425	0.2875	-	0.0623	*0.0177*	*0.0201*	0.1975	*0.0275*	0.0669	*0.0052*	0.0851	0.0771	*0.0209*	*0.0012*
KG	0.1627	0.0843	0.1294	0.1071	0.0598	0.2949	-	*0.0001*	*0.0077*	0.0946	*0.0026*	0.0481	*0.0005*	0.0369	0.1035	*0.0010*	*0.0000*
CR	0.5533	0.5073	0.5079	0.4739	0.4455	0.7532	0.4414	-	*0.0000*	*0.0038*	*0.0001*	*0.0001*	*0.0000*	*0.0000*	*0.0002*	*0.0000*	*0.0000*
IN	0.3157	0.2451	0.2767	0.2500	0.2072	0.4912	0.2143	0.5685	-	*0.0255*	*0.0022*	*0.0058*	*0.0002*	*0.0101*	*0.0140*	*0.0004*	*0.0000*
SO	0.2913	0.1945	0.2417	0.2096	0.1536	0.5714	0.1651	0.6319	0.3493	-	*0.0255*	0.0892	*0.0023*	0.0982	0.1245	*0.0120*	*0.0001*
GV	0.3767	0.3079	0.3333	0.3034	0.2619	0.5844	0.2668	0.6254	0.4141	0.4288	-	*0.0027*	*0.0001*	*0.0053*	*0.0135*	*0.0009*	*0.0000*
AS	0.1627	0.0843	0.1294	0.1071	0.0598	0.2949	0.0714	0.4414	0.2143	0.1651	0.2668	-	*0.0009*	*0.0346*	0.0962	*0.0008*	*0.0000*
LZ	0.3349	0.2766	0.3023	0.2791	0.2424	0.4766	0.2479	0.5260	0.3675	0.3636	0.4119	0.2479	-	*0.0027*	*0.0024*	*0.0000*	*0.0000*
VU	0.1777	0.0974	0.1429	0.1197	0.0714	0.3188	0.0831	0.4666	0.2305	0.1826	0.2857	0.0831	0.2628	-	0.1053	*0.0006*	*0.0000*
MR	0.1667	0.0826	0.1307	0.1070	0.0569	0.3115	0.0692	0.4719	0.2224	0.1692	0.2805	0.0692	0.2564	0.0814	-	*0.0021*	*0.0000*
ED	0.5350	0.4815	0.4834	0.4464	0.4140	0.7899	0.4107	0.7307	0.5536	0.6326	0.6206	0.4107	0.5129	0.4372	0.4421	-	*0.0000*
KE	0.8601	0.8478	0.8176	0.7810	0.7744	1.0000	0.7571	0.9109	0.8464	0.9454	0.8925	0.7571	0.7389	0.7893	0.8085	0.9466	-

Appendix 2. continued

Tripneustes gratilla

a. Φ_{ST}

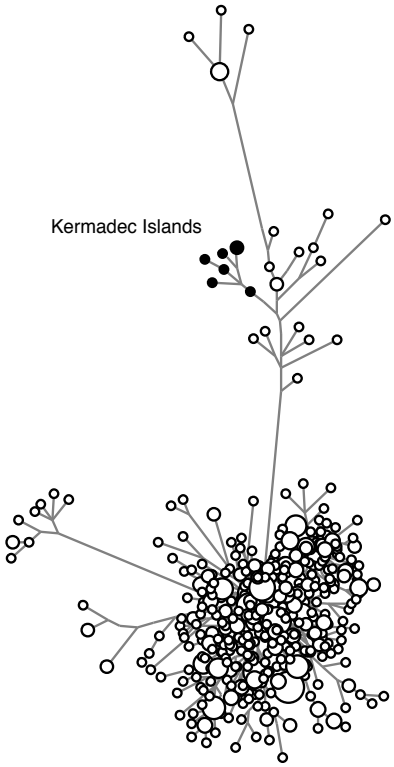
	JP	HW	GM	PH	CP	PM	ML	CR	KR	GP	KV	SO	MQ	PG	LZ	MO	CL	KE
JP	-	*0.0000*	*0.0310*	*0.0281*	*0.0002*	*0.0009*	*0.0082*	*0.0001*	0.0569	*0.0080*	*0.0236*	*0.0134*	*0.0171*	*0.0084*	0.0813	*0.0005*	*0.0067*	*0.0000*
HW	0.3259	-	0.5886	*0.0000*	*0.0002*	*0.0051*	*0.0341*	*0.0012*	*0.0258*	*0.0080*	*0.0000*	*0.0029*	*0.0043*	*0.0000*	*0.0059*	*0.0000*	0.0401	*0.0008*
GM	0.4955	0.0315	-	*0.0266*	*0.0284*	0.0673	0.1064	*0.0141*	0.0609	0.1026	*0.0094*	0.0442	0.2036	*0.0112*	*0.0355*	*0.0044*	0.0475	*0.0303*
PH	0.0907	0.2855	0.3400	-	*0.0000*	*0.0022*	*0.0004*	*0.0000*	*0.0000*	*0.0040*	0.2904	0.0696	0.0390	*0.0249*	0.4985	*0.0352*	*0.0039*	0.0675
CP	0.4541	0.3819	0.4210	0.4446	-	*0.0045*	*0.0021*	*0.0001*	*0.0008*	*0.0175*	*0.0000*	*0.0000*	*0.0005*	*0.0000*	*0.0026*	*0.0000*	*0.0008*	*0.0000*
PM	0.4424	0.5191	1.0000	0.2673	0.5282	-	*0.0059*	0.9999	*0.0033*	0.7263	*0.0000*	*0.0017*	*0.0094*	*0.0000*	*0.0046*	*0.0002*	*0.0036*	0.9999
ML	0.3137	0.1765	0.1853	0.3086	0.3410	0.6040	-	*0.0001*	0.2061	*0.0213*	*0.0004*	*0.0097*	*0.0073*	*0.0005*	*0.0183*	*0.0002*	*0.0184*	*0.0004*
CR	0.4794	0.5349	0.8745	0.3312	0.5752	-0.1215	0.6385	-	*0.0000*	0.2163	*0.0000*	*0.0000*	*0.0001*	*0.0000*	*0.0008*	*0.0000*	*0.0005*	0.9999
KR	0.1031	0.1390	0.2805	0.1602	0.3448	0.3952	0.0624	0.4351	-	*0.0280*	*0.0000*	0.0797	0.1335	*0.0005*	0.1011	*0.0001*	0.1235	*0.0010*
GP	0.2188	0.2721	0.4253	0.1580	0.3277	-0.0341	0.3185	0.0358	0.1641	-	*0.0004*	*0.0142*	*0.0194*	*0.0001*	*0.0470*	*0.0001*	*0.0128*	*0.0040*
KV	0.1019	0.3613	0.5702	0.0145	0.5152	0.4420	0.4140	0.4511	0.1728	0.2199	-	0.5775	*0.0349*	0.6211	0.7472	0.9038	*0.0005*	*0.0064*
SO	0.0856	0.1803	0.3051	0.0389	0.4223	0.3066	0.2317	0.3375	0.0532	0.1349	-0.0081	-	0.5566	0.5930	0.9500	0.2660	0.1808	*0.0055*
MQ	0.1188	0.1504	0.3442	0.0697	0.4128	0.3821	0.2729	0.4116	0.0632	0.1474	0.0606	-0.0112	-	0.1569	0.5783	*0.0149*	0.5362	*0.0130*
PG	0.0936	0.3000	0.4767	0.0591	0.5129	0.3794	0.3637	0.3837	0.1453	0.2073	-0.0090	-0.0072	0.0305	-	0.5565	0.4636	*0.0051*	*0.0123*
LZ	0.1033	0.2257	0.4747	-0.0164	0.4308	0.4576	0.3241	0.4821	0.0843	0.1475	-0.0339	-0.0629	-0.0354	-0.0157	-	0.6730	0.1562	*0.0052*
MO	0.1920	0.4272	0.6800	0.0608	0.5667	0.5467	0.4936	0.5250	0.2214	0.2930	-0.0304	0.0109	0.1032	-0.0006	-0.0319	-	*0.0001*	*0.0071*
CL	0.1863	0.1059	0.3652	0.1386	0.4045	0.4800	0.2867	0.4944	0.0781	0.1943	0.1574	0.0258	-0.0160	0.1036	0.0472	0.2115	-	0.2981
KE	0.4128	0.5230	1.0000	0.1100	0.6347	0.0000	0.6775	-0.3499	0.3752	0.1778	0.1767	0.2167	0.1747	0.1432	0.1808	0.2338	0.0281	-

Appendix 2. continued

Tripneustes gratilla

b. F_{ST}

	JP	HW	GM	PH	CP	PM	ML	CR	KR	GP	KV	SO	MQ	PG	LZ	MO	CL	KE
JP	-	*0.0050*	*0.0304*	*0.0011*	*0.0004*	*0.0020*	*0.0015*	*0.0000*	*0.0058*	*0.0040*	*0.0002*	*0.0013*	*0.0000*	*0.0000*	*0.0276*	*0.0000*	*0.0177*	*0.0000*
HW	0.1222	-	*0.0272*	*0.0033*	*0.0026*	*0.0011*	*0.0007*	*0.0001*	*0.0126*	*0.0030*	*0.0005*	*0.0036*	*0.0002*	*0.0019*	0.0645	*0.0002*	*0.0198*	*0.0001*
GM	0.3403	0.2955	-	*0.0274*	*0.0066*	0.0660	0.0570	*0.0157*	*0.0327*	*0.0341*	*0.0087*	*0.0156*	*0.0383*	*0.0089*	0.0746	*0.0115*	0.0689	*0.0273*
PH	0.1081	0.0761	0.2696	-	*0.0004*	*0.0000*	*0.0003*	*0.0000*	*0.0035*	*0.0013*	*0.0005*	*0.0018*	*0.0000*	*0.0004*	0.0520	*0.0000*	*0.0169*	*0.0000*
CP	0.1539	0.1225	0.3247	0.1090	-	*0.0006*	*0.0009*	*0.0000*	*0.0026*	*0.0013*	*0.0000*	*0.0004*	*0.0000*	*0.0001*	*0.0259*	*0.0000*	*0.0062*	*0.0001*
PM	0.4457	0.4077	1.0000	0.3731	0.4144	-	*0.0060*	*0.0009*	*0.0010*	*0.0094*	*0.0002*	*0.0009*	*0.0016*	*0.0000*	*0.0049*	*0.0000*	*0.0020*	*0.0036*
ML	0.2371	0.2017	0.4815	0.1829	0.2290	0.5852	-	*0.0000*	*0.0011*	*0.0040*	*0.0001*	*0.0002*	*0.0005*	*0.0000*	*0.0180*	*0.0000*	*0.0023*	*0.0008*
CR	0.4778	0.4444	0.8263	0.4055	0.4390	0.8549	0.5989	-	*0.0000*	*0.0006*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0003*	*0.0000*	*0.0003*	*0.0000*
KR	0.1222	0.0889	0.2955	0.0761	0.1225	0.4077	0.2017	0.4444	-	*0.0029*	*0.0003*	*0.0040*	*0.0000*	*0.0013*	0.0687	*0.0003*	*0.0188*	*0.0000*
GP	0.2622	0.2258	0.5385	0.2052	0.2517	0.6418	0.3650	0.6418	0.2258	-	*0.0015*	*0.0024*	*0.0003*	*0.0004*	*0.0321*	*0.0004*	*0.0049*	*0.0006*
KV	0.2366	0.2050	0.4350	0.1874	0.2291	0.5112	0.3184	0.5268	0.2050	0.3438	-	*0.0003*	*0.0000*	*0.0000*	*0.0060*	*0.0000*	*0.0010*	*0.0000*
SO	0.1144	0.0827	0.2759	0.0706	0.1150	0.3755	0.1884	0.4054	0.0827	0.2105	0.1923	-	*0.0000*	*0.0001*	*0.0352*	*0.0000*	*0.0146*	*0.0000*
MQ	0.2963	0.2624	0.5494	0.2400	0.2823	0.6273	0.3932	0.6296	0.2624	0.4252	0.3683	0.2442	-	*0.0000*	*0.0049*	*0.0000*	*0.0032*	*0.0001*
PG	0.1725	0.1427	0.3387	0.1291	0.1700	0.4136	0.2435	0.4276	0.1427	0.2650	0.2416	0.1346	0.2911	-	*0.0113*	*0.0000*	*0.0036*	*0.0000*
LZ	0.1154	0.0788	0.3115	0.0652	0.1159	0.4563	0.2052	0.4987	0.0788	0.2333	0.2079	0.0724	0.2758	0.1380	-	*0.0020*	0.0986	*0.0009*
MO	0.2892	0.2589	0.4896	0.2389	0.2775	0.5484	0.3697	0.5538	0.2589	0.3947	0.3527	0.2429	0.4130	0.2852	0.2680	-	*0.0003*	*0.0000*
CL	0.1151	0.0805	0.2949	0.0675	0.1157	0.4206	0.1984	0.4620	0.0805	0.2238	0.2020	0.0744	0.2633	0.1368	0.0692	0.2588	-	*0.0003*
KE	0.5243	0.4892	1.0000	0.4446	0.4775	1.0000	0.6667	0.8818	0.4892	0.7219	0.5700	0.4437	0.6919	0.4627	0.5625	0.5940	0.5130	-



Appendix 3. Median-joining haplotype network for *Acanthaster planci* based on mitochondrial control region sequences published in Vogler et al. (2013). Haplotypes found in the Kermadec Islands (black) are all unique to the Kermadec Islands, and are likely to be locally derived (based on the topology of the network). Note: relative frequency of haplotypes is represented by circle size; mutational steps between haplotypes are not represented.

CHAPTER SEVEN. General discussion

This thesis contributes empirical case studies as well as methods of study design and analysis for the field of seascape genetics. Overall, this body of work advances our understanding of population connectivity of Indo-Pacific coral reef organisms. In this chapter I draw together the major findings of the empirical data chapters contributing to this thesis. Last, avenues for future research are proposed.

Major findings and contributions

Chapters Four, Five and Six test spatial genetic hypotheses in marine systems across various seascapes and species using purpose-designed and innovative methods. In **Chapter Four** we co-sampled four coral reef fishes (*Acanthurus triostegus*, *Halichoeres hortulanus*, *Pomacentrus coelestis*, and *Dascyllus trimaculatus*) across a common seascape allowing comparative study of spatial genetic patterns. The overall aim was to identify the historical and contemporary seascape features and processes responsible for any congruent spatial genetic patterns across species. Specifically, we tested the hypotheses that species of similar dispersal potential (based on egg type and pelagic larval duration, PLD) have similar spatial genetic patterns, and that the effect of historical seascape features is more evident in the spatial genetic patterns of low dispersal species. We found that species dispersal potential, based on egg type and PLD, does not reliably predict which species have similar spatial genetic patterns across our studied seascape. This result contrasts with recent inferences from meta-analyses (Riginos et al. 2011, 2014, Selkoe and Toonen 2011). Furthermore, we found that dispersal potential does not predict which seascape features or processes underlie the spatial genetic patterns, despite previous studies that have demonstrated such a relationship (Pelc et al. 2009).

Chapter Four is a valuable methodological contribution to seascape genetics. A recent review of published population genetic data for Indo-Pacific marine fauna identified five clusters of locations that tended to be co-sampled, with little sampling among clusters (Keyse et al. 2014). In Chapter Four, our sampling covered all five of the identified clusters. Furthermore, we present the first genetic survey of the reef-associated fauna of Timor-Leste. Our two-tiered analysis approach whereby we analyzed co-sampled locations (i.e. those that had been sampled for all four fish species), and also the wider sampled locations available for each species, enabled effective use of all our data. Our spatially explicit co-sampling allowed quantitative comparison of spatial genetic patterns across species rarely seen in comparative phylogeography. Furthermore, results for each species were considered within the context of genetic patterns found across their broader sampled ranges. This represents the first such attempt to cross-infer comparative spatial genetic patterns using the broader "genetic-background" as a reference (Chapter Three, Bowen et al. 2014). We find that the processes inferred to

have influenced the spatial genetic patterns of species differed according to the extent of the species range that was included in the analysis.

Two findings from Chapter Four raise important considerations for spatial genetic investigations. First, our findings challenge the general assumption that shared genetic breaks and spatial structure are a prerequisite for revealing the common influence of physical processes on the dispersal and gene flow of species. Shifting our focus from patterns of structure to quantitative pairwise measures of genetic differentiation enabled us to identify common responses to the same physical features, despite differences in patterns of spatial genetic structure (e.g. the effect of the Lydekker/Weber's line on *A. triostegus* and *P. coelestis*). Second, we found that patterns of spatial genetic structure, and the inferred processes underlying patterns of genetic differentiation, changed according to the study area and the populations included in the analysis. Thus, caution must be taken when inferring common spatial structures across species that are not strictly co-sampled.

Chapter Five used extensive sampling of one species (*P. coelestis*) in combination with previously studied populations (Liu et al. 2008, 2010, 2012) to investigate how spatial genetic patterns differ across the range of a species. This chapter tested the applicability of the primary range-wide spatial genetic hypothesis, the core-periphery hypothesis (CPH) in a tropical marine species. We extended the basic expectations of the CPH in two ways. First, we considered population history that may be particularly relevant to organisms that have cryptic admixture of lineages (an issue of particular concern in the context of tropical Indo-Pacific organisms). Second, we used the latitudinal extremes (replicated in the northern and southern hemispheres) of this coral reef fish to test the generality of peripheral patterns in tropical marine organisms. We suggested two likely demographic scenarios toward the range periphery of a tropical marine species and described how these would be reflected using conventional genetic measures and alternative measures more commonly used in community ecology.

In *P. coelestis*, we established that the presence of a cryptic clade within the core of the species range had a strong influence on the inferred latitude-wide genetic patterns (Chapter Five). We also found that despite having similar levels of genetic diversity based on conventional genetic measures, the genetic composition and spatial organization of genetic diversity differed between the northern and southern periphery. Using measures of β-diversity and nestedness we discovered that the northern periphery had more unique genetic diversity than the southern periphery, and the spatial distribution of genetic diversity in the southern periphery reflects contemporary population connectivity and likely asymmetrical immigration from the core of the range. This study represents the first latitude-wide investigation of genetic patterns for a tropical marine fish and suggests that the genetic composition of populations likely vary, not only from the core to the periphery of a species range, but also at the different range peripheries of a species. Furthermore, we effectively demonstrate that measures more commonly used in other scientific disciplines can extend the analytical toolkit

currently used in phylogeography, population, and landscape genetics (Diniz-Filho et al. 2012, discussed in Chapter Three).

With the intention of understanding the generality of peripheral genetic patterns in certain seascapes, **Chapter Six** focused on one peripheral location in the range of two coral reef species (*Acanthaster planci* and *Tripneustes gratilla*). This chapter investigated whether populations of the Kermadec Islands in the South Pacific were self-sustaining, via local retention, or sustained by immigration from the core of the species range. We found that the genetic diversity of both populations at the Kermadec Islands was very low, and therefore indicative of isolated colonization events and little subsequent immigration from core populations. The pattern of low genetic diversity and contemporary isolation for these two echinoderm populations at the southern periphery of their range contrasts with the genetic patterns and inferred demographic processes in the southern periphery of *P. coelestis*' range (Chapter Five). Instead, the inferred demographic conditions for these echinoderm populations are more similar to populations in the northern periphery of *P. coelestis*' range, in which unique genetic diversity is likely to arise.

In summary, our empirical chapters suggest instances of congruence across species in their spatial genetic patterns and inferred processes, and several cases of incongruence. We found that both echinoderm populations studied at the Kermadec Islands (*A. planci* and *T. gratilla*, Chapter Six) had remarkably similar patterns of low genetic diversity. Furthermore, we found similarity in the spatial genetic patterns of two coral reef fishes, *P. coelestis* and *H. hortulanus*, that share a similar pelagic larval duration, yet have differing egg types (benthic versus pelagic, respectively; Chapter Four). The similar spatial genetic patterns observed for these fish species are likely due to a common influence of oceanographic distance on their patterns of genetic differentiation. The effect of historical barriers were also observed to have a common influence on both *P. coelestis* and *A. triostegus*, despite their varying dispersal potential, and apparently different spatial genetic patterns (Chapter Four). However, cases of incongruence in spatial genetic patterns were also frequent. Quantitative inference of overall congruence (Kendall's coefficient, Legendre and Lapointe 2004) across four coral reef fish species was very low, and did not relate to the predicted biological predictors (egg type and PLD; Chapter Four). Even within the range of a single species, a consistent relationship to geography was not observed (Chapter Five). In the northern and southern periphery of *P. coelestis* range, genetic patterns were quite different. These conflicting patterns across the range of *P. coelestis* range likely reflect differences in the population history in of the north versus the south, such as timing of colonization.

Having broad geographic coverage via our own sampling efforts and previously published work strengthened our inference. For example, it is rare to have the entire latitudinal range of a species represented in a single study (Chapter Five). The latitudinal coverage we had for *P. coelestis* allowed us to test hypotheses relating to range-wide genetic patterns with replication of the range periphery in the north and the south. Moreover, placing results into a broader geographic context was

important for Chapter Six, where the anomalous genetic patterns of populations at the Kermadec Islands could only be understood in reference to levels of genetic diversity and differentiation found at other locations for these species (*A. planci* and *T. gratilla*). Leveraging the broad geographic coverage available, we found evidence that spatial genetic patterns vary, and the importance of certain physical processes change, over different spatial scales for several coral reef fish species (Chapter Four), and in different parts of a species range (Chapter Five). For example, despite *D. trimaculatus* having chaotic genetic structure over its smaller co-sampled range, oceanographic distance was associated with the genetic patterns over its broader range (Chapter Four). In contrast, for *P. coelestis* despite genetic patterns of differentiation having a relationship with geographic distance in the northern periphery of its range, there was no relationship in the southern periphery (Chapter Five).

Overall, although replication was limited (e.g. only four species were strictly co-sampled in Chapter Four), biological traits appeared to be a poor predictor of spatial genetic patterns, incidences of congruence, and the effects of certain physical processes on spatial genetic patterns. Nonetheless, this thesis contributes several species replicates aimed at understanding the importance of historical and contemporary processes in shaping genetic patterns across the juncture of the Indian and Pacific Oceans (Chapter Four) and its peripheral reaches (Chapters Five and Six); an approach advocated by others (Avise 2000, Dawson 2014). Moreover, this thesis contributes novel methodological approaches for the discipline of seascape genetics. In particular we present methods for the quantitative comparison of spatial genetic patterns across species and we promote an analytical framework focused on identifying seascape features and processes that have a common influence across species, rather than focusing on concordant spatial genetic patterns *per se* (Chapter Four). We present examples of how measures more commonly used in other scientific disciplines might be useful in a spatial genetic context (Chapter Five). With the acquisition of more case studies, and wider application of these methods we will continue to improve our understanding of marine population connectivity.

Future directions

Open data and collaboration

The major impediments to broadscale and synthetic analyses of genetic patterns in marine species across the Indo-Pacific Ocean include data accessibility and methods of analysis. The ranges of many Indo-Pacific species are vast (Spalding 2007, Briggs and Bowen 2012). Thus, sampling that covers a species entire range is challenging and beyond the scope of any individual labs' research agenda (discussed in Keyse et al. 2014). However, if data are shared and the appropriate metadata is supplied, there can be communal benefits far greater than that afforded by an individual's research efforts (Barber et al. 2014). Such open-data movements require a means to exchange and access data, and a data form that can be easily re-appropriated for other research questions.

In the field of Indo-Pacific population genetics, an inclusive international collaborative network of scientists has come together to coordinate a data sharing policy, mechanisms to archive and share genetic data, and collaborate on projects (Diversity of the Indo-Pacific Network, DIPnet, http://indopacificnetwork.wikispaces.com; led by C. Riginos, E. D. Crandall, and R. Toonen; discussed in Crandall and Riginos 2014). I have been an active member of DIPnet, helping to build bioinformatic pipelines to facilitate data upload and data checking. All the data within my thesis meets the metadata requirements of the collaborative network and has been uploaded to the database for other DIPnet members to use. As of July 2014, DIPnet has 48 members and the database holds over 200 population genetic datasets, representing over 1,500 locations throughout the Indo-Pacific region, and comprising over 30,000 mtDNA sequences. The network endeavors to extend the geographic coverage for individual species and increase replication across species enabling better opportunities for hypothesis testing, study design, research coordination and collaboration.

Mitochondrial DNA is the marker of choice in the initial coordination of DIPnet's research activity. Although the use of mtDNA to address questions of marine population connectivity can be coarse (discussed in Chapter Three), this locus can be very useful in specific contexts (Bohonak and Vandergast 2011, Bowen et al. 2014) and particularly in synthetic analyses and collaborative ventures (Keyse et al. 2014). Most studies of Indo-Pacific marine organisms use mtDNA and these studies tend to have broader geographic coverage than those based on other genetic markers (Keyse et al. 2014). Furthermore, mtDNA sequence data can be exchanged among collaborators with fewer concerns about reliability (Keyse et al. 2014). For example, the most comprehensive seascape genetic investigation for a single species to date was facilitated via a cross-laboratory collaboration made possible by the common use of the mtDNA locus (Cytochrome Oxidase 1, *Linckia laevigata*, Crandall et al. 2014 - Appendix Two).

Although there are advantages to using genome wide genetic markers for questions related to marine population (discussed below), these methods are prohibitively technical and/or expensive for some researchers, and especially for questions requiring a comparative approach (Bowen et al. 2014). Instead the use of mtDNA and next-generation sequencing (NGS) methods should be considered as complementary. For example, Bowen et al. (2014) suggest a progressive research program, whereby comparative genetic surveys of multiple species based on mtDNA or a few loci are used to identify questions, locations and species of interest, which can then be pursued using targeted population genomic investigations. Thus, although the data used within this thesis is restricted to mtDNA, it sets the stage for follow up studies using NGS methods. Moreover, the hypothetical frameworks, and analyses used herein are appropriate for any type of genetic data.

Methodological and analytical advances

Ultimately, progression in the field of comparative spatial genetics is contingent on developing quantitative frameworks for analyzing data across multiple taxa. Regardless of the genetic marker

used, the analytical challenges of combining genetic data across multiple species and inferring influential biological and physical factors persist. Many of these challenges were considered in Chapter Four of this thesis and have also been discussed elsewhere (Andrew et al. 2013, Bowen et al. 2014, Dawson 2014, Selkoe et al. 2014). Some of these challenges can be circumvented via experimental design. For example, using the same molecular marker/gene region across species, using closely related species or phylogenetically distinct species (depending on your justification of confidence), and/or sampling all species from the same locations (discussed in Chapters Two and Three, Dawson 2014, Dawson et al. in press).

There are two distinct benefits of uniformly co-sampling species across a seascape. First, the seascape features considered to have a role in shaping the spatial genetic patterns can be held constant across datasets. Second, such a study design allows the investigator to access a wider variety of statistical methods, without having to accommodate missing data or uneven sampling effects. Where species have been co-sampled, the methods presented herein could prove particularly useful. In Chapter Four we used a matrix comparison approach (congruence among distance matrices, Legendre and Lapointe 2004) to infer congruence in the spatial genetic patterns of four species of coral reef fish. This method was designed to test the hypothesis that several matrices, containing different types of variables about the same objects (e.g. populations), are congruent with one another, in order to address whether they may be jointly analyzed (Legendre and Lapointe 2004). With the application of this method to many more matrices (i.e. species) it is likely that suites of species with congruent spatial genetic patterns will emerge. These different suites may then be independently analyzed in order to deduce the biological and physical factors that may underlie their common spatial genetic patterns, such as dispersal-related traits, reproductive phenology, ecological traits, or population histories, for example. Analyzing suites of species, rather than *all* species, may reduce the noise associated with different responses across species to a common underlying factor, and provide higher statistical confidence in subsequent analyses.

Unfortunately very few spatial genetic datasets contain several species that are co-sampled. Nonetheless, with the use of many datasets across the same seascape, presumably there is increased power to accommodate issues associated with disparate sampling schemes. Traditionally, spatially-implicit summary statistics have been used to compare spatial genetic patterns across species that have no been strictly co-sampled, such as the IBD slope (e.g. Selkoe and Toonen 2011) and global F_{ST} (Riginos et al. 2011). These measures are very informative when testing hypotheses related to the role of species biology in determining genetic patterns (e.g. Riginos et al. 2011), but are less useful when testing spatially-explicit hypotheses. We used a two-tier approach to make use of both the co-sampled genetic data across species as well as the information from other studied populations across each species range in Chapter Four, however there are several alternative approaches that warrant investigation.

The synthesis of genetic data across species that have not been strictly co-sampled could be facilitated via the use of alternate genetic measures. For example, conditional genetic distances (*c*GD, Dyer and Nason 2004) could enable focal pairwise relationships (i.e. among co-sampled populations) to be conditioned on known relationships with other sampled populations within each species range. Alternatively, methods for imputing or interpolating values can be used to visualize regions of the landscape or seascape where species have congruent and incongruent spatial genetic patterns (e.g. Vandergast et al. 2011). However, caution should be taken when interpolating genetic "surfaces" between patchy reef habitats where the study organisms do not reside. Indeed, the validity of both these approaches requires thorough testing, and sensitivity analyses to guide appropriate thresholds for their use (e.g. minimum number of populations, variance in number of populations used across species for focal genetic relationships to be conditioned upon; and appropriate interpolation distance, respectively). In contrast, some methods that are well established in single species investigations may be sensibly adjusted for multi-species investigations. One example is the use of Hierarchical Approximate Bayesian Computation (HABC) that uses individual coalescent models for each species as the units of replication to infer community dynamics (Hickerson and Meyer 2008, Beaumont 2010). With the forthcoming increase in data availability via synthetic and collaborative initiatives, and technological advances (discussed below), it will necessary to develop and refine methods of analysis that are capable of handling large spatial genetic datasets, and that accommodate the testing of hypotheses relevant to the field of seascape genetics and our understanding of marine population connectivity.

Technological advances

Recent technological advances in sequencing and the availability of environmental data offer unparalleled opportunities for understanding the processes determining population connectivity, and the importance of migration versus other evolutionary forces in forming spatial genetic patterns. Next-generation sequencing technologies provide large numbers of loci (usually single nucleotide polymorphisms, SNPs) allowing unprecedented resolution in seascape genetic and genomic studies. These markers enable the simultaneous analysis of background genetic patterns created by neutral processes and patterns indicative of natural selection (Joost et al. 2007, Beaumont 2010). Even for non-model systems, such as most marine organisms, it is now feasible to link loci suspected to be under selection to functional genes using databases of transcriptome sequences (Andrew et al. 2013). Whereas this thesis has emphasized the utility of seascape genetics in a neutral context, there will likely be a greater focus on adaptive genetic variation in future, as is the case in the discipline of landscape genetics (Lowry 2010, Manel et al. 2010, Schoville et al. 2012).

An understanding of the role of ecology and selective forces in determining population connectivity and spatial genetic patterns in marine systems is vital. It is well accepted that biogeography is determined by both ecological and historical parameters (de Candolle 1820). While it has been recognized that ecology is also likely to play an important role in determining spatial genetic

patterns in marine systems (e.g. Choat 2006, Selkoe et al. 2010), investigations have been restricted by coarse genetic markers and limited environmental data (Andrew et al. 2013). However, with the increased power of hundreds and thousands of loci offered by next-generation methods and newly available environmental databases for marine ecosystems (Sbrocco and Barber 2013, Sbrocco 2014) the field of seascape genetics is now well-placed to address the role of ecology and selection in forming spatial genetic patterns.

With the recent surge in data availability there is the temptation for the field of seascape genetics to become increasingly descriptive. The large amounts of hyper variable loci and endless number of environmental variables to which their spatial patterns can be correlated will not cease to provide an interesting story in the absence of hypothesis testing. However, such an approach is unlikely to progress our understanding of population connectivity in marine systems in a way that can inform fisheries management, conservation decisions, and climate-change mitigation. There are several outstanding questions in the field of seascape genetics and concerning the population connectivity of reef-associated organisms that warrant consideration within appropriate hypothesis testing frameworks. To my mind, the most simple and consequential of these is the role of gene flow (i.e. larval dispersal for a reef-associated marine organism) in promoting and constraining local selection across a species range. This question may be of particular importance at the leading-edge of a species range where ocean currents and or environmental conditions may be in flux. To date, this question has received much theoretical attention (e.g. Holt and Gomulkiewicz 1997, Debarre et al. 2013), but there has been little empirical evaluation despite the recent availability of data to address such needs.

LITERATURE CITED

Ackiss AS, Pardede S, Crandall ED, Ablan-Lagman MCA, Ambariyanto, Romena N, Barber PH and Carpenter KE (2013) Pronounced genetic structure in a highly mobile coral reef fish, *Caesio cuning*, in the Coral Triangle. *Marine Ecology Progress Series* **480**: 185–197.

Alberto F, Raimondi PT, Reed DC, Watson JR, Siegel DA, Mitarai S, Coelho N and Serrao EA (2011) Isolation by oceanographic distance explains genetic structure for *Macrocystis pyrifera* in the Santa Barbara Channel. *Molecular Ecology* **20**: 2543–2554.

Aljanabi SM and Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA from PCR-based techniques. *Nucleic Acids Research* **25**: 4692–4693.

Allen GR (1991). *Damselfishes of the World*, Mergus Verlag, Melle, Germany.

Almany GR, Berumen ML, Thorrold SR, Planes S and Jones GP (2007) Local replenishment of coral reef fish populations in a marine reserve. *Science* **316**: 742–744.

Almany GR, Hamilton RJ, Bode M, Matawai M, Potuku T, Saenz-Agudelo P, Planes S, Berumen ML, Rhodes KL, Thorrold SR, Russ GR and Jones GP (2013) Dispersal of grouper larvae drives local resource sharing in a coral reef fishery. *Current Biology : CB* **23**: 626–630.

Almany GR and Webster MS (2006) The predation gauntlet: early post-settlement mortality in reef fishes. *Coral Reefs* **25**: 19–22.

Almeida-Neto M, Guimarães P, Guimarães PR, Loyola RD and Ulrich W (2008) A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117**: 1227–1239.

Alvarez N, Thiel-Egenter C, Tribsch A, Holderegger R, Manel S, Schonswetter P, Taberlet P, Brodbeck S, Gaudeul M, Gielly L, Kupfer P, Mansion G, Negrini R, Paun O, Pellecchia M, Rioux D, Schupfer F, Van Loo M, Winkler M and Gugerli F (2009) History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecology Letters* **12**: 632–640.

Anderson CD, Epperson BK, Fortin M-J, Holderegger R, James PMA, Rosenberg MS, Scribner KT and Spear S (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology* **19**: 3565–3575.

Andrew RL, Bernatchez L, Bonin A, Buerkle CA, Carstens BC, Emerson BC, Garant D, Giraud T, Kane NC, Rogers SM, Slate J, Smith H, Sork VL, Stone GN, Vines TH, Waits L, Widmer A and Rieseberg LH (2013) A road map for molecular ecology. *Molecular Ecology* **22**: 2605–2626.

Antonovics J (1976) The nature of limits to natural selection. *Annals of the Missouri Botanical Garden*: 224-247.

Araujo MB (2002) Biodiversity hotspots and zones of ecological transition. *Conservation Biology* **16**: 1662-1663.

Assis J, Castilho Coelho N, Alberto F, Valero M, Raimondi P, Reed D and Serrao EA (2013) High and distinct range-edge genetic diversity despite local bottlenecks. *PLoS ONE* **8**: e68646.

Austin D, Bowen WD and McMillan JI (2004) Intraspecific variation in movement patterns: modeling individual behaviour in a large marine predator. *Oikos* **105**: 15–30.

Avise JC (1992) Molecular popualtion structure and the biogeographic history of a regional fauna: a case study with lessons for conservation biology. *Oikos* **63**: 62–76.

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

Avise JC and Ball RMJ (1990) Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surveys in Evolutionary Biology* **7**: 45–67.

Ayre DJ and Hughes TP (2004) Climate change, genotypic diversity and gene flow in reef-building corals. *Ecology Letters*: 273–278.

Ayre DJ, Minchinton TE and Perrin C (2009) Does life history predict past and current connectivity for rocky intertidal invertebrates across a marine biogeographic barrier? *Molecular Ecology* **18**: 1887–1903.

Baird AH, Sommer B and Madin JS (2012) Pole-ward range expansion of *Acropora* spp. along the east coast of Australia. *Coral Reefs* **31**: 1063–1063.

Balkenhol N, Gugerli F, Cushman SA, Waits LP, Coulon A, Arntzen JW, Holderegger R and Wagner HH (2009) Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology* **24**: 455–463.

Bandelt H-J, Forster P and Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**: 37–48.

Banks SC, Ling SD, Johnson CR, Piggott MP, Williamson JE and Beheregaray LB (2010) Genetic structure of a recent climate change-driven range extension. *Molecular Ecology* **19**: 2011–2024.

Banks SC, Piggott MP, Williamson JE, Bové U, Holbrook NJ and Beheregaray LB (2007) Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin. *Ecology* **88**: 3055–3064.

Barber P, Erdmann MV and Palumbi SR (2006) Comparative phylogeography of three codistributed stomatopods: origins and timing of regional lineage diversification in the Coral Triangle. *Evolution* **60**: 1825–1839.

Barber P, Palumbi SR, Erdmann MV and Moosa MK (2002a) Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. *Molecular Ecology* **11**: 659–674.

Barber PH, Ablan-Lagman MCA, Berlinck RGS, Cahyani D, Crandall ED, Ravago-Gotanco R, Juinio-Meñez MA, Mahardika IGN, Shanker K, Starger CJ, Toha AHA, Anggoro AW and Willette DA (2014) Advancing biodiversity research in developing countries: the need for changing paradigms. *Bulletin of Marine Science* **90**: 187–210.

Barber PH, Moosa MK and Palumbi SR (2002b) Rapid recovery of genetic diversity of stomatopod populations on Krakatau: temporal and spatial scales of marine larval dispersal. *Proceedings of the Royal Society B: Biological Sciences* **269**: 1591–1597.

Barton NH (2001). Adaptation at the edge of a species' range. In: *Integrating Ecology and Evolution in a Spatial Context.* J Silvertown and J Antonovics (eds.). Oxford, UK, Blackwell Science.

Baselga A (2010) Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography* **19**: 134–143.

Baselga A (2012) The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Global Ecology and Biogeography* **21**: 1223–1232.

Baselga A and Orme CDL (2012) betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution* **3**: 808–812.

Bataillon T (2003) Shaking the 'deleterious mutations' dogma? *Trends in Ecology and Evolution* **18**: 315–317.

Baums IB, Miller MW and Hellberg ME (2005) Regionally isolated populations of an imperiled Caribbean coral, *Acropora palmata*. *Molecular Ecology* **14**: 1377–1390.

Baums IB, Paris CB and Cherubin LM (2006) A bio-oceanographic filter to larval dispersal in a reef-building coral. *Limnology and Oceanography* **51**: 1969–1981.

Bay LK and Caley MJ (2011) Greater genetic diversity in spatially restricted coral reef fishes suggests secondary contact among differentiated dineages. *Diversity* **3**: 483–502.

Bay LK, Choat JH, van Herwerden L and 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? *Marine Biology* **144**: 757–767.

Beale CM, Lennon JJ, Yearsley JM, Brewer MJ and Elston DA (2010) Regression analysis of spatial data. *Ecology Letters* **13**: 246–264.

Beaumont MA (2010) Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics* **41**: 379–406.

Beerli P and Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* **152**: 763–773.

Beger M, Selkoe KA, Treml E, Barber PH, von der Heyden S, Crandall ED, Toonen RJ and Riginos C (2014) Evolving coral reef conservation with genetic information. *Bulletin of Marine Science* **90**: 159–185.

Bellwood DR, Renema W and Rosen BR (2012). Biodiversity hotspots, evolution and coral reef biogeography: a review. In: *Biotic Evolution and Environmental Change in Southeast Asia*. DJ Gower, K Johnson, J Richardson (eds.) Cambridge, UK, Cambridge Univeristy Press.

Benjamini Y and Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)* **57**: 289–300.

Benzie JAH (1999) Genetic structure of coral reef organisms: ghosts of dispersal past. *American Zoologist* **39**: 131–145.

Bernardi G, Holbrook SJ and Schmitt RJ (2001) Geneflow at three spatial scales in a coral reef fish, the three-spot dascyllus, *Dascyllus trimaculatus*. *Marine Biology* **138**: 457–465.

Bernardi G, Holbrook SJ, Schmitt RJ and Crane NL (2003) Genetic evidence for two distinct clades in a French Polynesian population of the coral reef three-spot damselfish Dascyllus trimaculatus. *Marine Biology* **143**: 485–490.

Bierne N, Welch J, Loire E, Bonhomme F and David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**: 2044–2072.

Billot C, Engel CR, Rousvoal S, Kloareg B and Valero M (2003) Current patterns, habitat discontinuities and population genetic structure: the case of the kelp *Laminaria digitata* in the English Channel. *Marine Ecology Progress Series* **253**: 111–121.

Bird CE, Holland BS, Bowen BW and Toonen RJ (2007) Contrasting phylogeography in three endemic Hawaiian limpets (Cellana spp.) with similar life histories. *Molecular Ecology* **16**: 3173–3186.

Bird CE, Karl SA, Smouse PE and Toonen RJ (2011) Detecting and measuring genetic differentiation. *Crustacean Issues* **19**: 31–55.

Birkeland C and Lucas JS (1990). Acanthaster planci: *major management problem of coral reefs*. Boca Raton, CRC Press.

Bodmer WF and Cavalli-Sforza LL (1967) A migration matrix model for the study of random genetic drift. *Genetics* **59**: 565–592.

Bohonak AJ and Vandergast AG (2011) The value of DNA sequence data for studying landscape genetics. *Molecular Ecology* **20**: 2477–2479.

Bongaerts P, Riginos C, Hay KB, van Oppen MJ, Hoegh-Guldberg O and Dove S (2011) Adaptive divergence in a scleractinian coral: physiological adaptation of *Seriatopora hystrix* to shallow and deep reef habitats. *BMC Evolutionary Biology* **11**: 303.

Bongaerts P, Riginos C, Ridgway T, Sampayo EM, van Oppen MJH, Englebert N, Vermeulen F and Hoegh-Guldberg O (2010) Genetic divergence across habitats in the widespread coral *Seriatopora hystrix* and its associated *Symbiodinium*. *PLoS ONE* **5**: e10871.

Booth DJ, Bond N and Macreadie P (2011) Detecting range shifts among Australian fishes in response to climate change. *Marine and Freshwater Research* **62**: 1027–1042.

Booth DJ, Figueira WF, Gregson MA, Brown L and Beretta G (2007) Occurrence of tropical fishes in temperate southeastern Australia: Role of the East Australian Current. *Estuarine, Coastal and Shelf Science* **72**: 102–114.

Bowen BW, Rocha LA, Toonen RJ and Karl SA (2013) The origins of tropical marine biodiversity. *Trends in Ecology and Evolution* **28**: 359–366.

Bowen BW, Shanker K, Yasuda N, Celia M, Malay MCD, von der Heyden S, Paulay G, Rocha LA, Selkoe KA, Barber PH, Williams ST, Lessios HA, Crandall ED, Bernardi G, Meyer CP, Carpenter KE and Toonen RJ (2014) Phylogeography unplugged: comparative surveys in the genomic era. *Bulletin of Marine Science* **90**: 13–46.

Bradbury IR, Laurel B, Snelgrove PV, Bentzen P and Campana SE (2008) Global patterns in marine dispersal estimates: the influence of geography, taxonomic category and life history. *Proceedings of the Royal Society B: Biological Sciences* **275**: 1803–1809.

Briggs JC (1999) Modes of speciation: marine Indo-West Pacific. *Bulletin of Marine Science* **65**: 645–656.

Briggs JC and Bowen BW (2012) A realignment of marine biogeographic provinces with particular reference to fish distributions. *Journal of Biogeography* **39**: 12–30.

Brook FJ (1998) The coastal molluscan fauna of the northern Kermadec Islands, Southwest Pacific Ocean. *Journal of the Royal Society of New Zealand* **28**: 185–233.

Brook FJ (1999) The coastal scleractinian coral fauna of the Kermadec Islands, southwestern Pacific Ocean. *Journal of the Royal Society of New Zealand* **29**: 435–460.

Brussard PF (1984) Geographic patterns and environmental gradients: the central-marginal model in *Drosophila* revisited. *Annual Review of Ecology and Systematics* **15**: 25–64.

Buston PM, Fauvelot C, Wong MYL and Planes S (2009) Genetic relatedness in groups of the humbug damselfish *Dascyllus aruanus*: small, similar-sized individuals may be close kin. *Molecular Ecology* **18**: 4707–4715.

Caley MJ, Carr MH, Hixon MA, Hughes TP, Jones GP and Menge BA (1996) Recruitment and the local dynamics of open marine populations. *Annual Review of Ecology and Systematics* **27**: 477–500.

Calinski T and Harabasz J (1974) A dendrite method for cluster analysis. *Communications in Statistics* **3**: 1–27.

Carpenter KE, Barber PH, Crandall ED, Carmen M, Ablan-Lagman A, Ambariyanto, Mahardika GN, Manjaji-Matsumoto BM, Juinio-Meñez MA, Santos MD, Starger CJ and Toha AHA (2010) Comparative phylogeography of the Coral Triangle and implications for marine management. *Journal of Marine Biology* **2011**: 1–14.

Carr MH, Niegel JE, Estes JA, Andelman S, Warner RR and Largier J (2003) Comparing marine and terrestrial ecosystems: implications for the design of coastal marine reserves. *Ecological Applications* **13**: S90–S107.

Carvalho JC, Cardoso P and Gomes P (2012) Determining the relative roles of species replacement and species richness differences in generating beta-diversity patterns. *Global Ecology and Biogeography* **21**: 760–771.

Casale P, Laurent L, Gerosac G and Arganod R (2002) Molecular evidence of male-biased dispersal in loggerhead turtle juveniles. *Journal of Experimental Marine Biology and Ecology* **267**: 139–145.

Case RAJ, Hutchinson WF, Hauser L, Van Oosterhout C and Carvalho GR (2005) Macro-and micro-geographic variation in *pantophysin* (*Pan I*) allele frequencies in NE Atlantic cod *Gadus morhua*. *Marine Ecology Progress Series* **301**: 267–278.

Caswell (1989). *Matrix population models: construction, analysis and interpretation*. Sunderland, Sinauer.

Channell R and Lomolino MV (2001) Trajectories to extinction: spatial dynamics of the contraction of geographical ranges. *Journal of Biogeography* **27**: 169–179.

Chappell J and Shackleton NJ (1986) Oxygen isotopes and sea level. *Nature* **324**: 137–140.

Choat JH (2006) Phylogeography and reef fishes: bringing ecology back into the argument. *Journal of Biogeography* **33**: 967–968.

Clarke RT, Rothery P and Raybould AF (2002) Confidence limits for regression relationships between distance matrices: Estimating gene flow with distance. *Journal of Agricultural, Biological, and Environmental Statistics* **7**: 361–372.

Cole RG, Creese RG, Grace RV, Irving P and Jackson BR (1992) Abundance patterns of subtidal benthic invertebrates and fishes at the Kermadec Islands. *Marine Ecology Progress Series* **82**: 207–218.

Condie SA, Waring J, Mansbridge JV and Cahill ML (2005) Marine connectivity patterns around the Australian continent. *Environmental Modelling and Software* **20**: 1149–1157.

Connolly SR and Baird AH (2010) Estimating dispersal potential for marine larvae: dynamic models applied to scleractinian corals. *Ecology* **91**: 3572–3583.

Cowen RK (2000) Connectivity of marine populations: open or closed? *Science* **287**: 857–859.

Cowen RK (2006) Scaling of connectivity in marine populations. *Science* **311**: 522–527.

Cowen RK, Gawarkiewicz G, Pineda J, Thorrold SR and Werner FE (2007) Marine connectivity in marine systems: an overview. *Oceanography* **20**: 14–21.

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

Cowman PF and Bellwood DR (2012) The historical biogeography of coral reef fishes: global patterns of origination and dispersal. *Journal of Biogeography* **40**: 209–224.

Coykendall DK, Johnson SB, Karl SA, Lutz RA and Vrijenhoek RC (2011) Genetic diversity and demographic instability in *Riftia pachyptila* tubeworms from eastern Pacific hydrothermal vents. *BMC Evolutionary Biology* **11**: 96.

Crandall ED, Frey MA, Grosberg RK and Barber PH (2008) Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. *Molecular Ecology* **17**: 611–626.

Crandall ED and Riginos C (2014) Magnificent dimensions, varied forms, and brilliant colors: the molecular ecology and evolution of the Indian and Pacific oceans. *Bulletin of Marine Science* **90**: 1–11.

Crandall ED, Treml EA and Barber PH (2012) Coalescent and biophysical models of stepping-stone gene flow in Neritid snails. *Molecular Ecology* **21**: 5579–5598.

Crandall ED, Treml EA, Liggins L, Gleeson L, Yasuda N, Barber PH, Wörheide G and Riginos C (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.

Cushman SA and Landguth EL (2010) Spurious correlations and inference in landscape genetics. *Molecular Ecology* **19**: 3592–3602.

da Cunha AB, Burla H and Dobzhansky T (1950) Adaptive chromosomal polymorphism in *Drosophila willistoni*. *Evolution* **4**: 212–235.

Dale MRT and Fortin MJ (2010) From graphs to spatial graphs. *Annual Review of Ecology, Evolution, and Systematics* **41**: 21–38.

Darlington PJ (1957). *Zoogeography: the geographical distribution of animals*, Wiley.

Darriba D, Taboada GL, Doallo R and Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature methods* **9**: 772.

Daubenmire R (1975) Floristic plant geography of eastern Washington and northern Idaho. *Journal of Biogeography* **2**: 1–18.

David GK, Marshall DJ and Riginos C (2010) Latitudinal variability in spatial genetic structure in the invasive ascidian, *Styela plicata*. *Marine Biology* **157**: 1955–1965.

Dawson MN (2005) Rapid evolutionary radiation of marine zooplankton in peripheral environments. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 9235–9240.

Dawson MN (2012) Parallel phylogeographic structure in ecologically similar sympatric sister taxa. *Molecular Ecology* **21**: 987–1004.

Dawson MN, Grosberg RK, Stuart YE and Sanford E (2010) Population genetic analysis of a recent range expansion: mechanisms regulating the poleward range limit in the volcano barnacle *Tetraclita rubescens*. *Molecular Ecology* **19**: 1585–1605.

Dawson MN and Hamner WM (2008) A biophysical perspective on dispersal and the geography of evolution in marine and terrestrial systems. *Journal of the Royal Society Interface* **5**: 135–150.

Dawson MN, Hays CG, Grosberg RK and Raimondi PT. Dispersal potential and population genetic structure in the marine intertidal of the eastern North Pacific. *Ecological Monographs*: in press

Dawson MN and Phillimore A (2014) Natural experiments and meta-analyses in comparative phylogeography. *Journal of Biogeography* **41**: 52–65.

de Candolle AP (1820). *Géographie botanique*. Strasbourg and Paris, F.G. Levrault, imprimeur du Roi.

Debarre F, Ronce O and Gandon S (2013) Quantifying the effects of migration and mutation on adaptation and demography in spatially heterogeneous environments. *Journal of Evolutionary Biology* **26**: 1185–1202.

DeBoer TS, Naguit MRA, Erdmann MV, Ablan-Lagman MCA, Carpenter KE, Toha AHA and Barber PH (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.

Defeo O and Cardoso RS (2004) Latitudinal patterns in abundance and life-history traits of the mole crab *Emerita brasiliensis* on South American sandy beaches. *Diversity and Distributions* **10**: 89–98.

Dekshenieks MM, Hofmann EE, Klinck JM and Powell EN (1996) Modeling the vertical distribution of oyster larvae in response to environmental conditions. *Marine Ecology Progress Series* **136**: 97–110.

Diniz-Filho JA, Collevatti RG, Soares TN and de Campos Telles MP (2012) Geographical patterns of turnover and nestedness-resultant components of allelic diversity among populations. *Genetica* **140**: 189–195.

Diniz-Filho JAF and Bini LM (2011) Geographical patterns in biodiversity: towards an integration of concepts and methods from genes to species diversity. *Natureza and Conservação* **9**: 179–187.

Doherty PJ, Planes S and Mather P (1995) Gene flow and larval duration in seven species of fish from the Great Barrier Reef. *Ecology* **76**: 2373–2391.

Doney SC, Ruckelshaus M, Duffy JE, Barry JP, Chan F, English CA, Galindo HM, Grebmeier JM, Hollowed AB, Knowlton N and al. e (2012) Climate change impacts on marine ecosystems. *Annual Review of Marine Science* **4**: 11–37.

Drew J, Allen GR, Kaufman L and Barber PH (2008) Endemism and regional color and genetic differences in five putatively cosmopolitan reef fishes. *Conservation Biology* **22**: 965–975.

Drew J and Barber PH (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.

Drew JA and Barber PH (2012) Comparative phylogeography in Fijian coral reef fishes: a multi-taxa approach towards marine reserve design. *PLoS ONE* **7**: e47710.

Dyer RJ and Nason JD (2004) Population graphs: the graph theoretic shape of genetic structure. *Molecular Ecology* **13**: 1713–1727.

Dyer RJ, Nason JD and Garrick RC (2010) Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology* **19**: 3746–3759.

Eckert CG, Samis KE and Lougheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology* **17**: 1170–1188.

El Mousadik A and Petit EJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics* **92**: 832–839.

Elith J and Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* **40**: 677–697.

Epling C and Dobzhansky T (1942) Genetics of natural populations. VI. Microgeographic races in *Linanthus parryae*. *Genetics* **27**: 317–332.

Epperson BK, McRae BH, Scribner KIM, Cushman SA, Rosenberg MS, Fortin M-J, James PMA, Murphy M, Manel S, Legendre P and Dale MRT (2010) Utility of computer simulations in landscape genetics. *Molecular Ecology* **19**: 3549–3564.

Excoffier L and Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**: 564–567.

Excoffier L, Smouse P and Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes : application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.

Farquhar H (1897) A contribution to the history of New Zealand echinoderms. *Journal of the Linnean Society of London, Zoology* **26**: 186–198.

Faugeron S, Martínez EA, Correa JA, Cardenas L, Destombe C and Valero M (2004) Reduced genetic diversity and increased population differentiation in peripheral and overharvested populations *Ofgigartina skottsbergii* (Rhodophyta, Gigartinales) in Southern Chile. *Journal of Phycology* **40**: 454–462.

Faurby S and Barber PH (2012) Theoretical limits to the correlation between pelagic larval duration and population genetic structure. *Molecular Ecology* **21**: 3419–3432.

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

Feary DA, Pratchett MS, J Emslie M, Fowler AM, Figueira WF, Luiz OJ, Nakamura Y and Booth DJ (2013) Latitudinal shifts in coral reef fishes: why some species do and others do not shift. *Fish and Fisheries*: http://dx.doi.org/10.1111/faf.12036.

Feder ME and Mitchell-Olds T (2003) Evolutionary and ecological functional genomics. *Nature Reviews* **4**: 649–655.

Fievet V, Touzet P, Arnaud JF and Cuguen J (2007) Spatial analysis of nuclear and cytoplasmic DNA diversity in wild sea beet (*Beta vulgaris* ssp *maritima*)populations: do marine currents shape the genetic structure? *Molecular Ecology* **16**: 1847–1864.

Figueira WF, Biro P, Booth DJ and Valenzuela VC (2009) Performance of tropical fish recruiting to temperate habitats: role of ambient temperature and implications of climate change. *Marine Ecology Progress Series* **384**: 231–239.

Figueira WF and Booth DJ (2010) Increasing ocean temperatures allow tropical fishes to survive overwinter in temperate waters. *Global Change Biology* **16**: 506–516.

Finn MD and Kingsford MJ (1996) Two-phase recruitment of Apogonids (Pisces) on the Great Barrier Reef. *Marine and Freshwater Research* **47**: 423–432.

Foll M and Gaggiotti O (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics* **174**: 875–891.

Fontaine MC, Baird SJE, Piry S, Ray N, Tolley KA, Duke S, Birkun A, Ferreira M, Jauniaux T, Llavona Á, Öztürk B, A Öztürk A, Ridoux V, Rogan E, Sequeira M, Siebert U, Vikingsson GA, Bouquegneau J-M and Michaux JR (2007) Rise of oceanographic barriers in continuous populations of a cetacean: the genetic structure of harbour porpoises in Old World waters. *BMC Biology* **5**: doi: 10.1186/1741-7007-1185-1130.

Forcada J and Robinson SL (2006) Population abundance, structure and turnover estimates for leopard seals during winter dispersal combining tagging and photo-identification data. *Polar Biology* **29**: 1052–1062.

Fortuna MJ, Albaladejo RG, Fernández L, Aparicio A and Bascompte J (2009) Networks of spatial genetic variation across species. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 19044–19049.

Foster NL, Paris CB, Kool JT, Baums IB, Stevens JR, Sanchez JA, Bastidas C, Agudelo C, Bush P, Day O, Ferrari R, Gonzalez P, Gore S, Guppy R, McCartney MA, McCoy C, Mendes J, Srinivasan A, Steiner S, Vermeij MJA, Weil E and Mumby PJ (2012) Connectivity of Caribbean coral populations: complementary insights from empirical and modelled gene flow. *Molecular Ecology* **21**: 1143–1157.

Francis MP, Grace RV and Paulin CD (1987) Coastal fishes of the Kermadec Islands. *New Zealand Journal of Marine and Freshwater Research* **21**: 1–13.

Francis MP, Worthington CJ, Saul P and Clements KD (1999) New and rare tropical and subtropical fishes from northern New Zealand. *New Zealand Journal of Marine and Freshwater Research* **33**: 571–586.

Frey MA and 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. *Molecular Phylogenetics and Evolution* **48**: 1067–1086.

Gaither MR, Bowen BW, Bordenave TR, Rocha LA, Newman SJ, Gomez JA, van Herwerden L and Craig MT (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**: 189.

Gaither MR, Toonen RJ, Robertson DR, Planes S and Bowen BW (2009) Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira* and *Lutjanus fulvus*). *Journal of Biogeography* **37**: 133–147.

Galarza JA, Carreras-Carbonell J, Macpherson E, Pascual M, Roques S, Turner GF and Rico C (2009) The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 1473–1478.

Galindo H, Olson D and Palumbi SR (2006) Seascape genetics: A coupled oceanographic-genetic model predicts population structure of Caribbean corals. *Current Biology* **16**: 1622–1626.

Galindo HM, Pfeiffer-Herbert AS, McManus MA, Chao Y, Chai F and Palumbi SR (2010) Seascape genetics along a steep cline: using genetic patterns to test predictions of marine larval dispersal. *Molecular Ecology* **19**: 3692–3707.

Gallego A, North EW and Petitgas P (2007) Introduction: status and future of modelling physical-biological interactions during the early life of fishes. *Marine Ecology Progress Series* **347**: 121–126.

Galtier N, Nabholz B, Glemin S and Hurst GD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology* **18**: 4541–4550.

García-Ramos G and Kirkpatrick M (1997) Genetic models of adaptation and gene flow in peripheral populations. *Evolution* **51**: 21–28.

Gardner JPA, Curwen MJ, Long J, Williamson RJ and Wood AR (2006) Benthic community structure and water column characteristics at two sites in the Kermadec Islands Marine Reserve, New Zealand. *New Zealand Journal of Marine and Freshwater Research* **40**: 179–194.

Garroway CJ, Bowman J, Carr D and Wilson PJ (2008) Applications of graph theory to landscape genetics. *Evolutionary Applications* **1**: 620–630.

Gell FR and Roberts CM (2003) Benefits beyond boundaries: the fishery effects of marine reserves. *Trends in Ecology and Evolution* **18**: 448–455

Gerber S, Chabrier P and Kremer A (2003) FAMOZ: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Molecular Ecology Notes* **3**: 479–481.

Gerlach G, Atema J, Kingsford MJ, Black KP and Miller-Sims V (2007) Smelling home can prevent dispersal of reef fish larvae. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 858–863.

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

Gilg MR and Hilbish TJ (2003) The geography of marine larval dispersal: coupling genetics with fine-scale physical oceanography. *Ecology* **84**: 2989–2998.

Gill BJ (1997) Records of turtles and sea snakes in New Zealand,1837–1996. *New Zealand Journal of Marine and Freshwater Research* **31**: 477–486.

Gilman S (2005) A test of Brown's principle in the intertidal limpet *Collisella scabra* (Gould, 1846). *Journal of Biogeography* **32**: 1583–1589.

Goldberg CS and Waits LP (2010) Quantfication and reduction of bias from sampling larvae to infer population and landscape genetic structure. *Molecular Ecology Resources* **10**: 304–313.

Gomulkiewicz R, Holt RD and Barfield M (1999) The effects of density dependence and immigration on local adaptation and niche evolution in a black-hole sink environment. *Theoretical Population Biology* **5**: 283–296.

González-Wangüemert M, Canovas F, Marcos C and Perez-Ruzafa A (2009) Phosphoglucose isomerase variability of *Cerastoderma glaucum* as a model for testing the influence of environmental conditions and dispersal patterns through quantitative ecology approaches. *Biochemical Systematics and Ecology* **37**: 325–333.

Gray JS (1997) Marine biodiversity: patterns, threats and conservation needs. *Biodiversity and Conservation* **6**: 153–175.

Grimm V and Railsback SF (2012) Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **367**: 298–310.

Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke HH, Weiner J, Wiegand T and DeAngelis DL (2005) Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science* **310**: 987–991.

Guillemaud T, Mieuzet L and Simon JC (2003) Spatial and temporal genetic variability in French populations of the peach-potato aphid, *Myzus persicae*. *Heredity* **91**: 143–152.

Guillot G, Leblois RL, Coulon AL and Frantz AC (2009) Statistical methods in spatial genetics. *Molecular Ecology* **18**: 4734–4756.

Guo Q (2012) Incorporating latitudinal and central–marginal trends in assessing genetic variation across species ranges. *Molecular Ecology* **21**: 5396–5403.

Habel JC, Ulrich W, Assmann T and Masters J (2013) Allele elimination recalculated: nested subset analyses for molecular biogeographical data. *Journal of Biogeography* **40**: 769–777.

Hamilton SL, Regetz J and Warner RR (2008) Postsettlement survival linked to larval life in a marine fish. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 1561–1566.

Hannah CG (2007) Future directions in modelling physical–biological interactions. *Marine Ecology Progress Series* **347**: 301–306.

Hansen MM and Hemmer-Hansen J (2007) Landscape genetics goes to sea. *Journal of Biology* **6**: 6.

Hardie DC and Hutchings JA (2010) Evolutionary ecology at the extremes of species' ranges. *Environmental Reviews* **18**: 1–20.

Hartig F, Calabrese JM, Reineking B, Wiegand T and Huth A (2011) Statistical inference for stochastic simulation models-theory and application. *Ecology Letters* **14**: 816–827.

Hasselman DJ, Ricard D and Bentzen P (2013) Genetic diversity and differentiation in a wide ranging anadromous fish, American shad (*Alosa sapidissima*), is correlated with latitude. *Molecular Ecology* **22**: 1558–1573.

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.

Hedrick PW (1999) Highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**: 313–318.

Heesch S, Broom JES, Neill KF, Farr TJ, Dalen JL and Nelson WA (2009) *Ulva*, *Umbraulva*, and *Gemina*: genetic survey of New Zealand taxa reveals diversity and introduced species. *European Journal of Phycology* **44**: 143–154.

Hellberg ME (2007) Footprints on water: the genetic wake of dispersal among reefs. *Coral Reefs* **26**: 463–473.

Hellberg ME (2009) Gene flow and isolation among populations of marine animals. *Annual Review of Ecology, Evolution and Systematics* **40**: 291–310.

Hellberg ME, Burton RS, Neigel JE and Palumbi SR (2002) Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science* **70**: 273–290.

Hengeveld R and Haeck J (1982) The distribution of abundance. I. Measurements. *Journal of Biogeography*: 303–316.

Hewitt GM (1993). Postglacial distribution and species substructure: lessons from pollen, insects and hybrid zones. *Evolutionary Patterns and Processes*. DR Lees and DL Edwards, Academic Press**:** 97–103.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**: 247–276.

Hey J (1991) The structure of genealogies and the distribution of fixed differences between DNA sequences from natural populations. *Genetics* **128**: 831–840.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution* **27**: 905–920.

Hey J and Machado CA (2003) The study of structured populations - new hope for a difficult and divided science. *Nature Reviews Genetics* **4**: 535–543.

Hickerson MJ and Meyer C (2008) Testing comparative phylogeographic models of marine vicariance and dispersal using a hierarchical Bayesian approach. *BMC Evolutionary Biology* **8**: 322.

Hidas EZ, Russell KG, Ayre DJ and Minchinton TE (2013) Abundance of Tesseropora rosea at the margins of its biogeographic range is closely linked to recruitment, but not fecundity. *Marine Ecology Progress Series* **483**: 199–208.

Hinckley S, Hermann AJ and Megray BA (1996) Development of a spatially explicit, individual-based models of marine fish early life history. *Marine Ecology Progress Series* **139**: 47–68.

Hoban SM, Borkowski DS, Brosi SL, McCleary TS, Thompson LM, McLachlan JS, Pereira MA, Schlarbaum SE and Romero-Severson J (2010) Range-wide distribution of genetic diversity in the North American tree *Juglans cinerea*: a product of range shifts, not ecological marginality or recent population decline. *Molecular Ecology* **19**: 4876–4891.

Hohenlohe PA, Phillips PC and Cresko WA (2010) Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *International Journal of Plant Sciences* **171**: 1059–1071.

Holderegger R and Wagner HH (2008) Landscape Genetics. *BioScience* **59**: 199–207.

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**: 65–70.

Holt RD (2003) On the evolutionary ecology of species' ranges. *Evolutionary Ecology Research* **5**: 159–178.

Holt RD and Gomulkiewicz R (1997) How does immigration influence local adaptation? A reexamination of a familiar paradigm. *The American Naturalist* **149**: 563–572.

Horne JB, van Herwerden L, Choat JH and Robertson DR (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.

Hossaert-McKey M, Valero M, Magda D, Jarry M, Cuguenand J and Verne P (1996) The evolving genetic history of a population of *Lathyrus sylvestris*: evidence from temporal and spatial genetic structure. *Evolution* **50**: 1808–1821.

Huelsken T, Keyse J, Liggins L, Penny S, Treml EA and 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**: e80858.

Hutchison DW and Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*: 1898–1914.

Jensen JL, Bohonak AJ and Kelley ST (2005) Isolation by distance, web service. *BMC Genetics* **6**: 13.

Johannesson K, Panova M, Kemppainen P, André C, Rolán-Alvarez E and Butlin RK (2010) Repeated evolution of reproductive isolation in a marine snail: unveiling mechanisms of speciation. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* **365**: 1735–1747.

Johansson ML, Banks MA, Glunt KD, Hassel‐Finnegan HM and Buonaccorsi VP (2008) Influence of habitat discontinuity, geographical distance, and oceanography on fine-scale population genetic structure of copper rockfish (*Sebastes caurinus*). *Molecular Ecology* **17**: 3051–3061.

Johnson MS and Black R (1982) Chaotic genetic patchiness in an intertidal limpet, *Siphonaria* sp. *Marine Biology* **70**: 157–164.

Jokiel P and Martinelli FJ (1992) The vortex model of coral reef biogeography. *Journal of Biogeography* **19**: 449–458.

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

Jones GP, Milicich MJ, Emslie MJ and Lunow C (1999) Self-recruitment in a coral reef fish population. *Nature* **402**: 802–804.

Jones GP, Planes S and R. TS (2005) Coral reef fish larvae settle close to home. *Current Biology* **15**: 1314–1318.

Joost S, Bonin A, Bruford MW, Després L, Conord C, Erhardt G and Taberlet P (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* **16**: 3955–3969.

Jordan DS and Starks EC (1902) A review of the Hemibranchiate fishes of Japan. *Proceedings of the United States National Museum* **26**: 57–73.

Jorgensen HBH, Hansen MM, Bekkevold D, Ruzzante DE and Loeschcke V (2005) Marine landscapes and population genetic structure of herring (*Clupea harengus* L.) in the Baltic Sea. *Molecular Ecology* **14**: 3219–3234.

Jost L (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* **17**: 4015–4026.

Karl SA, Toonen RJ, Grant WS and Bowen BW (2012) Common misconceptions in molecular ecology: echoes of the modern synthesis. *Molecular Ecology* **21**: 4171–4189.

Kawecki TJ and Holt RD (2002) Evolutionary conserquences of asymmetric dispersal rates. *The American Naturalist* **160**: 333–347.

Kelly RP and Palumbi SR (2010) Genetic structure among 50 species of the Northeastern Pacific rocky intertidal community. *PLoS ONE* **5**: e8594.

Kendall MG and Smith BB (1939) The problem of m rankings. *The Annals of Mathematical Statistics* **10**: 275–287.

Keyse J, Crandall ED, Toonen RJ, Meyer CP, Treml EA and 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**: 47–78.

Kingsford MJ, Leis JM, Shanks A, Lindeman KC, Morgan SG and Pineda J (2002) Sensory environments, larval abilities and local self-recruitment. *Bulletin of Marine Science* **70**: 309–340.

Kininmonth S, van Oppen MJH and Possingham HP (2010) Determining the community structure of the coral *Seriatopora hystrix* from hydrodynamic and genetic networks. *Ecological Modelling* **221**: 2870–2880.

Kinlan BP and Gaines SD (2003) Propagule dispersal in marine and terrestrial environments: a community perspective. *Ecology* **84**: 2007–2020.

Kirkpatrick M and Barton NH (1997) Evolution of a species' range. *The American Naturalist* **150**: 1–23.

Klanten OS, Choat JH and van Herwerden L (2007) Extreme genetic diversity and temporal rather then 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*: 189–216.

Knutsen H, Jorde PE, Sannaes H, Rus Hoelzel A, Bergstad OA, Stefanni S, Johansen T and Stenset NC (2009) Bathymetric barriers promoting genetic structure in the deepwater demersal fish tusk (*Brosme brosme*). *Molecular Ecology* **18**: 3151–3162.

Koehn RK, Bayne BL, Moore MN and Siebenalller JF (1980a) Salinity related physiological and genetic differences between populations of *Mytilus edulis*. *Biological Journal of the Linnean Society* **14**: 319–334.

Koehn RK, Newell RIE and Immermann F (1980b) Maintenance of an aminopeptidase allele frequency cline by natural selection. *Proceedings of the National Academy of Sciences of the United States of America* **77**: 5385–5389.

Koehn RK and Siebenaller JF (1981) Biochemical studies of amino-peptidase polymorphism in *Mytilus edulis* II. Dependence of reaction rate on physical factors and enzyme concentration. *Biochemical Genetics* **19**: 1143–1162.

Kool JT, Moilanen A and Treml EA (2012) Population connectivity: recent advances and new perspectives. *Landscape Ecology* **28**: 165–185.

Kool JT, Paris CB, Andréfouët S and Cowen RK (2010) Complex migration and the development of genetic structure in subdivided populations: an example from Caribbean coral reef ecosystems. *Ecography* **33**: 597–606.

Kool JT, Paris CB, Barber PH and Cowen RK (2011) Connectivity and the development of population genetic structure in Indo-West Pacific coral reef communities. *Global Ecology and Biogeography* **20**: 695–706.

Lacson JM (1992) Minimal genetic variation among samples of six species of coral reef fishes collected at La Parguera, Puerto Rico, and Discovery Bay, Jamaica. *Marine Biology* **112**: 327–331.

Ladd HS (1960) Origin of the Pacific island molluscan fauna. *American Journal of Science* **258**: 137–150.

Laikre L, Allendorf FW, Aroner LC, Baker CS, Gregovich DP, Hansen MM, Jackson JA, Kendall KC, McKelvey K, Neel MC, Olivieri I, Ryman N, Schwartz MK, Bull RS, Stetz JB, Tallmon DA, Taylor BL, Vojta CD, Waller DM and Waples RS (2010) Neglect of genetic diversity in implementation of the Convention on Biological Diversity. *Conservation Biology* **24**: 86–88.

Lammi A, Siikamäki P and Mustajärvi K (1999) Genetic diversity, population size, and fitness in central and peripheral populations of a rare plant *Lychnis viscaria*. *Conservation Biology* **13**: 1069–1078.

Landguth EL, Cushman SA and Johnson NA (2012a) Simulating natural selection in landscape genetics. *Molecular Ecology Resources* **12**: 363–368.

Landguth EL, Fedy BC, Oyler-McCance SJ, Garey AL, Emel SL, Mumma M, Wagner HH, Fortin M-J and Cushman SA (2012b) Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. *Molecular Ecology Resources* **12**: 276–284.

Lavery S, Moritz C and Fielder DR (1996) Indo-Pacific population structure and evolutionary history of the coconut crab *Birgus latro*. *Molecular Ecology* **5**: 557–570.

Lawrence JM and Agatsuma Y (2001). The ecology of *Tripneustes*. In: *Edible sea urchins: biology and ecology*. JM Lawrence (Ed.) Amsterdam, Elsevier**.** pp. 395–413.

Lawton RJ, Messmer V, Pratchett MS and Bay LK (2011) High gene flow across large geographic scales reduces extinction risk for a highly specialised coral feeding butterflyfish. *Molecular Ecology* **20**: 3584–3598.

Lee W-J, Conroy J, Howell WH and Kocher TD (1995) Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution* **41**: 54–66.

Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**: 1659–1673.

Legendre P and Fortin M-J (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 and Lapointe F-J (2004) Assessing congruence among distance matrices: single-malt scotch whiskies revisited. *Australian and New Zealand Journal of Statistics* **46**: 615–629.

Legendre P, Lapointe F-J and Casgrain P (1994) Modeling brain evolution from behaviour: a permutational regression aproach. *Evolution* **48**: 1487–1499.

Leggett WC and Frank KT (1997) A comparative analysis of recruitment variability in North Atlantic flatfishes–testing the species range hypothesis. *Journal of Sea Research* **37**: 281–299.

Leis JM (1991). The pelagic stage of reef fishes: the larval biology of coral reef fishes. In: *The ecology of fishes on coral reefs*. PF Sale (Ed.)**.** pp. 183–230.

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

Leis JM, van Herwerden L and Patterson HM (2011) Estimating connectivity in marine fish populations: what works best? *Oceanography and Marine Biology* **49**: 193–234.

Leis JM, Wright KJ and Johnson RN (2007) Behaviour that influences dispersal and connectivity in the small, young larvae of a reef fish. *Marine Biology* **153**: 103–117.

Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology and Evolution* **17**: 183–189.

Leray M, Beldade R, Holbrook SJ, Schmitt RJ, Planes S and Bernardi G (2010) Allopatric divergence and speciation in coral reef fish: the three-spot Dascyllus, *Dascyllus trimaculatus*, species complex. *Evolution* **64**: 1218–1230.

Lesica P and Allendorf FW (1995) When are peripheral populations valuable for conservation? *Conservation Biology* **9**: 753–760.

Lessios HA, Kane J and Robertson DR (2003) Phylogeography of the pantropical sea urchin *Tripneustes*: contrasting patterns of popualtion structure between oceans. *Evolution* **57**: 2026–2036.

Lessios HA, Kessing BD and Robertson DR (1998) Massive gene flow across the world's most potent marine biogeographic barrier. *Proceedings of the Royal Society B: Biological Sciences* **265**: 583–588.

Lessios HA and Robertson DR (2006) Crossing the impassable: genetic connections in 20 reef fishes across the eastern Pacific barrier. *Proceedings of the Royal Society B: Biological Sciences* **273**: 2201–2208.

Lester SE, Ruttenberg BI, Gaines SD and Kinlan BP (2005) The relationship between dispersal ability and geographic range size. *Ecology Letters* **10**: 1461–0248.

Lett C, Ayata S-D, Huret M and Irisson J-O (2010) Biophysical modelling to investigate the effects of climate change on marine population dispersal and connectivity. *Progress In Oceanography* **87**: 106–113.

Levin SA, Muller-Landau HC, Nathan R and Chave J (2003) The ecology and evolution of seed dispersal: a theoretical perspective. *Annual Review of Ecology, Evolution, and Systematics* **34**: 575–604.

Levinton JS and Haefner PA (2002). *Marine biology: function, biodiversity, ecology*. Oxford, Oxford University Press.

Liang WD, Yang YJ, Tang TY and Chuang WS (2008) Kuroshio in the Luzon Strait. *Journal of Geophysical Research: Oceans (1978-2012)*: 113(C118).

Liggins L, Booth DJ, Figueira WF, Tonk L, Harris DA, Ridgway T, Treml EA and Riginos C. Genetic patterns across the latitudinal range of a coral reef fish reveal historical effects and contrasting patterns of turnover and nestedness at the range peripheries. *in prep.*

Liggins L, Gleeson L and Riginos C (2014) Evaluating edge-of-range genetic patterns for tropical echinoderms, *Acanthaster planci* and *Tripneustes gratilla*, of the Kermadec Islands, southwest Pacific. *Bulletin of Marine Science* **90**: 379–397.

Liggins L, Treml EA, Possingham HP and Riginos C. The roles of historical and contemporary seascape processes in forming spatial genetic patterns of four Indo-Pacific reef fishes. *in prep.*

Liggins L, Treml EA and Riginos C (2013) Taking the plunge: an introduction to undertaking seascape genetic studies and using biophysical models. *Geography Compass* **7**: 173–196.

Liu S-YV, Ho H-CH and Dai C-F (2013) A new species of *Pomacentrus* (Actinopterygii: Pomacentridae) from Micronesia, with comments on its phylogenetic relationships. *Zoological Studies* **52**: 1–8.

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

Liu S-YV, Wang C-H, Shiao J-C and Dai C-F (2010) Population connectivity of neon damsel, *Pomacentrus coelestis*, inferred from otolith microchemistry and mtDNA. *Marine and Freshwater Research* **61**: 1416–1424.

Liu S-YV, Dai CF, Allen GR and Erdmann MV (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.

Lowe WH and Allendorf FW (2010) What can genetics tell us about population connectivity? *Molecular Ecology* **19**: 3038–3051.

Lowry DB and Willis JH (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology* **8**: doi:10.1371/journal.pbio.1000500.

Luikart G, England PR, Tallmon D, Jordan S and Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* **4**: 981–994.

Lukoschek V, Waycott M and Marsh H (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.

Maggs CA, Castilho R, Foltz D, Henzler CM, Jolly MT, Kelly J, Olsen J, Perez KE, Stam W, Väinölä R, Viard F and Wares J (2008) Evaluating signitures of glacial refugia for North Atlantic benthic marine taxa. *Ecology* **89**: S108–S122.

Manel S, Joost S, Epperson BK, Holderegger R, Storfer A, Rosenberg MS, Scribner KT, Bonin A and Fortin M-J (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology* **19**: 3760–3772.

Manel S, Schwartz MK, Luikart G and Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* **18**: 189–197.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**: 209–220.

Margvelashvili N, Saint-Cast F and Condie S (2008) Numerical modelling of the suspended sediment transport in Torres Strait. *Continental Shelf Research* **28**: 2241–2256.

Marko PB and Hart MW (2011) The complex analytical landscape of gene flow inference. *Trends in Ecology and Evolution* **26**: 448–456.

Marshall DJ and Morgan SG (2011) Ecological and evolutionary consequences of linked life-history stages in the sea. *Current Biology* **21**: R718–R725.

Marshall DJ, Munro K, Bode M, Keough MJ and Swearer S (2010) Phenotype-environment mismatches reduce connectivity in the sea. *Ecology Letters* **13**: 128–140.

Matschiner M, Hanel R and Salzburger W (2009) Gene flow by larval dispersal in the Antarctic notothenioid fish *Gobionotothen gibberifrons*. *Molecular Ecology* **18**: 2574–2587.

McCormick MI, Makey L and Dufour V (1997) Comparative study of metamorphosis in tropical reef fishes. *Marine Biology* **141**: 841–853.

McKnight DG (1968) Some echinoderms from the Kermadec Islands. *New Zealand Journal of Marine and Freshwater Research* **2**: 505–526.

McKnight DG (1978) *Acanthaster planci* (Linneaus)(Asteroidea: Echinodermata) at the Kermadec Islands. *New Zealand Oceanographic Institute* **4**: 17–20.

McMillan WO and Palumbi SR (1995) Concordant evolutionary patterns among Indo-West Pacific butterflyfishes. *Proceedings of the Royal Society B: Biological Sciences* **260**: 229–236.

McRae BH (2006) Isolation by resistance. *Evolution* **60**: 1551–1561.

McRae BH (2008) Using circuit theory to model connectivity in ecology, evolution and conservation. *Ecology* **89**: 2712–2724.

Meirmans P and van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Heredity* **93**: 43–50.

Meirmans PG and Hedrick PW (2011) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources* **11**: 5–18.

Mendez M, Rosenbaum HC, Subramaniam A, Yackulic C and Bordino P (2010) Isolation by environmental distance in mobile marine species: molecular ecology of franciscana dolphins at their southern range. *Molecular Ecology* **19**: 2212–2228.

Metaxas A and Saunders M (2009) Quantifying the "bio-" components in biophysical models of larval transport in marine benthic invertebrates: advances and pitfalls. *Biological Bulletin* **216**: 257–272.

Meyers LS, Gamst G and Guarino AJ (2006). *Applied Multivariate Research: Design and Interpretation*. Sage, CA, Thousand Oaks.

Miller MJ, Bermingham E, Klicka J, Escalante P and Winker K (2010) Neotropical birds show a humped distribution of within population genetic diversity along a latitudinal transect. *Ecology Letters* **13**: 576–586.

Mirams GK, Treml EA, Shields JL, Liggins L and Riginos C (2011) Vicariance and dispersal across an intermittent barrier: popualtion genetic structure of marine animals across the Torres Strait land bridge. *Coral Reefs* **30**: 937–949.

Moen T, Hayes B, Nilsen F, Delghandi M, Fjalestad KT, Fevolden SE, Berg RK and Lien S (2008) Identification and characterisation of novel SNP markers in Atlantic cod: evidence for directional selection. *BMC Genetics* **26**: 9–18.

Mokhtar-Jamai K, Pascual M, Ledoux JB, Coma R, Feral JP, Garrabou J and Aurelle D (2010) From global to local genetic structuring in the red gorgonian *Paramuricea clavata*: the interplay between oceanographic conditions and limited larval dispersal. *Molecular Ecology* **20**: 3291–3305.

Mora C, Chittaro PM, Sale PF, Kritzer JP and Ludsin SA (2003) Patterns and processes in reef fish diversity. *Nature* **421**: 933–936.

Mora C and Sale PF (2002) Are populations of coral reef fish open or closed? *Trends in Ecology and Evolution* **17**: 422–428

Mora C, Treml EA, Roberts J, Crosby K, Roy D and Tittensor DP (2011) High connectivity among habitats precludes the relationship between dispersal and range size in tropical reef fishes. *Ecography* **35**: 89–96.

Moritz C and Faith DP (1998) Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology* **7**: 419–429.

Mortensen T (1937) Contributions to the study of the development and larval forms of echinoderms III. *D K Dan Vidensk Selsk Skr Nat Math Afd* **9**: 1–65.

Nakajima Y, Nishikawa A, Iguchi A and Sakai K (2012) Regional genetic differentiation among northern high-latitude island populations of a broadcast-spawning coral. *Coral Reefs* **31**: 1125–1133.

Nei M (1972) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* **70**: 3321–3323.

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

Nei M and Kumar S (2000). *Molecular evolution and phylogenetics*. New York, Oxford University Press, Inc.

Ni G, Li QI, Kong L and Yu H (2014) Comparative phylogeography in marginal seas of the northwestern Pacific. *Molecular Ecology* **23**: 534–548.

Nielsen EE, Hemmer-Hansen J, Larsen PF and Bekkevold D (2009a) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology* **18**: 3128–3150.

Nielsen EE, Hemmer-Hansen J, Poulsen NA, Loeschcke V, Moen T, Johansen T, Mittelholzer C, Taranger G-L, Ogden R and Carvalho GR (2009b) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology* **9**: 276.

North EW, Gallego A and Petitgas P, (Eds.) (2009). *Manual of recommended practices for modelling physical–biological interactions during fish early life.* Cooperative Research Report. Denmark, International Council for Exploration of the Sea.

Nosil P, Vines TH and J. FD (2005) Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59**: 705–719.

O'Connor MI, Bruno JF, Gaines SD, Halpern BS, Lester SE, Kinlan BP and Weiss JM (2007) Temperature control of larval dispersal and the implications for marine ecology, evolution and conservation. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 1266–1271.

Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ and Suggests MASS (2007) The vegan package. Community ecology package.

Oreskes N, Shrader-Frechette K and Belitz K (1994) Verification, validation, and confirmation of numerical models in the earth sciences. *Science* **263**.

Palma-Silva C, Lexer C, Paggi GM, Barbará T, Bered F and Bodanese-Zanettini MH (2009) Range-wide patterns of nuclear and chloroplast DNA diversity in *Vriesea gigantea* (Bromeliaceae), a neotropical forest species. *Heredity* **103**: 503–512.

Palumbi SR (1992) Marine speciation on a small planet. *Trends in Ecology and Evolution* **7**: 114–118.

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics* **25**: 547–572.

Palumbi SR (2004) Marine reserves and ocean neighborhoods: the spatial scale of marine populations and their management. *Annual Review of Environment and Resources* **29**: 31–68.

Palumbi SR, Grabowsky G, Duda T, Geyer L and Tachino N (1997) Speciation and population genetic structure in tropical Pacific sea urchins. *Evolution* **51**: 1506–1517.

Pampoulie C, Ruzzante DE, Chosson V, Jörundsdóttir TD, Taylor L, Thorsteinsson V, Daníelsdóttir AK and Marteinsdóttir G (2006) The genetic structure of Atlantic cod (*Gadus morhua*) around Iceland: insight from microsatellites, the *Pan I* locus, and tagging experiments. *Canadian Journal of Fisheries and Aquatic Sciences* **63**: 2660–2674.

Pande E and Simon-Bouhet B (2013) marmap: a package for importing, plotting and analyzing bathymetric and topographic data in R. *PLoS ONE* **8**: e73051.

Papetti C, Pujolar JM, Mezzavilla M, La Mesa M, Rock J, Zane L and Patarnello T (2012) Population genetic structure and gene flow patterns between populations of the Antarctic icefish *Chionodraco rastrospinosus*. *Journal of Biogeography* **39**: 1361–1372.

Paquette SR and Lapointe F-J (2009) A statistical procedure to assess the significance level of barriers to gene flow. *Journal of Genetics and Genomics* **36**: 685–693.

Paradis E, Claude J and Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**: 289–290.

Paris CB and Chérubin LM (2007) Surfing, spinning, or diving from reef to reef: effects on population connectivity. *Marine Ecology Progress Series* **347**: 285–300.

Paris CB and Cowen RK (2004) Direct evidence of a biophysical retention mechanism for coral reef fish larvae. *Limnology and Oceanography* **49**: 1964–1979.

Pelc RA, Warner RR and Gaines SD (2009) Geographical patterns of genetic structure in marine species with contrasting life histories. *Journal of Biogeography* **36**: 1881–1890.

Petit R, J., Aguinagalde I, de Beaulieu J, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S and Vendramin GG (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**: 1563–1565.

Petit R, J., El Mousadik A and Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* **12**: 844–855.

Pfeifer M, Schatz B, Xavier Pico F, Passalacqua NG, Fay MF, Carey PD and Jeltsch F (2009) Phylogeography and genetic structure of the orchid *Himantoglossum hircinum* (L.) Spreng. across its European central–marginal gradient. *Journal of Biogeography* **36**: 2353–2365.

Pineda J, Hare JA and Sponaugle S (2007) Larval transport and dispersal in the coastal ocean and consequences for population connectivity. *Oceanography* **20**: 22–40.

Pinheiro F, de Barros F, Palma-Silva C, Fay MF, Lexer C and Cozzolino S (2011) Phylogeography and genetic differentiation along the distributional range of the orchid *Epidendrum fulgens*: a Neotropical coastal species not restricted to glacial refugia. *Journal of Biogeography* **38**: 1923–1935.

Planes S (1993) Genetic differentiation in relation to restricted larval dispersal of the convict surgeonfish *Acanthurus triostegus* in French Polynesia. *Marine Ecology Progress Series* **98**: 237–246.

Planes S, Galzin R and Bonhomme F (1996) A genetic metapopulation model for reef fishes in oceanic islands: the case of the surgeonfish, *Acanthurus triostegus*. *Journal of Evolutionary Biology* **9**: 103–117.

Planes S, Jones GP and Thorrold SR (2009) Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 5693–5697.

Planes S and Lenfant P (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.

Planes S, Parroni M and Chauvet C (1998) Evidence of limited gene flow in three species of coral reef fishes in the lagoon of New Caledonia. *Marine Biology* **130**: 361–368.

Pogson GH (2001) Nucleotide polymorphism and natural selection at the pantophysin (*Pan I*) locus in the Atlantic cod, *Gadus morhua* (L.). *Genetics* **157**: 317–330.

Pogson GH and Mesa KA (2004) Positive Darwinian selection at the *Pantophysin* (*Pan I*) locus in marine gadid fishes. *Molecular Biology and Evolution* **21**: 65–75.

Polzin T and Daneshmand SV (2003) On Steiner trees and minimum spanning trees in hypergraphs. *Operations Research Letters* **31**: 12–20.

Possingham HP and Roughgarden J (1990) Spatial population dynamics of a marine organism with a complex life cycle. *Ecology* **71**: 973–985.

Powell AWB (1976) On the considerable influx of warm water molluscs that have invaded northern New Zealand waters within recent years. *Records of the Auckland Institute and Museum* **13**: 141–166.

Prada C, Schizas NV and Yoshioka PM (2008) Phenotypic plasticity or speciation? A case from a clonal marine organism. *BMC Evolutionary Biology* **8**: 47.

Pritchard JK (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Pujol B and Pannell JR (2008) Reduced responses to selection after species range expansion. *Science* **321**: 96.

Puritz JB and Toonen RJ (2011) Coastal pollution limits pelagic larval dispersal. *Nature Communications* **2**: 226.

R Core Team, Ed. (2012). R: a language and environment for statistical computing. R Foundation for Statistical Computing. *Available at* http://wwwR-projectorg/

Rambaut AE (1996) Se-Al: sequence alignment editior v2.0a11. *Available at* http://evolvezoooxacuk/.

Randall JE (1956) A revision of the surgeonfish genus *Acanthurus*. *Pacific Science* **10**: 159–235.

Randall JE (1998) Zoogeography of shore fishes of the Indo-Pacific region. *Zoological Studies* **37**: 227–268.

Reeves JM, Chivas AR and Garcia A (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.

Reid DG, Lal K, Mackenzie-Dodds J, Kaligis F, Littlewood DTJ and Williams ST (2006) Comparative phylogeography and species boundaries in Echinolittorina snails in the central Indo-West Pacific. *Journal of Biogeography* **33**: 990–1006.

Reinig WF (1938). *Elimination und Selektion*. Jena, Gustav Fischer.

Richards ZT and Liggins L. Hermatypic corals and crown-of-thorns starfish of the Kermadec Islands. *Kermadec Biodiscovery Expedition 2011*. T Trnski and H Schlumpf. Auckland, Bulletin of the Auckland Museum **20:** *in press*

Riginos C, Buckley YM, Blomberg SP and Treml EA (2014) Dispersal capacity predicts both population genetic structure and species richness in reef fishes. *The American Naturalist* **184**: 52–64.

Riginos C, Douglas KE, Jin Y, Shanahan DF and Treml EA (2011) Effects of geography and life history traits on genetic differentiation in benthic marine fishes. *Ecography* **34**: 566–575.

Riginos C and Liggins L (2013) Seascape genetics: populations, individuals, and genes marooned and adrift. *Geography Compass* **7**: 197–216.

Riginos C and Nachman MW (2001) Population subdivision in marine environments: the contributions of biogeography, geographical distance and discontinuous habitat to genetic differentiation in a blennioid fish, *Axoclinus nigricaudus*. *Molecular Ecology* **10**: 1429–1453.

Roberts CM, McClean CJ, Veron JE, Hawkins JP, Allen GR, McAllister DE, Mittermeier CG, Schueler FW, Spalding M, Wells F, Vynne C and Werner TB (2002) Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* **295**: 1280–1284.

Roberts JJ, Best BD, Dunn DC, Treml EA and Halpin PN (2010) Marine Geospatial Ecology Tools: an integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environmental Modelling and Software* **25**: 1197–1207.

Rocha LA, Bass AL, Robertson DR and Bowen BW (2002) Adult habitat preferences, larval dispersal, and the comparative phylogeography of three Atlantic surgeonfishes (Teleostei: Acanthuridae). *Molecular Ecology* **11**: 243–252.

Rocha LA and Bowen BW (2008) Speciation in coral-reef fishes. *Journal of Fish Biology* **72**: 1101-1121.

Rocha LA, Craig MT and Bowen BW (2007) Phylogeography and the conservation of coral reef fishes. *Coral Reefs* **26**: 501–512.

Rogerson P (2001). *Statistical methods for geography*. London, Sage.

Roman J and Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in Ecology and Evolution* **22**: 454–464.

Roughgarden J, Gaines S and Possingham H (1988) Recruitment dynamics in complex life cycles. *Science* **241**: 1460–1466.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**: 58–62.

Saenz-Agudelo P, Jones GP, Thorrold SR and Planes S (2009) Estimating connectivity in marine populations: an empirical evaluation of assignment tests and parentage analysis under different gene flow scenarios. *Molecular Ecology* **18**: 1765–1776.

Saenz-Agudelo S, Jones GP, Thorrold SR and Planes S (2011) Connectivity dominates larval replenishment in a coastal reef fish metapopulation. *Proceedings of the Royal Society B: Biological Sciences* **278**: 2954–2961.

Sagarin RD and Gaines SD (2002) The 'abundant centre'distribution: to what extent is it a biogeographical rule? *Ecology Letters* **5**: 137–147.

Sbrocco EJ (2014) Paleo-MARSPEC: gridded ocean climate layers for the mid-Holocene and Last Glacial Maximum. *Ecology* **95**: 1710.

Sbrocco EJ and Barber P (2013) MARSPEC: ocean climate layers for marine spatial ecology. *Ecology* **94**: 979.

Schiel DR, Kingsford MJ and Choat JH (1986) Depth distribution and abundance of benthic organisms and fishes at the subtropical Kermadec Islands. *New Zealand Journal of Marine and Freshwater Research* **20**: 521–535.

Schmidt PS and Rand DM (2001) Adaptive maintenance of genetic polymorphism in an intertidal barnacle: habitat and life stage specific survivorship of *MPI* genotypes. *Evolution* **55**: 1336–1344.

Schmidt PS, Serrão EA, Pearson GA, Riginos C, Rawson PD, Hilbish TJ, Brawley SH, Trussell GC, Carrington E, Wethey DS, Grahame JW, Bonhomme F and Rand DM (2008) Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. *Ecology* **89**: S91–S107.

Schoville SD, Bonin A, Francois O, Lobreaux S, Melodelima C and Manel S (2012) Adaptive genetic variation on the landscape: methods and cases. *Annual Review of Ecology, Evolution and Systematics* **43**: 23–43.

Schwartz MK and McKelvey KS (2008) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics* **10**: 441–452.

Scribner KT, Smith MH, Garrott RA and Carpenter LH (1991) Temporal, spatial, and age-specific changes in genotypic composition of mule deer. *Journal of Mammalogy* **72**: 126–137.

Selkoe KA, Gaggiotti OE, Bowen BW and Toonen RJ (2014) Emergent patterns of population genetic structure for a coral reef community. *Molecular Ecology* **23**: 3064–3079.

Selkoe KA, Gaines SD, Caselle JE and Warner RR (2006) Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology* **87**: 3082–3094.

Selkoe KA, Henzler CM and Gaines SD (2008) Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries* **9**: 363–377.

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

Selkoe KA, Watson JR, White C, Horin TB, Iacchei M, Mitarai S, Siegel DA, Gaines SD and 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. *Molecular Ecology* **19**: 3708–3726.

Sexton JP, McIntyre PJ, Angert AL and Rice KJ (2009) Evolution and ecology of species range limits. *Annual Review of Ecology, Evolution and Systematics* **40**: 415–436.

Shanks A (2009) Pelagic larval duration and dispersal distance revisited. *Biological Bulletin* **216**: 373–385.

Sherman CDH, Hunt A and Ayre DJ (2008) Is life history a barrier to dispersal? Contrasting patterns of genetic differentiation along an oceanographically complex coast. *Biological Journal of the Linnean Society* **95**: 106–116.

Shima JS and Swearer SE (2010) The legacy of dispersal: larval experience shapes persistence later in the life of a reef fish. *The Journal of Animal Ecology* **79**: 1308–1314.

Shulman MJ and Bermingham E (1995) Early life histories, ocean currents, and the population genetics of Caribbean reef fishes. *Evolution* **49**: 897–910.

Siegel DA, Kinlan BP, Gaylord B and Gaines SD (2003) Langrarian descriptions of marine larval dispersion. *Marine Ecology Progress Series* **260**: 83–96.

Simpson GG (1943) Mammals and the nature of continents. *American Journal of Science* **241**: 1–31.

Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics* **16**: 393–430.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**: 264–279.

Slatkin M (2004) Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Molecular Ecology* **14**: 67–73.

Smouse P (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology* **35**: 627–632.

Sokal RR and Rohlf FJ (1981). *Biometry*. New York, W.H. Freeman.

Sørensen T (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application in analysis of the vegetation on Danish commons. *Biologiske Skifter Det Konglige Danske Videnskabernes Selskab* **5**: 1–34.

Spalding M, Ravilious C and Green EP (2001). *World atlas of coral reefs*. Berkeley, University of California Press.

Spalding MD, Fox HE, Allen GR, Davidson N, FerdaÑA ZA, Finlayson MAX, Halpern BS, Jorge MA, Lombana AL, Lourie SA, Martin KD, McManus E, Molnar J, Recchia CA and Robertson J (2007) Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *BioScience* **57**: 573.

Spear SF, Balkenhol N, Fortin M-J, McRae BH and Scribner KIM (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology* **19**: 3576–3591.

Spencer HG, Waters JM and Eichhorst TE (2007) Taxonomy and nomenclature of black nerites (Gastropoda:Neritimorpha:*Nerita*) from the South Pacific. *Invertebrate Systematics* **21**: 229–237.

Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L and Waits LP (2007) Putting the 'landscape' in landscape genetics. *Heredity* **98**: 128–142.

Storfer A, Murphy MA, Spear SF, Holderegger R and Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology* **19**: 3496–3514.

Strathmann RR (1990) Why life histories evolve differently in the sea. *American Zoologist* **30**: 197–207.

Suthers IM, Young JW, Baird ME, Roughan M, Everett JD, Brassington GB, Byrne M, Condie SA, Hartog JR, Hassler CS, Hobday AJ, Holbrook NJ, Malcolm HA, Oke PR, Thompson PA and Ridgway K (2011) The strengthening East Australian Current, its eddies and biological effects - an introduction and overview. *Deep Sea Research Part II: Topical Studies in Oceanography* **58**: 538–546.

Sutton P, Chiswell S, Gorman R, Kennan S and Rickard G (2012). Physical marine environment of the Kermadec Islands region. *Science for Conservation*. Wellington, Department of Conservation**:** 1–17.

Swearer S, Caselle JE, Lea DW and Warner RR (1999) Larval retention and recruitment in an island population of a coral-reef fish. *Nature* **402**: 799–802.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **125**: 585–593.

Tajima F (1996) The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. *Genetics* **143**: 1457–1465.

Taylor MS and Hellberg ME (2003) Genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. *Science* **299**: 107–109.

Thiel-Egenter C, Holderegger R, Brodbeck S and Gugerli F (2009) Concordant genetic breaks, identified by combining clustering and tessellation methods, in two co-distributed alpine plant species. *Molecular Ecology* **18**: 4495–4507.

Thomassen HA, Cheviron ZA, Freedman AH, Harrigan RJ, Wayne RK and Smith TB (2010) Spatial modelling and landscape-level approaches for visualizing intra-specific variation. *Molecular Ecology* **19**: 3532–3548.

Thorrold SR, Zacherl DC and Levin LA (2007) Population connectivity and larval dispersal–using geochemical signatures in calcified structures. *Oceanography* **20**: 80–89.

Thorson G (1971). *Life in the Sea*. London, Weidenfeld and Nicolson.

Thresher RE, Colin PL and Bell LJ (1989) Planktonic duration, distribution and population structure of Western and Central Pacific damselfishes. *Copeia* **2**: 420–434.

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

Timmers MA, Bird CE, Skillings DJ, Smouse PE and 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**: e31159.

Tittensor DP, Mora C, Jetz W, Lotze HK, Ricard D, Berghe EV and Worm B (2010) Global patterns and predictors of marine biodiversity across taxa. *Nature* **466**: 1098–1101.

Toonen RJ, Andrews KR, Baums IB, Bird CE, Concepcion GT, Daly-Engel TS, Eble JA, Faucci A, Gaither MR, Iacchei M, Puritz JB, Schultz JK, Skillings DJ, Timmers MA and Bowen BW (2011) Defining boundaries for ecosystem-based management: a multispecies case study of marine connectivity across the Hawaiian archipelago. *Journal of Marine Biology* **2011**: 1–13.

Toonen RJ and Grosberg RK (2011). Causes of chaos: spatial and temporal genetic heterogeneity in the intertidal anomuran crab *Petrolisthes cinctipes*. *Phylogeography and Population Genetics in Crustacea*. S Koenemann, C Held and C Schubart. Boca Raton, CRC Press. Crustacean Issues Series: 75–107.

Treml EA and Halpin PN (2012) Marine population connectivity identifies ecological neighbors for conservation planning in the Coral Triangle. *Conservation Letters* **5**: 441-449.

Treml EA, Halpin PN, Urban DL and Pratson LF (2008) Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. *Landscape Ecology* **23**: 19–36.

Treml EA, Roberts JJ, Chao Y, Halpin PN, Possingham HP and Riginos C (2012) Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. *Integrative and Comparative Biology* **52**: 525–537.

Tuya F, Wernberg T and Thomsen MS (2008) Testing the 'abundant centre' hypothesis on endemic reef fishes in south-western Australia. *Marine Ecology Progress Series* **372**: 225–230.

Ulrich W, Almeida-Neto M and Gotelli NJ (2009) A consumer's guide to nestedness analysis. *Oikos* **118**: 3–17.

Underwood JN, Smith LD, van Oppen MJH and Gilmour JP (2009) Ecologically relevant dispersal of corals on isolated reefs: implications for managing resilience. *Ecological Applications* **19**: 18–29.

Urban DL, Minor ES, Treml EA and Schick RS (2009) Graph models of habitat mosaics. *Ecology Letters* **12**: 260–273.

Uthicke S, Schaffelke B and Byrne M (2009) A boom-bust phylum? Ecological and evolutionary consequences of density variations in echinoderms. *Ecological Monographs* **79**: 3–24.

Vandergast AG, Perry WM, Lugo RV and Hathaway SA (2011) Genetic landscapes GIS Toolbox: tools to map patterns of genetic divergence and diversity. *Molecular Ecology Resources* **11**: 158–161.

Vellend M and Geber MA (2005) Connections between species diversity and genetic diversity. *Ecology Letters* **8**: 767–781.

Veron JEN, DeVantier LM, Turak E, Green AL, Kininmonth S, Stafford-Smith M and Peterson N (2009) Delineating the Coral Triangle. *Galaxea* **11**: 91–100.

Vigliola L, Doherty PJ, Meekan MG, Drown DM, Jones ME and Barber PH (2007) Genetic identity determines risk of post-settlement mortality of a marine fish. *Ecology* **88**: 1263–1277.

Vogler C, Benzie J, Lessios H, Barber P and Worheide G (2008) A threat to coral reefs multiplied? Four species of crown-of-thorns starfish. *Biology Letters* **4**: 696–699.

Vogler C, Benzie JAH, Tenggardjaja K, Ambariyanto, Barber PH and Wörheide G (2013) Phylogeography of the crown-of-thorns starfish: genetic structure within the Pacific species. *Coral Reefs* **32**: 515–525.

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

Vuilleumier S and Possingham HP (2006) Does colonization asymmetry matter in metapopulations? *Proceedings of the Royal Society B: Biological Sciences* **273**: 1637–1642.

Wakeley J (1999) Nonequilibrium migration in human history. *Genetics* **153**: 1863–1871.

Wang P, Song YT, Chao Y and Zhang HC (2005) Parallel computation of the regional ocean modeling system. *The International Journal of High Performance Computing Applications* **19**: 375–385.

Waples RS (1987) A multispecies approach to the analysis of gene low in marine shore fishes. *Evolution* **41**: 385–400.

Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Heredity* **89**: 438–450.

Ward RD, Woodward M and Skibinski DOF (1995) A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of Fish Biology* **44**: 213–232.

Wares JP, Gaines SD and Cunningham CW (2001) A comparative study of asymmetric migration events across a marine biogeographic boundary. *Evolution* **55**: 295–306.

Waters JM, Fraser CI and Hewitt GM (2013) Founder takes all: density-dependent processes structure biodiversity. *Trends in Ecology and Evolution* **28**: 78–85.

Waters JM, King TM, O'Loughlin PM and Spencer HG (2005) Phylogeographical disjunction in abundant high-dispersal littoral gastropods. *Molecular Ecology* **14**: 2789–2802.

Waters JM, McCulloch GA and Eason JA (2007) Marine biogographical structure in two highly dispersive gastropods: implications for trans-Tasman dispersal. *Journal of Biogeography* **34**: 678–687.

Watt JC (1975) A general introduction to terrestrial Arthropoda of the Kermadec Islands. *New Zealand Entomologist* **6**: 32–45.

Weersing K and Toonen RJ (2009) Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series* **393**: 1–12.

White C, Selkoe KA, Watson J, Siegel DA, Zacherl DC and Toonen RJ (2010) Ocean currents help explain population genetic structure. *Proceedings of the Royal Society B: Biological Sciences* **277**: 1685–1694.

Whitlock MC and McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. *Heredity* **82**: 117–125.

Whittaker RH (1960) Vegetation of the Siskiyou mountains, Oregon and California. *Ecological Monographs* **30**: 279–338.

Wicks LC, Gardner JPA and Davy SK (2010a) Spatial patterns and regional affinities of coral communities at the Kermadec Islands Marine Reserve, New Zealand—a marginal high-latitude site. *Marine Ecology Progress Series* **400**: 101–113.

Wicks LC, Sampayo E, Gardner JPA and Davy SK (2010b) Local endemicity and high diversity characterise high-latitude coral–Symbiodinium partnerships. *Coral Reefs* **29**: 989–1003.

Wilkinson-Herbots HM and Ettridge R (2004) The effect of unequal migration rates on F_{ST}. *Theoretical Population Biology* **66**: 185–197.

Williams ST and Duda TF (2008) Did tectonic activity stimulate Oligo-Miocene speciation in the Indo-West Pacific? *Evolution* **62**: 1618–1634.

Wolanski E, Ridd P and Inoue M (1988) Currents through the Torres Strait. *Journal of Physical Oceanography* **18**: 1535–1545.

Wood AR and Gardner JPA (2007) Small spatial scale population genetic structure in two limpet species endemic to the Kermadec Islands, New Zealand. *Marine Ecology Progress Series* **349**: 159–170.

Wright DH and Reeves JH (1992) On the meaning and measurement of nestedness of species assemblages. *Oecologia* **92**: 416–428.

Wright S (1931) Evolution in mendelian populations. *Genetics* **16**: 97–159.

Wright S (1943) Isolation by distance. *Genetics* **28**: 114–138.

Yamaguchi M (1973). Early life-histories of coral reef asteroids, with special reference to *Acanthaster planci* (L.). In: *Biology and geology of coral reefs*. OA Jones and R Endean (Eds.) New York, Academic Press. pp 369–387.

Yasuda N, Nagai S, Hamaguchi M, Okaji KEN, Gerard K and Nadaoka K (2009) Gene flow of *Acanthaster planci*(L.) in relation to ocean currents revealed by microsatellite analysis. *Molecular Ecology* **18**: 1574–1590.

APPENDIX ONE. Vicariance and dispersal across an intermittent barrier: population genetic structure of marine animals across the Torres Strait land bridge

Published in Coral Reefs 30: 937–949. 2011

Anna G. K. Mirams, Eric A. Treml, Jody L. Shields, **Libby Liggins** and Cynthia Riginos

School of Biological Sciences, The University of Queensland

Abstract

Biogeographic barriers, some transitory in duration, are likely to have been important contributing factors to modern marine biodiversity in the Indo-Pacific region. One such barrier was the Torres Strait land bridge between continental Australia and New Guinea that persisted through much of the late Pleistocene and separated Indian and Pacific Ocean taxa. Here, we examine the pat- terns of mitochondrial DNA diversity for marine animals with present-day distributions spanning the Torres Strait. Specifically, we investigate whether there are concordant signatures across species, consistent with either vicariance or recent colonization from either ocean basin. We survey four species of reef fishes (*Apogon doederleini*, *Pomacentrus coelestis*, *Dascyllus trimaculatus*, and *Acanthurus triostegus*) for mtDNA cytochrome oxidase 1 and control region variation and contrast these results to previous mtDNA studies in diverse marine animals with similar distributions. We find substantial genetic partitioning (estimated from F-statistics and coalescent approaches) between Indian and Pacific Ocean populations for many species, consistent with regional persistence through the late Pleistocene in both ocean basins. The species-specific estimates of genetic divergence, however, vary greatly and for reef fishes we estimate substantially different divergence times among species. It is likely that Indian and Pacific Ocean populations have been isolated for multiple glacial cycles for some species, whereas for other species genetic connections have been more recent. Regional estimates of genetic diversity and directionality of gene flow also vary among species. Thus, there is no apparent consistency among historical patterns across the Torres Strait for these co-distributed marine animals.

Introduction

Absolute vicariance in the marine realm is a rare event (Palumbi 1994), as it depends on major geological restructuring of the landscape such as the movement of continental plates. Indisputably, such vicariant events are important generators of marine biodiversity, for example, geminate species on either side of the Isthmus of Panama have long been recognized (Jordan 1908; Mayr 1954). Some

barriers have been more transient, arising from sea level changes, sometimes in combination with locally dynamic geological processes. The imprint of these transient barriers (i.e., an absolute barrier in the past, but without complete impediment to dispersal in the present) can often be detected in the genetic patterns of co-distributed species. For example, Avise and coworkers documented the striking phylogeographic concordance of marine species distributed along the Florida coastline, associated with past sea level changes, climate fluctuations, and contemporary discontinuities in water masses (Avise 2000 and references therein). Similar genetic concordance has been observed across former land barriers, such as between Atlantic and Mediterranean marine species (Patarnello et al. 2007), and species on either side of the mainland Australia–Tasmania land bridge (Waters et al. 2005, 2007).

It is self-evident that landmasses can create barriers to gene exchange among marine animals. Once the barriers dissipate (due to sea level rise, for example), however, dispersal across the former barrier should lead to gene exchange and degradation of any genetic signal of past vicariance. Why the signatures of vicariance remain for many taxa is puzzling. It may be that marine animals disperse far less then generally assumed, and therefore, historical patterns persist for a long time once genetic differentiation is established. Another possibility is that boundaries between biogeographic regions are anomalous in some way; either dispersal is specifically reduced in such areas (for example, due to currents and water flow) or selection against cross-boundary migrants prevents them from reaching, settling, or reproducing in new destinations (Hellberg 2007). Whether the inherent dispersal and colonization ability of a species affects present-day gene flow across historical barriers is unresolved.

The reefs of tropical Australia together with reefs of the Indo-Malay Archipelago (the ''Coral Triangle'') contain the greatest concentration of coastal marine biodiversity on earth (Tittensor et al. 2010), with distinct biotas inhabiting the Indian and Pacific Oceans. The exact location and nature of many historic barriers between the Indian and Pacific Oceans are uncertain due to the extremely dynamic geology in this region (Voris 2000). One clear and well-dated barrier is the Torres Strait land bridge (Voris 2000, see Fig. 1). This land bridge connected northern Australia and New Guinea intermittently throughout the late Pleistocene until ~7,000 years ago, after sea levels rose to present levels (Voris 2000; Reeves et al. 2008). Thus, the Torres Strait land bridge offers an excellent setting for exploring how genetic differentiation has been shaped by both recent dispersal and historical vicariance (Chenoweth et al. 1998). There has been a spate of recent studies examining genetic patterns of marine animals distributed across the Torres Strait (Elliott 1996; Begg et al. 1998; Chenoweth et al. 1998; Gopurenko et al. 1999; Lessios et al. 2001; Williams et al. 2002; Chenoweth and Hughes 2003; Uthicke and Benzie 2003; Bay et al. 2004; Dethmers et al. 2006; Duncan et al. 2006; Reid et al. 2006; Imron et al. 2007; Klanten et al. 2007; Lukoschek et al. 2007; Dudgeon et al. 2009; van Herwerden et al. 2009); but there has been no explicit comparison among different species to date.

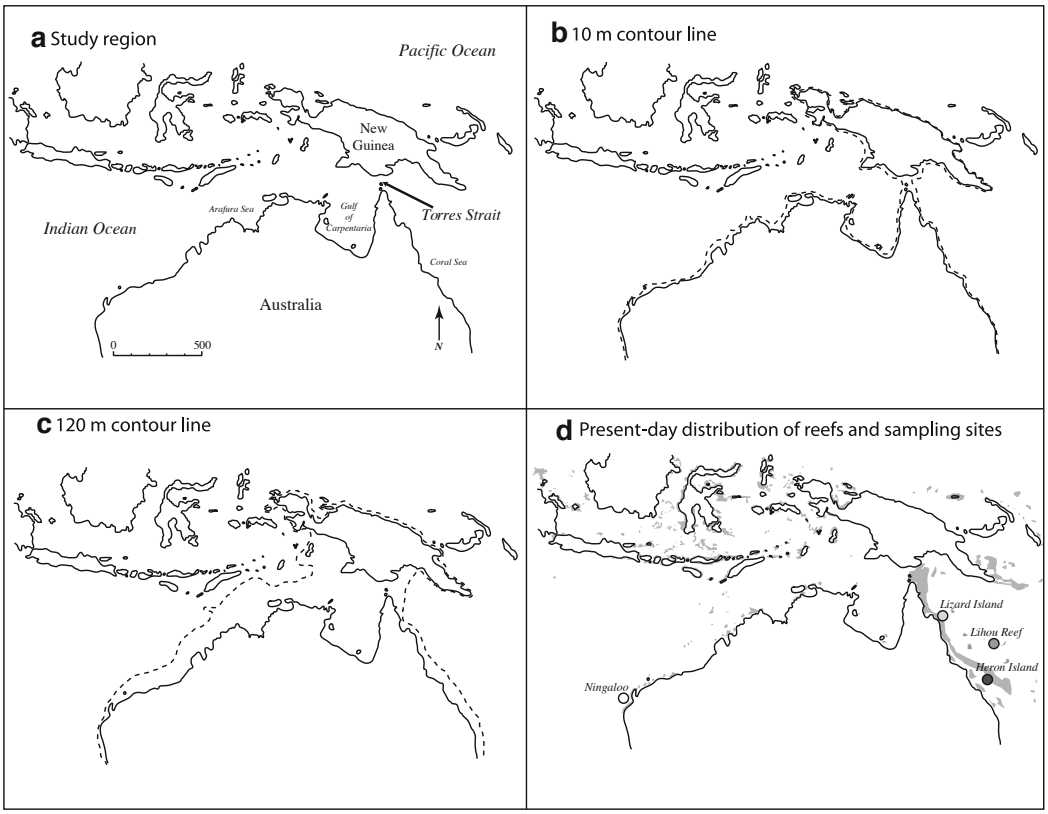


Figure 1. Australia and New Guinea. **a** study region, **b** the 10-m bathymetric contour is indicated by a dotted line; sea level has been at or below this level for most of the late Pleistocene, **c** the 120-m bathymetric contour is indicated by a dotted line; this represents the low sea level stand at ~17kyr, **d** study sites and present-day locations of coral reef; reef areas are shaded.

For marine species with present-day distributions spanning the Torres Strait, there are two possible historical scenarios since the inception of the barrier: vicariance caused by the land bridge with persistence of species in both Indian and Pacific Oceans or persistence in one region followed by recent colonization through the Torres Strait (Lukoschek et al. 2007). At the last glacial maximum (LGM), sea level was ~125 m lower than present. The exact geographic distribution of coral reefs (the habitat used by most species considered in the present study) in the Indo-Pacific at that time is uncertain (reviewed by (Montaggioni 2005). Although evidence for reef growth is found at many locations at the conclusion of the LGM, periods of growth were interspersed by reef drownings when sea level rise was too rapid for coral growth to match sea level rise (Montaggioni 2005). For example, the modern Great Barrier Reef (GBR) probably dates to no more than 10 kyr BP, as reef material is missing from the 18–9 kyr interval (Carter and Johnson 1986). Holocene strata in the GBR overlay older reef material from the previous interglacial, ~120 kyr (Marshall and Davies 1984). Similarly, the modern reef of Ningaloo in western Australia is less than 8 kyr and overlays reef strata from the previous interglacial (Collins et al. 2003). Coral reefs have been continuously present along the coastline of western Australia, however, at least since the late Pleistocene (128–121 kyr), albeit with northern contraction during the last glacial (Greenstein and Pandolfi 2008), whereas it is unknown whether there were regional refugia along the east coast of Australia.

To what degree marine animals can disperse across the modern Torres Strait is uncertain, as the oceanography is complex and not yet fully understood. Broadscale currents through this area are highly seasonal, with a complete reversal between summer and winter (Wolanski et al. 1988; Margvelashvili et al. 2008). The average westerly flow is 3-times stronger (0.25 Sv) and is sustained for approximately 3-times as long as the easterly current (Margvelashvili et al. 2008). The water depth in the Torres Strait ranges from 5 to 25 m (Fig. 1), and there are many small reef patches and islands, around which complex eddies can form (Wolanski et al. 1988; Margvelashvili et al. 2008). Overall, this oceanographic profile implies that there is limited sustained through-flow of water through this area. Gene flow through the Torres Strait is therefore likely to be restricted and directionality of dispersal may be seasonal. Genetic surveys taken from sites in the present-day Torres Strait indicate strong genetic affinities with GBR populations for coral trout (van Herwerden et al. 2009) and littorine snails (Reid et al. 2006). For samples from the Gulf of Carpentaria, sea snakes (Lukoschek et al. 2007) and barramundi (Chenoweth et al. 1998) also are genetically similar to GBR samples, indicative of east to west con- temporary movement or post-glacial colonization in a westerly direction. In contrast, for sea turtles, Gulf of Carpentaria populations are more similar to western Australian populations than to populations from the east (Dethmers et al. 2006).

In the present study, we evaluate historical and contemporary barriers to movement for marine animals distributed on either side of the Torres Strait. We survey four co-distributed coral reef fishes (see ‘‘Materials and methods’’) for mtDNA variation and compare these results with previously published mtDNA surveys for a variety of marine animals. First, we investigate whether mtDNA

variation is consistent with (a) vicariance across the Torres Strait, which would be expected if both western and eastern regional populations persisted since the last marine connection ~120 kyr ago, or (b) post-glacial colonization from either region. Second, in the instance that vicariance is most likely, we examine the age of genetic divergence within species and specifically test whether single or multiple episodes of vicariance typify observed genetic patterns. Finally, we use coalescent approaches to determine whether present-day gene flow is consistently asymmetrical in a westerly or easterly direction.

Materials and methods

In the present study, we combine new mtDNA sequence data from four species of coral reef fishes with data gathered from published literature from an additional nine species of marine animals with varying habitat requirements.

Focal study species

The four reef fish species chosen for this study vary in life history traits that may affect dispersal in reef fishes (Riginos and Victor 2001) including egg type, pelagic larval duration (PLD), and larval behavior (see Table 1 and Electronic Supplementary Material, ESM). Based on these qualitative assessments of dispersal potential, we estimated that dispersal is most restricted for *Apogon doederleini*, followed by *Pomacentrus coelestis*, *Dascyllus trimaculatus*, and *Acanthurus triostegus*. Previous genetic surveys for *Apogon doederleini* (Gerlach et al. 2007), *P. coelestis* (Gerlach et al. 2007), and *D. trimaculatus* (Bernardi et al. 2002; Leray et al. 2010) are consistent with these expectations, but for *Acanthurus triostegus* statistically significant genetic structure has been reported over at short distances (Planes 1993; Planes et al. 1998). Thus, we predict that genetic differentiation as estimated by mtDNA should be greatest for *Apogon doederleini*, followed by *P. coelestis*, *D. trimaculatus*, and *Acanthurus triostegus*, but that aberrant results for *A. triostegus* are possible.

Field collections and mtDNA sequencing

The study region included western and eastern Australia, with sites on either side of the Torres Strait. Fishes were collected from Heron (23° 260 S, 151°550 E) and Lizard (14°400 S, 145°280 E) Islands in the Great Barrier Reef (GBR), Lihou Reef (17°250 S, 151°400 E) in the Coral Sea (eastern populations), and Ningaloo Reef Marine Park (23°100 S, 113°450 E) in western Australia (western population) (Fig. 1). A limitation to most studies of coral reef organisms from this region (including the present study) is that sampling locations are distant from the putative barrier of the Torres Strait, with Ningaloo Reef and northern GBR as the usual flanking sites. This unsatisfactory sampling design arises from a number of factors that are difficult to circumvent. Reefs are diffuse or poorly developed along coastal Australia west of the Torres Strait (Fig. 1), with soft sediments typifying the Gulf of Carpentaria. The reefs of northern western Australia are poorly surveyed in general due to difficulties with access (requiring permissions from traditional owners) and dangerous working conditions

Table 1. Mitochondrial DNA surveys of animals distributed across the Torres Strait.

Species	Common name	Early life stages and habitat usage	Predicted dispersal	Live on/near coral reefs?	Data reference
Reef fishes					
Apogon doederleini	Doederlein's cardinalfish	Males brood eggs in their mouths; larvae are released and are planktonic for 16–27 days. Adults are free swimming and strongly reef associated	Low	Yes	Present study
Pomacentrus coelestis	Neon Damselfish	Eggs are benthic and larvae can settle on the outer reef in low quality habitat following an 18–20 day PLD. Adults are free swimming and reef associated	Medium	Yes	Present study
Dascyllus trimaculatus	Three-spot damselfish	Eggs are benthic and young recruit to large anemones following a 22–26 day PLD. Adults are free swimming and reef associated	Medium	Yes	Present study
Acanthurus triostegus	Convict surgeonfish	Eggs are pelagic with a long PLD of 40+ days. Adults are reef associated generalists but are often found in high-flow turbid environments such as coastal bays and harbors	High	Yes	Present study
Other vertebrates					
Aipysurus laevis	Olive sea snake	Adults free swimming, found near coral reefs and over soft bottoms; olive sea snakes are viviparous	Low	Yes	Lukoschek et al. (2007)
Stegostoma fasciatum	Zebra shark	Adults are free swimming and reef associated; zebra sharks are ovoviviparous	Low	Yes	Dudgeon et al. (2009)
Chelonia mydas	Green turtle	Adults are pelagic but probably return to natal area to breed and nest	Low	Sometimes	Dethmers et al. (2006)
Mollusks					
Haliotis asinina	Donkey ear abalone	Adults live on shallow hard substrate; lecithotrophic larvae settle after 3–4 days	Low	Yes	Imron et al. (2007)
Echinolittorina trochoides B	Periwinkle	Adults live on shallow hard substrate; planktotrophic larval phase up to 4 weeks	High	Yes	Reid et al. (2006)
Echinolittorina vidua	Periwinkle	Adults live on shallow hard substrate; planktotrophic larval phase up to 4 weeks.	High	Yes	Reid et al. (2006)
Other invertebrates					
Scylla serrata	Mud crab	Adults live in estuaries but migrate offshore to release eggs; planktotrophic larval phase.	High	No	Gopurenko and Hughes (2002)
Diadema setosum-a	Urchin	Adults live on shallow hard substrate; planktotrophic larval phase up to 6 weeks.	High	Yes	Lessios et al. (2001)
Holothuria nobilis	Sea cucumber	Adults live on coral reefs, planktotrophic larval phase greater than 10 days.	Medium	Yes	Uthicke and Benzie (2003)

(crocodiles). Although sampling sites adjacent to a suspected barrier are ideal for evaluating the effect of that barrier, the aforementioned considerations constrained out investigation (and many others included in Table 1). Therefore, we cannot decouple the effect of a large geographic distance (Ningaloo to northern GBR) from a historic barrier. From our study sites, tissue samples were collected on SCUBA or snorkel, with 15–20 individuals collected per species, per site. Fresh tissue was preserved in 100% ethanol and then stored at -80°C.

Total genomic DNA was extracted from muscle tissue using a salt extraction method (Aljanabi and Martinez 1997). Cytochrome oxidase 1 (CO1) and the control region of the mitochondria were amplified in 30µl polymerase chain reactions (PCR), with each reaction tube containing 1µl of DNA, 3µl each of NH_4 buffer and DNTPs (10mM), 1.5µl of $MgCl_2$ and 0.2µl of Taq-polymerase (Bioline). CO1 was amplified using the universal fish primers Fish F1, Fish F2, Fish R1, and Fish R2 (Ward et al. 2005). The control region was amplified using the universal primers CR-A and CR-E (Lee et al. 1995). PCR reactions were purified using a standard Exo-SAP protocol (using reagents and specific protocols from New England Biolabs) and outsourced to Macrogen (Korea) for DNA sequencing via capillary electrophoresis. Sequence trace files were manually checked and edited, using Codon Code Aligner v3.0.2.

Data from earlier published studies

In addition to the new mtDNA sequence data generated specifically for this study, we conducted a literature search to identify studies that had sampled marine animals on either side of the Torres Strait and used mtDNA sequences to determine population structure. In order to standardize the analyses across datasets, the published mtDNA datasets were reconstructed using GenBank accession information and information provided in the published papers. Several relevant studies were not included because there was insufficient information to recover the original sequence data.

Phylogeographic, population genetic, and coalescent analyses

For each of the four newly sampled reef fish species, we constructed haplotype networks to visualize the data (although our main conclusions derive from summary statistics and coalescent analyses). Haplotype networks (combined COI and region) were produced in TCS v1.18 (Clement et al. 2000) with a 95% connection limit. Because regional *Apogon doederleini* sequences were highly divergent and reciprocally monophyletic, we were concerned that western and eastern samples might not be sister taxa and that their divergence could be due to factors not associated with vicariance due to the Torres Strait land bridge. To verify that our sampled *Apogon* were sister taxa, we augmented our phylogenetic analyses for *Apogon* by searching GenBank for COI sequences for species within the same genus. All *Apogon* sequences were manually aligned with Se-AL v2.0a11 (Rambaut 1996). A maximum likelihood tree was constructed in PAUP version 4.0b10 (Swofford 1998) using a GTR+G+I model, as selected by Modeltest (Posada and Crandall 1998) using the Akaike information criterion as the most appropriate model of molecular evolution for the data. Branch support was assessed using

bootstrapping based on neighbor-joining (with the same GTR+G+I model) and parsimony trees (with maxtree per bootstrap replicate limited to 1,000) in PAUP, and also using the Bayesian approach implemented in MrBayes (ver. 3.1.2, Ronquist and Huelsenbeck 2003) using a GTR model of molecular evolution.

For our reef fish species, we assessed diversity by population and region. DNA polymorphism (number and diversity of haplotypes, polymorphic sites, and π, the average number of pairwise differences) and Tajima's *D* statistic (as an estimator of deviations from neutral, equilibrium expectations, Tajima 1989) were estimated for all populations sampled using DNAsp v5.00.07 (Rozas et al. 2003). In addition, population structure among all pairs of populations was evaluated using Meirman's F_{ST} (Meirmans 2006), with the significance assessed based on 999 permutations using GenoDive version 2.0b19 (Meirmans and Van Tienderen 2004).

Because exact sampling regimes differed among published studies, all comparisons across species had to be at a regional level. We consider our two regions to be western Australia (west of the Gulf of Carpentaria, with Ningaloo Reef the most commonly sampled) and eastern Australia (east of the Torres Strait, with the GBR region being most commonly sampled). To assess regional gene diversity (proportional to N_e, effective population size) within populations, we estimated π for each sampled population in a given study and then took the regional mean value across the sampled populations; this gave us an average measure of diversity per population per region. For some studies, sampling was diffuse within regions and not linked to specific populations—for instance, the two *Echinolittorina* species consisted of a few individuals from both the west and the east but collected from several locations. For such cases, we analyzed all individuals from a region as a single population (see ESM for sampling schemes in each study). Population structure was estimated using the AMOVA framework (Excoffier et al. 1992) as implemented in GenoDive. For studies where there were multiple populations sampled per region, we estimated Φ_{CT}, the partitioning of variation among regions. For all studies, we estimated Φ_{ST} for the geographically most proximate pair of populations spanning the Torres Strait (based on the 10m contour, but not sampled from within the Torres Strait itself; a few studies had 1-2 individuals from reef locations within the present-day Torres Strait region). The Torres Strait was chosen as the exact point of demarcation as it has the shallowest seabed and was the last land area to be inundated. For our fish samples, we used Ningaloo and Lizard Island populations as all four species were sampled from these two locations.

In addition to the standard estimates of genetic diversity and population structure described above, we used a coalescent approach to estimate genetic diversity, divergence time, and migration rates. *IMa* (Hey and Nielsen 2007) is based on a two population model, whereby populations have diverged at the same point in the past but migration since divergence is permitted (Nielsen and Wakeley 2001). This model captures the essential elements of our system, namely two regional populations that have diverged, yet migration between populations can occur. Again, we concentrated

on the geographically most proximate pair of populations spanning the Torres Strait for previously published studies and used Ningaloo and Lizard Island populations for our fish species. Initial runs used Bayesian priors of θ = [1, 10], eastward and westward migration = [0, 10], divergence time = [0, 5], with a burn-in of 1,000,000 steps and a total of 20,000,000 steps using the Computational Biology Service Unit at Cornell University (April 2008 version). Following initial runs, the total number of steps was adjusted such that all ESS estimates were greater than 200 (see *IMa* documentation). For the four fish species, we also conducted a series of runs with maximum divergence time adjusted such that the time period of 1–500,000 years was considered, assuming a molecular clock rate of 1.2% per million years for COI based on several reef fish species separated by the Isthmus of Panama (Bermingham et al. 1997). All running conditions per species were repeated in triplicate, and results were found to be consistent between runs.

Because different mtDNA gene regions have been used in different studies, estimators of $N\mu$ (heterozygosity or gene diversity) such as π or θ will be affected by gene-specific neutral mutation rates (μ) such that relative population sizes between species cannot be inferred across studies. π and θ values, however, can be compared between western and eastern regions within each species. For example, if these values were consistently greater in western Australia, we might conclude that population sizes (and by inference long-term population stability) have historically been greater in the western region. Among the fish species considered here, π and θ values can be compared among species under the assumption of a single common mutation rate for the same gene region (i.e., COI).

We also tested the specific hypothesis that all four reef fish species experienced simultaneous divergence in a hierarchical approximate Bayesian computational framework (Hickerson et al. 2006). Here, data from multiple species are considered in combination and the observed fit to parameters shared among species (such as divergence time) evaluated. A single mutation rate is implicitly assumed, although the exact rate does not have to be specified. The program msBayes (ver. 20081106, Hickerson et al. 2007) was used to simulate data with priors guided by *IMa* results. Divergence time was set to τ = [0, 5], and simulations were run both with migration fixed at zero and allowed to vary, m = [0, 10], and max θ equal to 50 or 100. Theta values were set both at a maximum equal to 50 and 100. The hyper-parameter Ψ, which reflects the number of discrete divergence events, was allowed to vary between 1 (i.e., simultaneous divergence across species) and 4 (separate divergence time for each of the four species). The top bounds for the ancestral population size relative to current population size was set to 1, matching a scenario of vicariance between regions, and we used 1 million draws per set of simulation conditions and a tolerance of 0.002. The observed data from Ningaloo and Lizard Island populations were compared against simulations using the estimate of Ω as a measure of fit, whereby $\Omega > 0.01$ can be considered as rejecting simultaneous divergence (Hickerson et al. 2006).

Results

MtDNA was sequenced from *Apogon doederleini*, *Acanthurus triostegus*, *Dascyllus trimaculatus*, and *Pomacentrus coelestis* populations from Ningaloo (western Australia), Lizard Island (GBR, eastern Australia), and some additional populations in the GBR and Coral Sea (exact populations varied among species for these latter locations). COI was successfully PCR-amplified and sequenced for all four species, and the control region was also amplified and sequenced for all species except *A. triostegus*. (Multiple attempts at PCR amplification with various primer combinations targeting the control region were unsuccessful for this species.) Sequences have been deposited in GenBank (Acc Nos. JF717878-JF718183). Because of the imbalance in the species representation for control region data, for some analyses we primarily present COI results, so comparisons can be made across fish species. From published literature, we were able to reconstruct datasets for nine species of marine animals; all sources of data are summarized in Table 1.

Haplotype networks showed a range of phylogeographic patterns for the reef fish, ranging from no obvious association with geographic location for *Acanthurus triostegus* and *D. trimaculatus* to reciprocal monophyly between western and eastern populations of *Apogon doederleini* (Fig. 2). The large number of fixed differences between *A. doederleini* clades prompted us to verify that these two clades are sister taxa. Using all *Apogon* COI sequences found on GenBank, we conducted several phylogenetic analyses and in all cases found high support for the monophyly of our two *A. doederleini* clades; that branch was present in the maximum likelihood tree and had a bootstrap support of 96% from neighbor-joining and a 97% posterior probability in the Bayesian analysis based on 1 million MCMC steps following a burn-in of 11.5 million steps (Fig. of phylogenetic tree available in ESM). Although it is possible that the western and eastern forms of *A. doederleini* are different species, they appear to be sister taxa and therefore their divergence time is relevant to this study.

Standard estimators of population diversity among our fish species showed no consistent regional patterns. The population locations with the highest and lowest levels of diversity differed by species and estimator. There were no significant deviations from neutral equilibrium expectations as measured by Tajima's *D*. There was significant genetic differentiation between populations as measured by F_{ST} for many comparisons involving the single western population of Ningaloo with other populations. Estimates of gene diversity and population structure between Ningaloo and Lizard Island fish populations are summarized in Table 2, where there was significant genetic differentiation for *A. doederleini* and *P. coelestis* but not for *D. trimaculatus* or *Acanthurus triostegus*. (Population-specific results for these fish species are available in the ESM.) Table 2 also summarizes the results from species previously studied by other investigators. Many species showed substantial and significant genetic partitioning between western and eastern regions, whereas other species had a complete lack of population genetic structure. Results from the coalescent *IMa* analyses matched standard estimates of population diversity and structure: neither western nor eastern regions were consistently more

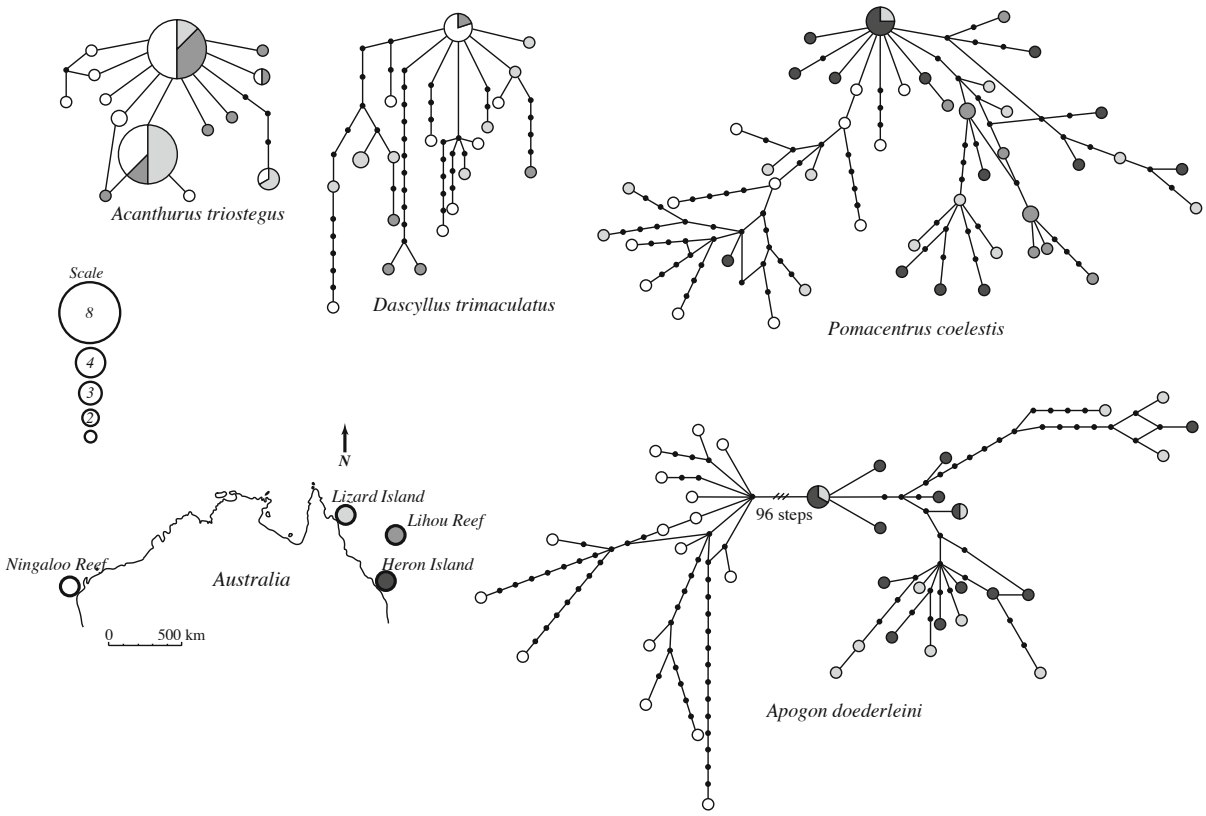


Figure 2. Haplotype networks for four reef fish species. Small black circles indicate haplotypes not observed in our samples. Haplotypes are shaded by location and the size is proportional to the observed number. Haplotypes are comprised of concatenated COI and control regions except for *Acanthurus triostegus* where only COI data are included.

Table 2. Gene diversity, genetic differentiation, and migration of animals distributed across the Torres Strait.

	Standard estimates			Coalescent estimates (90% credibility intervals)						
	Gene diversity		Genetic differentiation (*F*-statistics)	Gene diversity		Genetic differentiation (Divergence time scaled by *N*)	Migration rate		Number of migrants	
Species	π_{WA} (%)[a,b]	π_{EA} (%)[a,b]	ϕ_{CT} / ϕ_{ST}[d]	ϑ_{WA}[a,e]	ϑ_{WA}[a,e]	$t\mu$[d,e]	m_{WA}/μ[a]	m_{EA}/μ[a]	Nm_{WA}[a,f]	Nm_{EA}[a,f]
Reef fishes										
Apogon doederleini	0.23	0.43	0.905**	3.54 (0–8.97)	3.74 (0–7.78)	4.02** (0.47–5.00)	0	0	0	0
Pomacentrus coelestis	0.51	0.30	0.064*	51.37 (12.41–51.37)	51.37 (13.65–51.37)	0.83** (0.29–1.78)	0	4.05	0	208
Dascyllus trimaculatus	0.09	0.13	0.078	0.33 (0.04–6.38)	0.33 (0.05–6.50)	0.06 (0.00–4.43)	10.0 (0–10)[g]	10.0 (0–10)[g]	3.3	3.3
Acanthurus triostegus	0.37	0.32	0.059	0.50 (0.40–27.08)	21.77 (0.64–27.08)	0.23 (0.15–5.00)	0 (0–9.01)	4.54 (2.32–10.0)	0	99
Other vertebrates										
Aipysurus laevis	0.13	0.06	0.709**/0.409**	2.45 (0.74–26.36)[g]	0.09 (0–2.22)	1.21** (0.50–5.00)	0 (0–9.01)	0 (0–9.01)	0	0
Stegostoma fasciatum	0.15	0.07	0.290**	0.95 (0.18–9.21)	0.24 (0–1.01)	2.29** (0.48–5.00)	6.08 (1.52–10.00)	1.27 (0–8.38)	5.8	0.3
Chelonia mydas	0.83	0.85	0.549**/0.777**	0.47 (0–1.52)	1.87 (0–5.36)	0.62** (0–4.54)	0.06 (0–3.15)	0.18 (0–1.74)	<0.1	0.3
Mollusks										
Haliotis asinina	0.15	0.08	0.956/0.969**	1.12 (0–3.59)	0.17 (0–1.26)	1.36** (0.89–5)	0 (0–0.90)	0 (0–2.27)	0	0
Echinolittorina trochoides B	0.41	0.23	-0.021	62.88 (12.49–62.88)	7.64 (6.89–62.88)	1.20** (0.25–2.94)	0 (0–9.01)	0 (0–8.94)	0	0
Echinolittorina vidua	0.86	0.31	0.767**	95.89 (19.81–95.89)	12.81 (10.98–95.89)	0.99** (0.06–3.08)	0 (0–8.91)	0 (0–8.85)	0	0

Table 2. continued

Other invertebrates										
Scylla serrata	0.58	0.04	0.848**/0.678**	1.60 (0–3.55)	0.07 (0–0.55)	1.07** (0.10–4.15)	0 (0–1.62)	0 (0–7.48)	0	0
Diadema setosum-a	0.15	0.26	0.032	0.27 (0–11.05)	1.06 (0.35–14.48)	0.20 (0–4.46)	0 (0–8.67)	9.29 (2.44–10)	0	9.8
Holothuria nobilis	0.39	0.46	0.007/0.029	25.76 (14.60–80.83)	36.11 (19.21–80.83)	0.81** (0.53–1.18)	2.42 (1.13–10)	1.29 (0–9.44)	62	47

[a] Specific geographic regions are indicated and can include Western Australia (WA) and eastern Australia (EA). Migration (in the conventional sense) is into the population, e.g., m_{WA}/μ is the rate of migration (scaled by the mutation rate) into Western Australia.
[b] Average number of pairwise differences; where there are multiple populations sampled per region, π is the average value across the regional populations.
[c] Hierarchical AMOVA is used in cases where there are two or more populations per region. Otherwise pairwise ϕ_{ST} for the nearest two populations spanning the Torres Strait is reported.
[d] * indicates $P < 0.05$; ** indicates $P < 0.01$; based on 999 replicates for AMOVA and log-likelihood tests for $t\mu$. See text for more detail.
[e] Based on per locus estimates. These are not comparable across studies as different studies employed different mtDNA regions; the exception is our new fish data where the same region of COI was sequenced.
[f] Number of migrants per generation; the product of most probable point estimates for ϑ and m/μ.
[g] Discontinuous posterior probability distribution.

genetically diverse across species, and many species had significant genetic divergence between regions. Estimates of migration rates and numbers of migrants also differed among species with no consistent directionality inferred.

Due to the common sampling scheme for reef fishes (gene region and populations sampled), we were able to explicitly test the hypothesis of simultaneous divergence across these four species. For all four sets of conditions examined (max θ = 50, 100 and max m = 0, 10), the estimated value for Ω was equal to ~0.2, substantively greater than the suggested threshold of 0.01 for rejecting a single divergence time (Hickerson et al. 2006).

Discussion

By combining new mtDNA sequence data from four reef fish species with previously published data, we were able to examine patterns of genetic diversity and divergence for thirteen species of marine animals across the historic Torres Strait land bridge. We found that there was significant population structure across the Torres Strait for many species, consistent with long-standing vicariance. The degree of genetic structure, however, ranged from very high (Φ_{ST} > 0.7: Table 2, and reciprocal monophyly of mtDNA haplotypes between regions: Fig. 2) to negligible (Φ_{ST} not significant, no phylogeographic structure). Coalescent analyses of reef fish mtDNA data imply that there was no single divergence date across species, a pattern qualitatively matched in previous studies. Estimates of migration similarly varied among species and directionality was idiosyncratic.

Broadscale phylogeographic concordance

The Torres Strait land bridge was an intermittent barrier to dispersal in the late Pleistocene for approximately 90,000–100,000 years, until its most recent inundation ~7,000 years ago (Voris 2000). Evidence that this barrier and/or the large geographic distance between sampling sites (see ''Materials and methods'') contributed to population structure between western and eastern populations was strong for eight out of thirteen marine animals considered here (i.e., significant Φ_{ST} and matched by significant divergence times in *IMa* analyses: Table 2). For instance, genetic differentiation between Indian and Pacific Ocean populations was high for the reef fish *Apogon doederleini* and strong for *Pomacentrus coelestis*. These results match observations from previous studies representing a diverse cross-section of marine animals with perfect or nearly perfect reciprocal monophyly between western and eastern Australian populations, including coral trout (van Herwerden et al. 2009), sea snakes (Lukoschek et al. 2007), sea turtles (Dethmers et al. 2006), abalone (Imron et al. 2007), a littorine snail (Reid et al. 2006), and mud crabs (Gopurenko et al. 1999). (See also significant Φ_{ST} values for some of these same species in Table 2).

The repeated observation of high population structure (Table 2) across species with widely differing ecologies (Table 1) is the hallmark of past vicariance (Avise et al. 1987; Avise 2000). Thus,

our genetic data support a historical scenario whereby regional populations of many species persisted in both Indian and Pacific Ocean reefs through the late Pleistocene and Holocene, although the large geographic distance between sampling points would also be expected to contribute to population structure. If a species had become locally extinct in one region, then colonization would have been within the last 7,000 years, and we would expect to find no appreciable genetic divergence between populations and reduced diversity in the colonized region relative to the source (Hewitt 1996). Among the five panmictic species (based on *F*-statistics), mtDNA gene diversity was greater (at least two times difference for either conventional or coalescent estimators) in the western region for *E. trochoides B*, greater in the eastern region for *Acanthurus triostegus* and *Dascyllus setosum-a*, and equivalent for *D. trimaculatus* and *H. nobilis*. Whereas we cannot exclude colonization scenarios for these five species, there is no consistent pattern that would indicate that one region historically supported greater species diversity than the other.

Multiple ages of vicariance

Although the pattern of vicariance and secondary contact is common to many species distributed across tropical Australia, the age of divergence between Indian and Pacific Ocean populations appears to differ markedly among species. Because studies compiled from the literature employed different mtDNA gene regions, the estimated divergence times ($t\mu$: Table 2) can reflect gene or taxon-specific mutation rates (μ) and are not comparable. We can, however, make direct comparisons among our surveyed reef fishes under the assumptions of a molecular clock (without using a specific clock rate). The sequenced gene region was the same and the species are from fairly similar taxonomic groups, so the assumption of a common mutation rate is reasonable inasmuch that such an assumption is ever valid. For these reef fish species, coalescent analyses using *IMa* indicated different divergence times, evidenced by peak posterior probabilities for different $t\mu$ values among species (Fig. 3). The hierarchical approximate Bayesian computation framework also allowed us to explicitly reject the hypothesis of simultaneous divergence ($\Omega > 0.01$). In Fig. 3, we have used a molecular clock calibrated for fish COI to illustrate these varying divergence times in years, but regardless of any specific clock employed, these differences among species would remain. If we assume that the clock used is roughly correct, then divergence times for *P. coelestis* and *Apogon doederleini* appear to substantially pre-date the last glacial cycle.

Prior to the most recent flooding of the Torres Strait land bridge, there would have been a marine connection at approximately 120kyr and at other older times of high sea level stands (Chappell and Shackleton 1986; Voris 2000). The old divergence of *P. coelestis* and *A. doederleini* between western and eastern Australia implies that Holocene gene exchange has been insufficient to erase signatures of Pleistocene separation times and that regional populations have persisted in isolation from each other for a long time (albeit with some post-divergence gene flow for *P. coelestis*: Table 2). In contrast, the recent west–east divergence for *D. trimaculatus* and possible lack of divergence for *Acanthurus triostegus* (reflected in a flat probability surface: Fig. 3) suggest that there have been

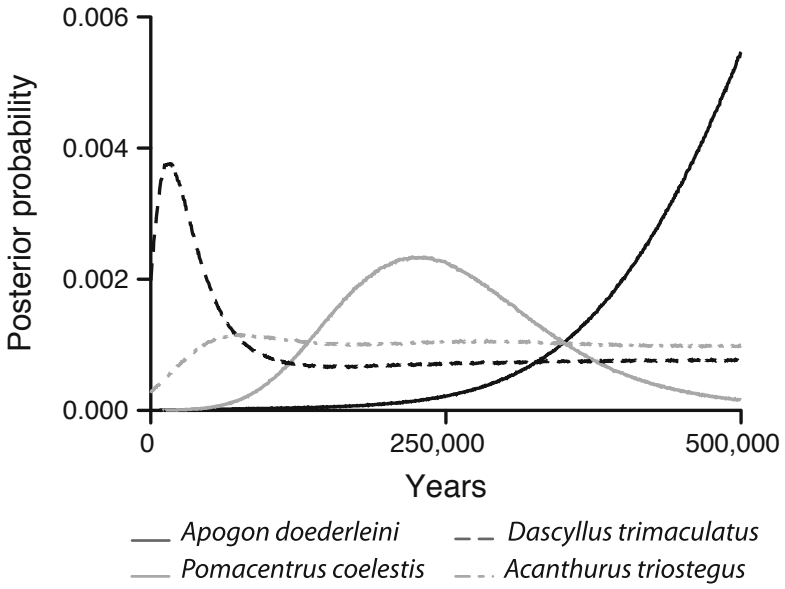


Figure 3. Posterior probabilities of divergence times between Indian and Pacific Ocean populations of reef fishes based on *IMa*.

strong genetic connections between regions, possibly entailing large-scale colonization in addition to continuing gene flow. One possibility is that *A. triostegus* and *D. trimaculatus* circumvented the Torres Strait barrier via stepping-stone connections north of New Guinea. For these species, recent (Holocene) events obscure the past, and therefore we are unable to make strong inferences about regional patterns of persistence. Similarly, our inference about regional patterns is somewhat hampered as it is based on sampling of a single site in western Australia (Ningaloo) that is quite distant from the Torres Strait itself; Ningaloo is the sampling site for our fishes and western site most proximate to the Torres Strait in most other studies. Thus, we have detected divergence between western and eastern Australian populations and have assumed that vicariance due to the past land bridge is the cause of this divergence and not some other factor, for example, specific to Ningaloo. The studies that have more extensive sampling (multiple populations per region) also showed substantial regional population structure (Φ_{CT}: Table 2), indicating that western and eastern populations of many species are distinct and that there is nothing unusual about Ningaloo as representative of western Australian populations.

Whereas the inclusion of additional loci per species (i.e., nuclear loci) would strengthen the conclusions for individual species (due to the inherent high variance in coalescence times among independently assorting loci), our emphasis was on uncovering regional patterns where species could be considered as independent realizations of shared geographic history. For such a comparative framework, mtDNA sequences deliver substantial information relative to effort and costs.

The overall pattern that emerges from fishes (and is consistent with other species surveyed: Table 2) is that the Torres Strait and large expanse of northern Australia has been a long-standing potent barrier for some species, whereas for other species these barriers have been transitory, with genetic connections between regions intermittently reestablished during periods of high sea level stands. A similar diversity in phylogeographic patterns has been described for marine species at the Atlantic–Mediterranean transition (Patarnello et al. 2007), where a complete barrier to movement has also been historically intermittent.

Life history traits, divergence times, and gene flow

One of the goals of historical ecology is to investigate how species composition of ecological communities has changed over time and to identify suites of life history traits that have affected community membership (Wares 2002; Hickerson et al. 2010; Tager et al. 2010). Similarly, how life history traits affect dispersal (often inferred from genetic patterns) among established populations is a topic of considerable interest. For marine animals, factors such as pelagic larval duration (PLD), egg type (for reef fishes), and larval feeding mode (for invertebrates) are obvious candidates for predicting dispersal and colonization ability. Species with non-planktonic larvae nearly always have greater population structure than those with planktonic larvae (reviewed by Weersing and Toonen 2009). Egg type and larval feeding mode can be correlated with differing patterns of genetic structure (Watts and

Thorpe 2006; Pelc et al. 2009; Riginos et al. 2011), whereas effects of PLD on genetic differentiation are probably less important (Weersing and Toonen 2009).

In the present study, we find that qualitative predictors of dispersal ability match rank order estimates of population genetic structure and gene flow for reef fishes. Among the reef fishes, the only species with a high posterior probability for zero gene flow was *Apogon doederleini*, and this lack of gene flow matched expectations based on larval biology. *A. doederleini* brood their young in their mouths (Takeyama et al. 2007), develop quickly, and larvae orient themselves toward natal reefs (Atema et al. 2002; Gerlach et al. 2007). The predicted high-dispersal species, *Acanthurus triostegus* (planktonic eggs, long PLD), was panmictic, matching predictions based on larval morphology. Unlike allozyme surveys of the same species (Planes 1993; Planes et al. 1998; Planes and Fauvelot 2002), however, we did not detect significant geographic differentiation of mtDNA. For the putative intermediate dispersers, *P. coelestis* and *D. trimaculatus* with benthic eggs and relatively short PLDs, the results were mixed with significant population structure for *P. coelestis* and not for *D. trimaculatus*. *P. coelestis* is known to orient toward reefs, which might contribute to genetic differentiation. Nothing specific has been published about *D. trimaculatus* larval behavior.

Alternative to egg and larval traits affecting genetic patterns, differences in divergence times could reflect the habitat preferences of each study species. In a recent study, Ayre et al. (2009) demonstrated that habitat specificity could be an important determinant of genetic differentiation for intertidal invertebrates. In the case of the Torres Strait, as sea levels rose and covered the land bridge, some species could have recolonized the newly submerged areas before others. Such a pattern has been demonstrated with the Isthmus of Panama, with divergence for coastal species post-dating those species that are limited to deep water or isolated islands (Knowlton and Weigt 1998). *Acanthurus triostegus* are generalists and are often found in turbid, high-flow environments such as coastal bays and harbors (pers. obs.). These characteristics may have allowed *A. triostegus* to quickly colonize new habitat. Of the other three fish species, *P. coelestis* has the least specific habitat preferences, as they settle on the outer reefs in low quality habitat (Gerlach et al. 2007), so this species would probably be able to colonize new habitat quickly. *Apogon doederleini* larvae settle deep in the reef (Gerlach et al. 2007) and *D. trimaculatus* settle onto large anemones (Bernardi et al. 2001) or coral heads in the absence of anemones (pers. obs.), indicating greater habitat specificity, especially for *D. trimaculatus* (the habitat specificity of *D. trimaculatus* may explain the lower gene diversity relative to the other fishes: Table 2). Although habitat specificity would predict that *P. coelestis* would be a better colonist than *D. trimaculatus*, we find greater divergence between western and eastern populations for *P. coelestis* as compared to *D. trimaculatus*.

For the reef fishes considered here, there are possible correlations between species characteristics and west–east connections over time; however, it is impossible to make strong inference with only four species. The additional nine marine animals examined represent a diverse

array of phyla with a wide variety of traits relevant to dispersal ability and habitat specificity. Our *a priori* predictions of relative dispersal ability do not match observed genetic patterns and no obvious differences emerge based on coarse habitat type (e.g., coral reef and pelagic). Fine-tuning our predictions regarding dispersal ability or habitat specificity (for instance, restricting analyses to closely related species that differ in single relevant traits, or using quantitative niche predictions) could potentially uncover correlations that are obscured by coarse analyses such as those presented here. Alternatively, it is quite possible that species and community responses to historical barriers and fluctuating local conditions are idiosyncratic and not predictable (Patarnello et al. 2007; Tager et al. 2010). Such a conclusion may be warranted when the predictive characteristics comprise broad categories that cut across evolutionarily diverse taxa.

Acknowledgements

Thanks to J. D. Aguirre-Davies and F. MacKenzie for assistance in the field, and two anonymous reviewers for comments. The Computational Biology Service Unit from Cornell University, which is partially funded by Microsoft Corporation, was used for some analyses. Funding for this work was provided by the Australian Research Council (DP0878306 to CR), and the World Wildlife Fund (to EAT). Fishes were collected under permits from the Department of the Environment, Water, Heritage, and the Arts (AU- COM2008042), Great Barrier Reef Marine Park Authority and Department of Environment and Resource Management (G08/ 28114.1), Queensland Government Dept. of Primary Industries and Fisheries (118636), and WA Department of Environment and Conservation (SF006619). This study complies with Animal Ethics standards for the University of Queensland (permit SIB/817/08/ARC).

References

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA from PCR-based techniques. Nucleic Acids Res 25:4692–4693

Atema J, Kingsford MJ, Gerlach G (2002) Larval reef fish could use odour for detection, retention and orientation to reefs. Mar Ecol Prog Ser 241:151–160

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

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

Ayre DJ, Minchinton TE, Perrin C (2009) Does life history predict past and current connectivity for rocky intertidal invertebrates across a marine biogeographic barrier? Mol Ecol 18:1887–1903

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 144:757–767

Begg GA, Keenan CP, Sellin MJ (1998) Genetic variation and stock structure of school mackerel and spotted mackerel in northern Australian waters. J Fish Biol 53:543–559

Bermingham E, McCafferty SS, Martin AP (1997) Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus Molecular Systematics. Academic Press, San Diego, pp 113–128

Bernardi G, Holbrook SJ, Schmitt RJ (2001) Gene flow at three spatial scales in a coral reef fish, the three-spot *dascyllus*, *Dascyllus trimaculatus*. Mar Biol 138:457–465

Bernardi G, Holbrook SJ, Schmitt RJ, Crane NL, DeMartini E (2002) Species boundaries, populations and colour morphs in the coral reef three-spot damselfish (*Dascyllus trimaculatus*) species complex. Proc R Soc Lond B Biol Sci 269:599–605

Carter RM, Johnson DP (1986) Sea-level controls on the post-glacial development of the Great Barrier Reef, Queensland. Mar Geol 71:137–164

Chappell J, Shackleton NJ (1986) Oxygen isotopes and sea level. Nature 324:137–140

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

Chenoweth SF, Hughes JM, Keenan CP, Lavery S (1998) When oceans meet: a teleost shows secondary intergradation at an Indian-Pacific interface. Proc R Soc Lond B Biol Sci 265: 415–420

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9:1657–1659

Collins LB, Zhu ZR, Wyrvoll KH, Eisenhauer A (2003) Late Quaternary structure and development of the northern Ningaloo Reef, Australia. Sediment Geol 159:81–94

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 15:3931–3946

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- Pacific. Mol Ecol 18:248–261

Duncan KM, Martin AP, Bowen BW, de Couet HG (2006) Global phylogeography of the scalloped hammerhead shark (*Sphyrna lewini*). Mol Ecol 15:2239–2251

Elliott NG (1996) Allozyme and mitochondrial DNA analysis of the tropical saddle-tail sea perch, *Lutjanus malabaricus* (Schneider), from Australian waters. Mar Freshw Res 47:869–876

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Gerlach G, Atema J, Kingsford MJ, Black KP, Miller-Sims V (2007) Smelling home can prevent dispersal of reef fish larvae. Proc Natl Acad Sci USA 104:858–863

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 Freshw Res 53:849–857

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

Greenstein BJ, Pandolfi JM (2008) Escaping the heat: range shifts of reef coral taxa in coastal Western Australia. Global Change Biol 14:513–528

Hellberg ME (2007) Footprints on water: the genetic wake of dispersal among reefs. Coral Reefs 26:463–473

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. Biol J Linn Soc 58:247–276

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc Natl Acad Sci USA 104:2785–2790

Hickerson MJ, Stahl EA, Lessios HA (2006) Test for simultaneous divergence using approximate Bayesian computation. Evolution 60:2435–2453

Hickerson MJ, Stahl E, Takebayeyashi N (2007) msBayes: Pipeline for testing comparative phylogeographic histories using hierarchical approximate Bayesian computation. BMC Bioinformatics 8:268

Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. Mol Phylogenet Evol 54:291–301

Imron, Jeffry B, 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 16:289-304

Jordan DS (1908) The law of geminate species. Am Nat 42:73–80 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 150:659–670 Knowlton N, Weigt LA (1998) New dates and new rates for divergence across the Isthmus of Panama. Proc R Soc Lond B Biol Sci 265:2257–226

Lee W-J, Conroy J, Howell WH, Kocher TD (1995) Structure and evolution of teleost mitochondrial control regions. J Mol Evol 41:54–66

Leray M, Beldade R, Holbrook SJ, Schmitt RJ, Planes S, Bernardi G (2010) Allopatric divergence and speciation in coral reef fish: the three-spot *dascyllus*, *Dascyllus trimaculatus*, species complex. Evolution 64:1218–1230

Lessios HA, Kessing BD, Pearse JS (2001) Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. Evolution 55:955–975

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 16:3406–3422

Margvelashvili N, Saint-Cast F, Condie S (2008) Numerical modeling of the suspended sediment transport in Torres Strait. Cont Shelf Res 28:2241–2256

Marshall JF, Davies PJ (1984) Last interglacial reef growth beneath modern reefs in the Southern Great Barrier Reef. Nature 307: 44–46

Mayr E (1954) Geographic speciation in tropical echinoids. Evolution 8:1–18

Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution 60: 2399–2402

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Mol Ecol Notes 4:792–794

Montaggioni LF (2005) History of Indo-Pacific coral reef systems since the last glaciation: development patterns and controlling factors. Earth Sci Rev 71:1–75

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov Chain Monte Carlo approach. Genetics 158:885–896

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. Annu Rev Ecol Syst 25:547–572

Patarnello T, Volckaert FAMJ, Castilho R (2007) Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographic break? Mol Ecol 16:4426–4444

Pelc RA, Warner RR, Gaines SD (2009) Geographical patterns of genetic structure in marine species with contrasting life histories. J Biogeogr 36:1881–1890

Planes S (1993) Genetic differentiation in relation to restricted larval

dispersal of the convict surgeonfish, *Acanthurus triostegus*, in French Polynesia. Mar Ecol Prog Ser 98:237–246

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

Planes S, Parroni M, Chauvet C (1998) Evidence of limited gene flow in three species of coral reef fishes in the lagoon of New Caledonia. Mar Biol 130:361–368

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818

Rambaut AE (1996) Se-Al: sequence alignment editor v2.0a11. Available at http://evolve.zoo.ox.ac.uk/

Reeves JM, Chivas AR, Garcia A, Holt S, Couapel MJJ, Jones BG, Cendon DI, Fink D (2008) The sedimentary record of palaeoenvironments and sea-level change in the Gulf of Carpentaria, Australia, through the last glacial cycle. Quaternary Int 183:3–22

Reid DG, Lal K, Mackenzie-Dodds J, Kaligis F, Littlewood DTJ, Williams ST (2006) Comparative phylogeography and species boundaries in *Echinolittorina* snails in the central Indo-West Pacific. J Biogeogr 33:990–1006

Riginos C, Victor BC (2001) Larval spatial distributions and other early life-history characteristics predict genetic differentiation in eastern Pacific blennioid fishes. Proc R Soc Lond B Biol Sci 268:1931–1936

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. doi:10.1111/j.1600- 0587.2010.06511.x

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Rozas J, Sánchez-Delbarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analysis by the coalescent and other methods. Bioinformatics 19:2496–2497

Swofford DL (1998) PAUP*, Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts

Tager D, Webster JM, Potts DC, Renema W, Braga JC, Pandolfi JM (2010) Community dynamics of Pleistocene coral reefs during alternative climatic regimes. Ecology 91:191–200

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Takeyama T, Okuda N, Yanagisawa Y (2007) Filial cannibalism as a conditional strategy in males of a paternal mouthbrooding fish. Evol Ecol 21:109–119

Tittensor DP, Mora C, Jetz W, Lotze HK, Ricard D, Vanden Berghe E, Worm B (2010) Global patterns and predictors of marine biodiversity across taxa. Nature 466:1098–1102

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 12:2635–2648

van Herwerden L, Choat JH, Newman SJ, Leray M, Hillersøy G (2009) Complex patterns of population structure and recruitment of *Plectropomus leopardus* (Pisces: Epinephelidae) in the Indo-West Pacific: implications for fisheries management. Mar Biol156:1595–1607

Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. J Biogeogr 27: 1153–1167

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. Philos Trans R Soc Lond B Biol Sci 360:1847–1857

Wares JP (2002) Community genetics in the Northwestern Atlantic intertidal. Mol Ecol 11:1131–1144

Waters JM, King TM, O'Loughlin PM, Spencer HG (2005) Phylogeographical disjunction in abundant high-dispersal littoral gastropods. Mol Ecol 14:2789–2802 Waters JM, McCulloch GA, Eason JA (2007) Marine biogographical structure in two highly dispersive gastropods: implications for trans-Tasman dispersal. J Biogeogr 34:678–687

Watts PC, Thorpe JP (2006) Influence of contrasting larval developmental types upon the population-genetic structure of cheilostome bryozoans. Mar Biol 149:1093–1101

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

Williams ST, Jara J, Gomez E, Knowlton N (2002) The marine Indo-West Pacific break: contrasting the resolving power of mitochondrial and nuclear genes. Integr Comp Biol 42:941–952

Wolanski E, Ridd P, Inoue M (1988) Currents through the Torres Strait. J Phys Oceanogr 18:1535–1545

APPENDIX TWO. Return of the ghosts of dispersal past: historical spread and contemporary gene flow in the blue sea star *Linckia laevigata*

Published in Bulletin of Marine Science 90: 399–425. 2014

Eric D. Crandall[1], Eric A. Treml[2], **Libby Liggins**[3], Lachlan Gleeson[3], Nina Yasuda[4], Paul H. Barber[5], Gert Wörheide[6] and Cynthia Riginos[3]

[1] Department of Ocean Sciences, University of California

[2] Department of Zoology, University of Melbourne

[3] School of Biological Sciences, The University of Queensland

[4] Department of Marine Biology and Environmental Sciences, University of Miyazaki

[5] Department of Ecology and Evolutionary Biology, University of California Los Angeles

[6] Department of Earth and Environmental Sciences, Division of Paleontology and Geobiology and GeoBio-Center LMU, Ludwig-Maximilians- Universität München

Abstract

Marine animals inhabiting the Indian and Pacific Oceans have some of the most extensive species ranges in the world, sometimes spanning over half the globe. These Indo-Pacific species present a challenge for study with both geographic scope and sampling density as limiting factors. Here, we augment and aggregate phylogeographic sampling of the iconic blue sea star, *Linckia laevigata* Linnaeus, 1758, and present one of the most geographically comprehensive genetic studies of any Indo-Pacific species to date, sequencing 392 base pairs of mitochondrial COI from 791 individuals from 38 locations spanning over 14,000 km. We first use a permutation based multiple-regression approach to simultaneously evaluate the relative influence of historical and contemporary gene flow together with putative barriers to dispersal. We then use a discrete diffusion model of phylogeography to infer the historical migration and colonization routes most likely used by *L. laevigata* across the Indo-Pacific. We show that estimates of genetic structure have a stronger correlation to geographic distances than to "oceanographic" distances from a biophysical model of larval dispersal, reminding us that population genetic estimates of gene flow and genetic structure are often shaped by historical processes. While the diffusion model was equivocal about the location of the mitochondrial most recent common ancestor (MRCA), we show that gene flow has generally proceeded in a step-wise manner across the Indian and Pacific Oceans. We do not find support for previously described barriers at the Sunda Shelf and within Cenderawasih Bay. Rather, the strongest genetic disjunction is found to the east of Cenderawasih Bay along northern New Guinea. These results underscore the importance of comprehensive range-wide sampling in marine phylogeography.

Introduction

Marine biodiversity is concentrated in the Indo-Pacific region, with species diversity reaching its highest values in the Coral Triangle, a region centered in the Indo- Malay-Philippines archipelago (Roberts et al. 2002, Carpenter and Springer 2005, Tittensor et al. 2010). The Coral Triangle has a complex geological history with much tectonic activity, including substantial reconfigurations of landmasses due to moving plates (Hall 2002). Fluctuating sea levels have also substantially restructured land and sea configurations, as the shallow continental shelf is extensive in this region (Voris 2000). For example, sea levels 18,000 yrs ago are thought to have been about 130 m lower than present day levels, resulting in greatly reduced area for most shallow marine habitats and thus severe reductions in local population sizes (Crandall et al. 2012a). At that time, the Makassar Strait was much narrower than it is today, almost completely blocking the marine connection between the Pacific and Indian Oceans (Chappell and Shackleton 1986). As sea levels rose over the next 10,000 yrs, continental shelf habitat of an area slightly smaller than the land area of the country of India (approximately 3.16 × 106 km^2) would have resubmerged and become avail- able again to local marine species (Voris 2000).

Perhaps unsurprisingly, phylogeographic studies have revealed the imprint of these sea level changes on population genetic patterns of Coral Triangle species (reviewed in Carpenter et al. 2011). some taxa show nearly reciprocal monophyly in mtDNA over relatively short distances, probably as a consequence of isolation during the Pleistocene (e.g., Barber et al. 2000, 2002). For many taxa, interestingly, the locations of likely vicariance due to Pleistocene sea level change are also associated with regions of persistent population genetic structure, most notably the Sunda Shelf (McMillan et al. 1999, DeBoer et al. 2008, Ackiss et al. 2013), Torres Strait (Mirams et al. 2011 – Appendix One), and Halmahera Eddy (Barber et al. 2006, 2011), although there is also evidence that this structure is being eroded by contemporary dispersal in some species (Gaither et al. 2011, Liu et al. 2012, DeBoer et al. 2014). Finally, nearly all loci and taxa that have been examined to date show signatures of Pleistocene-era population expansions onto newly submerged continental shelf habitats (e.g., Chenoweth et al. 1998, Lind et al. 2007, Crandall et al. 2008a,b, 2012a, Gaither et al. 2010).

Given the globally acknowledged value of the Coral Triangle in terms of marine biodiversity as well as the multitude of current threats to the region (Roberts et al. 2002, Burke et al. 2011), it is of practical importance to understand both the history of its marine communities (how and from where did genetic and species diversity arise? e.g., Renema et al. 2008, Williams and Duda 2008, reviewed in Bowen et al. 2013), as well as how the regional seascape is presently constructed (how are different parts of the region connected demographically by larval dispersal? reviewed in Riginos and Liggins 2013). Previous phylogeographic attempts to answer these questions have focused on the measurement of genetic structure (F_{ST} and its analogues, Wright 1950), and testing specific

hypotheses of population structure primarily with analyses of molecular variance (AMOVA, Excoffier et al. 1992). However, these approaches based on allele frequencies are poorly suited to organisms with evolutionarily high levels of gene flow (10–100 migrants per generation) and large coalescent effective population sizes (partially arising from high gene flow), which depress traditional estimates of population structure (such as F_{ST}, Hedrick 2005). Furthermore, because F_{ST} and AMOVA summarize a combined model of gene flow and effective population size (Whitlock 2011), these frequency-based approaches allow estimation of marine population structure only at a very coarse resolution, often resulting in substantial bias when parameters such as effective population size and timing of population divergence are not considered (Bird et al. 2011, Marko and Hart 2011, Faurby and Barber 2012). As an alternative, one might consider the inverse approach: testing explicit hypotheses of gene flow rather than genetic structure (Crandall et al. 2012b).

The substitution of gene flow for genetic structure as the parameter of interest in seascape genetics makes intuitive sense. Gene flow in most marine organisms is mediated by the planktonic larval stage, where millions of larvae disperse through a complex milieu of currents and environmental conditions. Thus, there are few impermeable barriers to larval dispersal in the ocean: for almost every physical or oceanographic entity that is thought to impede larval dispersal for some species (see examples in Rocha et al. 2007) one can find several species that show no evidence of isolation whatsoever (Lessios and Robertson 2006, Carpenter et al. 2011, Toonen et al. 2011). Although ocean currents and land masses may effectively act as barriers to gene flow, it is more appropriate to think in terms of probabilistic larval dispersal kernels for which the probability of a successful dispersal event (and therefore of gene flow) declines sharply with distance due to larval diffusion, behavior and mortality (Cowen et al. 2000, Gerlach et al. 2007, Buston et al. 2012). The most informed hypotheses of gene flow therefore come from modeling such dispersal through oceanographic current vectors while taking these additional factors into account (Kool et al. 2011, Treml et al. 2012). Empirical testing with genetic data has shown this to be the case using coalescent estimates of gene flow (Crandall et al. 2012b), assignment tests (Fievet et al. 2007), parentage-based tagging (Saenz-Agudelo 2012) or F_{ST} (Galindo and Palumbi 2006, White et al. 2010, Alberto et al. 2011, Foster et al. 2012). Until recently, gene flow estimates have been constrained to the unrealistic assumptions of Wright's island model by relying on the F_{ST} summary statistic (Whitlock and McCauley 1999), which assumes equal levels of gene flow throughout the sampled area and does not take historical factors (such as lineage sorting) into account. However, coalescent modeling approaches now allow flexible evaluation and selection of specific models of gene flow, which are estimated simultaneously with the genealogical history of genetic sequence data (Hey and Nielsen 2007, Lemey et al. 2009, Beaumont 2010, Beerli and Palczewski 2010).

A further challenge to understanding genetic diversity in the Coral Triangle arises from its location: it exists near the junction of the Indian and Pacific Oceans, at the center of the Indo-Pacific region, which is the largest biogeographic region on Earth (Spalding et al. 2007). Many species found

in the Coral Triangle have vast ranges that may include large portions of the Indian and Pacific Oceans. Although there have been considerable recent efforts to document genetic patterns within the Coral Triangle, there are few studies with dense sampling (many locations, many individuals per location) that include both the Coral Triangle and surrounding regions of the Indo-Pacific (see Keyse et al. 2014). Although logistically challenging, large-scale geographic coverage is necessary to determine the context of genetic variation of marine species. Without broad-scale sampling that includes both the Coral Triangle and other parts of the species' range, it is not possible to completely resolve the extent of divergent genetic lineages (Manel and Holdregger 2013).

The sea star *Linckia laevigata* Linnaeus, 1758, easily recognizable for its striking blue coloration, is one of the best-studied species in the Indo-Pacific region. Besides the well-known royal blue phenotype, several color variations are reported from different geographic regions without apparent differentiation by morphotype (Williams 2000). Like many other marine benthic taxa, adult *L. laevigata* are sedentary, but the larvae have a moderate (at least 22 d before metamorphosis, Yamaguchi 1973) pelagic larval duration (PLD) after external fertilization.

Early genetic surveys of *L. laevigata* were based on allozymes (Williams and Benzie 1993, Williams and Benzie 1996, 1998, Williams et al. 2002) or mtDNA (Williams and Benzie 1997, 1998, Williams et al. 2002, Crandall et al. 2008b, Kochzius et al. 2009) and have had broad coverage from the western Pacific to the Indian Ocean without much sampling in the Coral Triangle (e.g., Williams and Benzie 1998) or have exclusively focused sampling within the Coral Triangle (Crandall et al. 2008b, Kochzius et al. 2009). At the broadest scale, support was found for differentiation between the Indian and Pacific Oceans based on allozymes, albeit with individuals from Western Australia (Ningaloo) showing greater affinity to western Pacific populations than to western Thailand and South Africa (Williams and Benzie 1998, Williams et al. 2002, see also Vogler et al. 2013 for similar results in crown-of-thorns starfish). For mtDNA COI sequences, Williams et al. (2000) described two major clades: an "Indian Ocean" clade, which included both Pacific and Indian Ocean individuals and a "Pacific Ocean" clade that contains only Pacific Ocean individuals with the exception of a few western Australian individuals. Crandall et al. (2008b) as well as Kochzius et al. (2009) sampled comprehensively within the Coral Triangle and found that the Indian clade haplotypes dominate most Indonesian populations, declining in frequency from Aceh in the west to Jayapura in the east. Within the Coral Triangle, the greatest population structure was found by grouping Aceh with Krakatau against a cluster of remaining locations, whereas less support was found for a west-east delineation defined by the Sunda Shelf, and there was modest support for distinctiveness of Teluk Cenderawasih (Crandall et al. 2008b). Kochzius et al. (2009) suggested that *L. laevigata* has historically expanded populations into the western Pacific from eastern Indian Ocean origins, a route of colonization running counter to the Indonesian Through Flow (Kochzius et al. 2009).

In the present study we combine mtDNA data from some of the previous studies cited above with new sampling from 18 additional locations to evaluate the influence of putative historical and contemporary gene flow and spatial features on genetic patterns within *L. laevigata*. Although data from additional loci are desirable for future studies, the rich genealogical information in this mtDNA-only data set allows inference from the recent (approximately 100 kya) matrilineal history of this species and the cross compatibility of this DNA sequence data facilitates synergism across research groups (Bowen et al. 2014). The data set analyzed here represents the single most extensive population genetic survey both in geographic extent (diameter >14,000 km) and density for any species to date from the Indo-Pacific region (Keyse et al. 2014); its compilation was only possible by cooperation and collaboration among research groups.

We use this data set to consider the genealogical history of *L. laevigata* in terms of inferred mitochondrial gene flow in addition to estimates of genetic structure. We begin by parameterizing a biophysical model of contemporary larval dispersal to predict mean dispersal distances for *L. laevigata* larvae among sampling sites. We then use a multiple regression approach based on permutation (Legendre et al. 1994) to ask whether measurements of genetic structure in this species are better explained by the mean dispersal distances from the biophysical model, or by geographic distances. The biophysical model assesses whether mitochondrial patterns are best explained by equilibrium gene flow from contemporary larval dispersal whereas the geographic distance model assesses the relative importance of colonization processes and historical gene flow. We also use the biophysical model output to define modular geographic regions with greater larval connectivity within each region than between regions. These regions are then used in a phylogeographical model that reconstructs historical gene flow through time by treating each region as a character state that can be inferred for each node on the genealogy through ancestral state reconstruction (Lemey et al. 2009). This approach allows us to infer the vectors of gene flow that were important in the spread of this lineage throughout the Indo-Pacific region.

Methods

Study sites and COI sequencing

Mitochondrial cytochrome oxidase I sequences from blue morph *L. laevigata* were obtained from both newly sampled individuals and from previous studies (Williams 2000, Crandall et al. 2008b). New samples were obtained from locations in western and eastern Australia, Timor-Leste, Papua New Guinea, the Solomon Islands, Vanuatu, Fiji, and Tonga (see Fig. 1 and Table 1 for more details), using tube feet preserved in ethanol. New sequences were also obtained from locations in Thailand, Indonesia, Guam, and New Caledonia using old pyloric caecum samples preserved in DMSO, which were previously analyzed in an allozyme study (Williams et al. 2002). DNA preparation was undertaken at the University of Queensland (including NIN, ASH, TIM, KAV, MVO, MOT, LIZ, HER, MOO,

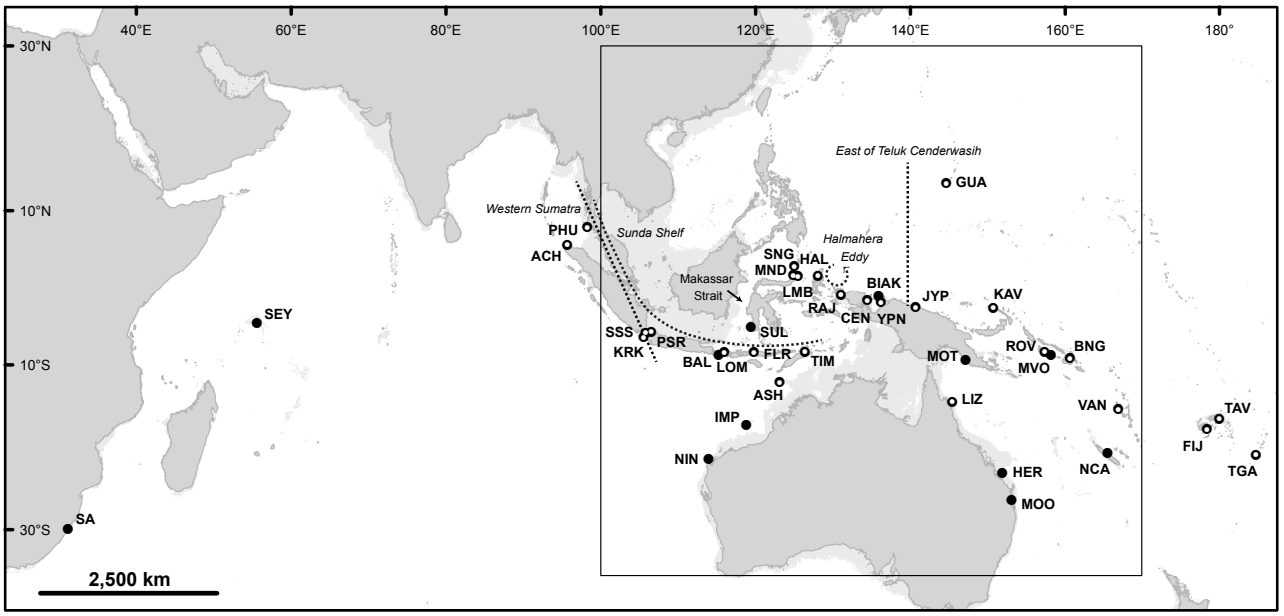


Figure 1. Sampling locations; see Table 1 for abbreviations. Open circles represent locations with sample size >10. The domain for the biophysical model is represented by a box. Four putative barriers to gene flow that were evaluated under the MRDM and AMOVA approaches are denoted as dotted lines.

Table 1. Sampling locations and summary statistics. Number of samples (*n*), Haplotype diversity (H), percent segregating sites (S), nucleotide diversity (π), Tajima's *D*, and Fu's F_s.

Location	CODE	Latitude	Longitude	Source	*n*	H	S (%)	π (%)	D	F_s
South Africa	SA	-30.1	31.2	Williams 2000, 2002	5	1.00	1.8	0.90	0.50	-1.90
Seychelles	SEY	-4.6	55.6	Williams 2000, 2002	4	0.25	0.0	0.00	0.00	0.00
Ningaloo Reef, Australia	NIN	-21.7	114.0	Present study: Riginos	7	0.86	2.8	1.20	0.44	-1.19
Imperieuse Reef, Australia	IMP	-17.5	118.8	Williams 2000, 2002	3	1.00	2.0	1.40	0.00	0.46
Ashmore Reef, Australia	ASH	-12.2	123.1	Present study: Riginos	15	0.80	6.1	1.90	-0.04	-3.12
Timor-Leste	TIM	-8.3	126.4	Present study: Riginos	19	0.68	4.8	1.10	-0.76	-4.90*
Aceh, Indonesia	ACH	5.6	95.7	Crandall et al. 2008	15	0.73	4.3	1.00	-0.96	-4.31**
Phuket, Thailand	PHU	7.9	98.3	Present study: Yasuda	28	0.75	6.9	1.00	-1.59	-15.47
Krakatau	KRK	-6.1	105.5	Crandall et al. 2008	48	0.65	9.4	1.00	-1.79	-25.36
Sebesi/Sebuku/Sangiang	SSS	-5.9	105.5	Crandall et al. 2008	51	0.51	9.4	1.40	-1.19	-11.13**
Palau Seribu	PSR	-5.7	106.6	Crandall et al. 2008	79	0.48	10.7	1.40	-1.19	-22.60
South Sulawesi	SUL	-5.1	119.4	Crandall et al. 2008	7	0.86	3.8	1.50	-0.28	-0.84
Bali	BAL	-8.7	115.3	Crandall et al. 2008	5	0.80	3.6	1.70	0.09	0.98
Lombok	LOM	-8.4	116.0	Crandall et al. 2008; Present study: Yasuda	23	0.70	4.3	1.20	0.05	-7.10***
Flores	FLR	-8.4	119.8	Crandall et al. 2008	14	0.79	4.8	1.40	-0.39	-3.53*
Manado	MND	1.6	124.9	Crandall et al. 2008	76	0.43	10.5	1.30	-1.21	-16.03***
Lembeh	LMB	1.5	125.2	Crandall et al. 2008	20	0.60	6.1	1.40	-0.70	-2.25
Sangihe	SNG	2.8	125.4	Crandall et al. 2008	17	0.47	3.1	0.80	-0.48	-1.20
Halmahera	HAL	1.5	128.0	Crandall et al. 2008	75	0.41	11.2	1.30	-1.37	-13.45***
Raja Ampat	RAJ	-0.9	131.1	Crandall et al. 2008	31	0.42	5.4	1.20	-0.39	-1.85
Teluk Cenderawasih	CEN	-1.7	134.5	Crandall et al. 2008	22	0.68	5.9	1.60	-0.10	-4.34*
Biak	BIAK	-1.1	136.0	Crandall et al. 2008	7	0.86	4.8	1.90	-0.24	-0.38

Table 1. continued

Yapan	YPN	-1.9	136.2	Crandall et al. 2008	19	0.42	3.8	0.90	-0.79	-0.62
Jayapura	JYP	-2.5	140.7	Crandall et al. 2008	19	0.58	4.1	1.10	-0.23	-2.54
Guam	GUA	13.5	144.7	Present study: Yasuda; Williams 2000, 2002	25	0.68	5.4	1.30	-0.21	-6.82**
Kavieng, PNG	KAV	-2.6	150.8	Present study: Riginos	16	0.69	4.1	1.30	0.06	-2.97
Roviana, SOL	ROV	-8.3	157.4	Present study: Riginos	18	0.72	5.9	1.30	-0.91	-4.46*
Marovo, SOL	MVO	-8.8	158.3	Present study: Riginos	5	1.00	2.8	1.30	-0.38	-1.35
Boneagi, SOL	BNG	-9.2	160.7	Present study: Crandall	10	0.60	2.8	1.10	0.59	0.09
Motupore, PNG	MOT	-9.4	147.2	Present study: Riginos	8	0.88	3.6	1.30	-0.11	-1.78
Lizard Island, GBR	LIZ	-14.7	145.5	Present study: Riginos	16	0.81	6.6	1.50	-0.92	-4.75*
Heron Island, GBR	HER	-23.5	151.9	Present study: Riginos	9	0.89	5.1	1.60	-0.70	-2.10
Mooloolaba, QLD	MOO	-26.6	153.1	Present study: Riginos	3	1.00	0.8	0.5	0.00	-0.69
New Caledonia	NCA	-21.0	165.6	Present study: Yasuda	5	0.80	2.0	1.00	0.29	0.13
Vanuatu	VAN	-15.6	167.0	Present study: Crandall	16	0.63	5.4	1.10	-1.31	-2.35
Viti Levu, Fiji	FIJ	-18.1	178.4	Present study: Crandall	20	0.40	4.1	1.00	-0.39	0.08
Tavenui, Fiji	TAV	-16.8	180.0	Present study: Crandall	10	0.80	3.8	1.50	0.37	-1.62
Tonga	TGA	-21.2	-175.3	Present study: Riginos	12	0.75	2.6	0.80	-0.42	-4.06*

*P < 0.05; ** P < 0.01; *** P < 0.001

TGA, ROV; see Table 1 for abbreviations), Boston University (including ROV, BNG, VAN, FIJ, and TAV) and Ludwig-Maximilians-Universität München (including LOM, PHU, NCA, and GUA). At the University of Queensland, genomic DNA was extracted using a modified salt extraction protocol (Aljanabi and Martinez 1997). The cytochrome oxidase I (COI) gene was amplified using polymerase chain reaction (PCR), with each reaction containing approximately 1 µl of DNA and using Titanium Taq polymerase (Clontech Laboratories, Inc.) and the universal COI primers from Folmer et al. (1994). Amplicons were purified with an exo/sap procedure (New England Biolabs) and sent to Macrogen (Korea) for capillary sequencing. At Boston University, sequencing protocols followed those outlined in Crandall et al. 2008b. At Ludwig-Maximilians-Universität München, newly developed *L. laevigata*-specific primers (LL-F1, 5′–ACCACCGGCTGGGTCGAA–3′ and LL-r1, 5′– TAATCTTTGGGGCGTGAGC–3′) were used for PCR and sequencing for three populations (PHU, GUA, and NCA), to increase PCR efficiency. The amplifications were made in 10 µl reactions with a final concentration of 3mM of MCl_2, 1× buffer, 0.3 µM of each primer, 0.8 mM of DNTP, 0.07 units of Kapa taq DNA Polymerase (Kapa Biosystems) and 1 µl of DNA template. Thermocycling consisted of denaturation of DNA at 94 °C for 3 min, 35 cycles of 94 °C for 30 s, 50 °C for 30 s and 72 °C for 30 s, followed by final extension of DNA at 72 °C for 7 min and cooling at 4 °C. The amplification of PCR products was confirmed by gel electrophoresis in a 1% agarose gel. Sequencing reactions were conducted on an ABI 3730 capillary sequencer using BigDye v.3.1. sequence was determined from both directions using LL-F1 and LL- r1 primers. trace files were manually checked and edited using CodonCode v3.0.2. These newly generated sequences were aligned against published sequences from Williams (2000), Williams et al. (2002), and Crandall et al. (2008b). Individual haplotypes from Crandall et al. (2008b) were assigned in the same manner as the original paper.

In total, we aggregated mtDNA haplotypes from 38 localities (Fig. 1) and data from all locations were represented in the haplotype network (Fig. 2). to obtain an accurate and unbiased estimate of pairwise genetic relationships between locations (Φ_{ST}, F_{ST}, and D_{est} with respect to Euclidean distance and overwater distances), we included only those locations which had at least 10 individuals sampled, leaving 26 populations for these analyses (Fig. 1, sites with white centers). Those locations within the domain of the biophysical model (box in Fig. 1) were used for two purposes. First, those sample sites with more than 10 individuals sampled, that also contained substantial reef habitat were used for reanalyzing the pairwise genetic relationships between locations with respect to dispersal distances derived from the biophysical model (20 total sites; all white sites within box in Fig. 1 excluding Jayapura, which does not have substantial reefs, and is thus not included in the biophysical model). second, all collection locations within the model domain were used, except Jayapura and Mooloolaba, due to reef representation to extract the geographic cluster membership of the remaining 29 sample locations.

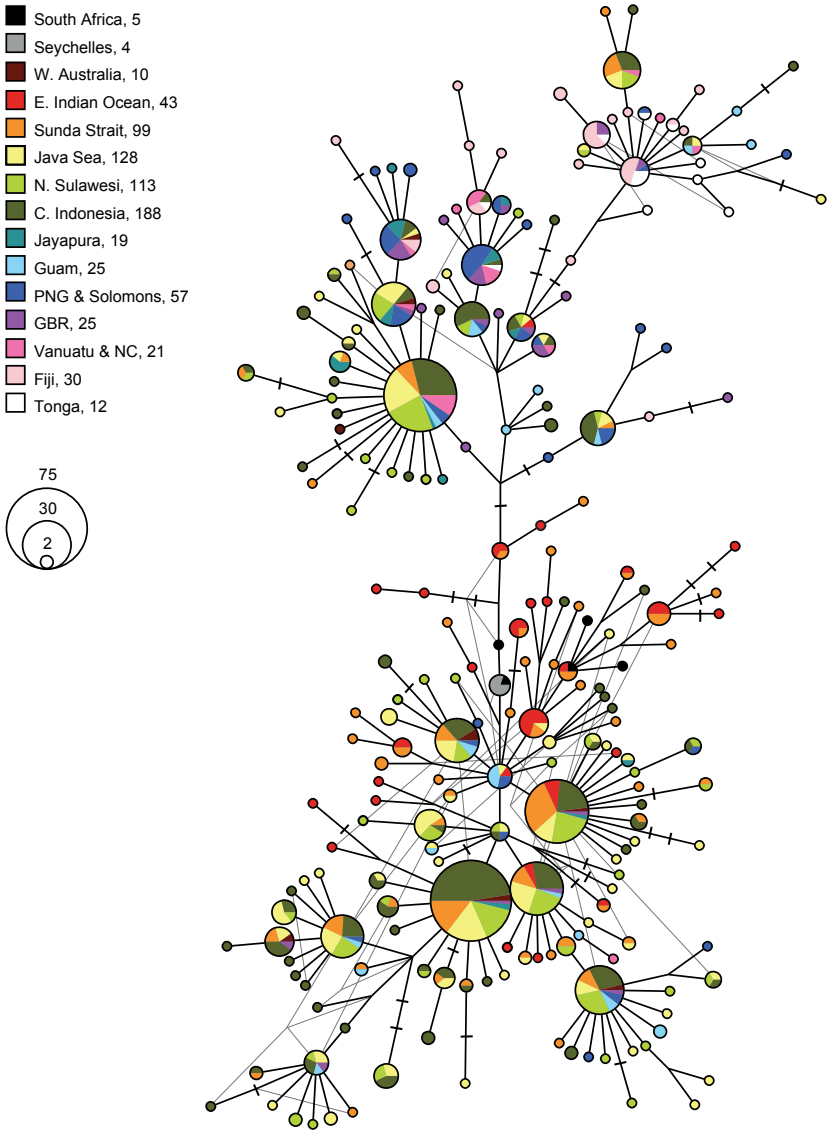


Figure 2. Median-joining haplotype network for *Linckia laevigata* colored according to modular clusters determined from the biophysical model. The number of individual sequences included in each cluster follow the name in the key. The frequency of each haplotype is indicated by size (see key, bottom right). Edges between haplotypes or small cross-bars indicate a mutational step. Black edges represent one of the maximum parsimony networks chosen at random; grey edges represent alternate relationships among haplotypes found in 29 other equally parsimonious networks.

Genetic diversity and differentiation

To visualize the total COI diversity, a median joining haplotype network was constructed in Network 4.611 and edited in Network Publisher 2.0 (fluxus-engineering.com, Bandelt et al. 1999). To reduce complexity, non-parsimonious links were deleted using the maximum parsimony calculation option (MP, Polzin and Daneshmand 2003). Colors across the haplotype network represented different biophysically derived clusters (see below) and other locations/regions beyond the scope of the biophysical model, but assumed to be distinct demographically due to their geographic isolation, and significant genetic structure (Online Table S1). Standard population summary statistics were calculated in Arlequin v3.5.1.3 (Excoffier and Lischer 2010). Statistics calculated included Watterson's θ (Watterson 1975), which estimates the average number of polymorphic sites, and π, which estimates the average number of differences between two random sequences from the same population. Tajima's D (Tajima 1989) and Fu's F_s (Fu 1997) were also measured to compare *L. laevigata* COI diversity against neutral, equilibrium expectations.

We used a variety of approaches to summarize genetic differentiation. Pairwise relationships among populations (for sampling locations including 10 individuals or more: 26 populations) were described by Φ_{ST} (based on the Tamura-Nei distance as selected by jModeltest; Posada 2009) as well as F_{ST} (based on haplotype identities) in Arlequin. To reduce the effects of high allelic variability, we also used D_{est} (Jost 2008) calculated in Genodive (Meirmans and Van Tienderen 2004) by reducing infiles to haplotype identity. Relationships among populations based on these pairwise statistics were visualized using non-metric multidimensional scaling (vegan package in R, Oksanen et al. 2012).

We used AMOVA (Excoffier et al. 1992) analysis in Arlequin to test *a priori* predictions about various spatial configurations of genetic differentiation arising from barriers to dispersal. To test the hypothesis that divergence between Indian and Pacific populations explains the most molecular variance in COI, we tried two different spatial configurations. We first separated locations in western Indonesia from central Indonesia (following Crandall et al. 2008b), and then delineated Indian and Pacific populations along the Sunda Shelf. The contribution of the Halmahera Eddy to population structure was assessed by configurations whereby central Indonesia locations were grouped separately from Cenderawasih locations, and these groupings were contrasted against configurations that combined central Indonesia, and Cenderawasih locations as a single group. Similarly the distinctiveness of Cenderawasih from the Pacific was evaluated by comparing separate and combined groupings. Furthermore, Guam was allowed to group by itself and with other Pacific locations due to preliminary analyses pointing to genetic distinctiveness of this location. These varying spatial configurations were evaluated using AMOVA based on Φ-statistics derived from Tamura-Nei distances between haplotypes and from haplotypes identified in Arlequin (as above). For AMOVAs based on haplotype identity, we also used F'- statistics (Meirmans and Van Tienderen 2004).

Geographic and biophysically informed predictions of gene flow among sample locations

Three pair-wise distance metrics were calculated as proxies for gene flow (i.e., realized dispersal over many generations) among sampled populations. As a simple null model of gene flow, we calculated the Euclidean distance among all 38 sampled locations. Although this model of gene flow implies that dispersal occurs relative to straight-line routes only, it is a common dispersal distance proxy used in marine population genetic studies (Riginos and Liggins 2013). The second dispersal distance proxy used for all locations was the shortest over- water distance calculated with a least-cost path algorithm. The distance calculations are similar to the Euclidean distance, but the least-cost path was forced around all land boundaries. These two measures of geographic distance are expected to explain more of the variance in genetic structure when historical processes such as colonization dominate the signal from gene flow (Selkoe 2008).

Finally, we used a biophysical model of larval dispersal (Treml et al. 2012) to quantify the relative dispersal strength among sampled populations within the model domain (Fig. 1). This dispersal model includes coral reef habitat (Spalding et al. 2001), oceanographic data describing sea surface currents for three years (ROMs, Wang et al. 2005), and several biological parameters describing the dispersal characteristics of *L. laevigata*: seasonal spawning periodicity, a 25-d maximum pelagic larval duration, a 1- to 2-d precompetency period, strong late-stage swimming/homing behavior, and a 30% d^{-1} larval mortality (Yamaguchi 1973). The model outputs the probability that larvae released in one location survive and settle in every other recipient location, summarized as a 1002 × 1002 source-reef by destination-reef matrix. This dispersal probability matrix was converted to a migration matrix representing the proportion of settlers to every reef patch that came from all upstream larval sources. See Treml et al. (2012) for model details and sensitivity analysis. The migration matrix, M, was converted to "oceanographic dispersal distance" using $\log(M^{-1})$ to transform the values to be the same rank-order as geographic distance (high proportion of settlers then have a short distance) required for many network-based algorithms. This inverse dispersal strength matrix was used as a proxy for dispersal distance, and is referred to as such throughout this paper. This oceanographic dispersal distance is expected to explain more of the variance in genetic structure when contemporary dispersal events dominate the signal relative to historical connections (White et al. 2010).

To identify the emergent geographic clustering of reef habitat (and sample locations) determined by the dispersal strengths represented in the migration matrix, we used a network-based leading eigenvector community detection algorithm (Newman 2006). This algorithm identifies the optimal clustering within a network by optimizing the network's modularity, or simply maximizing the density of within-cluster connections while minimizing between-cluster connections. The original asymmetric migration matrix was converted to a symmetric matrix by taking the maximum dispersal strength between all pairs of reefs. *Linckia laevigata* sample sites were overlaid with the network clustering results thereby revealing the potential clustering of sample sites based on dispersal

potential among all reefs. These spatial clusters were then used to aggregate sampling sites in the phylogeographic diffusion model described below.

Evaluating genetic differentiation using dispersal proxies

We used a multiple regression on distance matrices (MRDM, Legendre et al. 1994) to evaluate the relative influence geographic distance and multiple discrete landscape factors that might contribute to population genetic structure; this methodology performs favorably compared to many other methods (Balkenhol et al. 2009). In a simple model where the linear relationship between geographic distance and genetic distance are evaluated, MRDM is equivalent to a Mantel test. Dispersal distance proxies included Euclidean distances and overwater distance, as well as the biophysical dispersal distances between populations. The biophysical distances represent a hypothesis of contemporary dispersal. Euclidean and overwater distances could represent simpler (null) models of contemporary dispersal, but they also might capture historical averages of dispersal (that is both recurrent gene flow and colonization). The predictive contributions of four putative barriers/divisions were also evaluated including (1) Western Sumatra, as found in Crandall et al.'s (2008) earlier survey of *L. laevigata*; (2) Sunda Shelf, a focal point of Pleistocene marine disjunctions due to the expansion of the Sunda Shelf land mass at low sea level stands; (3) Halmahera Eddy, a contemporary hydrodynamic barrier; and (4) Pacific east of Cenderawasih, which might represent a location of habitat limitation. For barriers, dummy variables (0 vs 1) were coded with 0 for population pairs found on the same side of the putative barrier. Φ_{ST}, F_{ST}, and D_{est} values between pairs of populations were used as response variables. All variables were normalized and both forward and backward model selection were implemented as in Legendre et al. (1994). Models including Euclidean and overwater distances included all populations with sample size ≥10 (26 populations). Because biophysical distances were only available for some population pairs, the analyses involving these predictors was reduced to a more restricted subset of populations (20 populations). In addition, the decomposed matrices involving biophysical predictors were twice as large because the distances are asymmetric (distance from *X* to *Y* does not equal *Y* to *X*). Permutated probabilities for all matrix regression models were evaluated with lmPerm (Wheeler 2010).

Bayesian selection of important migration parameters

To estimate the historical gene flow required to explain the current distribution of mitochondrial genetic variation in *L. laevigata*, we modeled the phylogeographic history of COI as a discrete diffusion process following methods developed by Lemey et al. (2009). Using a coalescent perspective, we assigned a geographic location as a discrete character trait having one of 14 possible states (based on spatial clusters delineated by the biophysical model, see below) to each COI sequence, and reconstructed the most probable location of each ancestor back to the most recent common ancestor (MRCA) of the entire sample using BEAST 1.6.2 (Drummond and Rambaut 2007). Changes in location between ancestor and descendent nodes were modeled as a migration event, the probability of which was governed by a time-reversible matrix of migration rates among locations (similar to the

GTR model used for models of molecular evolution). To reduce the degrees of freedom in the matrix, this method uses Bayesian stochastic search variable selection to only allow *W* of the rates to be non-zero, where the prior on W is a truncated Poisson distribution with mean η (see Lemey et al. 2009 for full details). In a Bayesian framework, the geographic location of the MRCA, as well the migration events required to explain most topologies can be estimated simultaneously with models of nuisance parameters such as those for molecular evolution, demography and tree topology.

Because our sequences came from 38 different localities, many with relatively small sample sizes, we assigned their geographic location as one of the k = 9 spatial clusters resulting from the clustering algorithm on the migration matrix plus the following groups of sites based on geographic isolation: Fiji/Taveuni, Tonga, Phuket/ Aceh, Seychelles, and South Africa, for a total of 14 clusters (Fig. 3). We used a TN93 model of molecular evolution and an uncorrelated relaxed clock (Drummond et al. 2006) to model COI sequence evolution. Because this species and most other Indo- Pacific species appear to have a history of demographic fluctuation over evolutionary time, we implemented a Bayesian skyline model of demography as well (Drummond et al. 2005), which relaxes assumptions of any particular demographic history. We set a fairly uninformative truncated Poisson prior for the number of allowable migration rates within the matrix with an offset of k – 1 = 13 (the minimum number of rates required to connect all populations) and a mean of 10 (95% of the prior probability mass lies between 5 and 21 rates). All other priors were set to their defaults, and we elected not to use distance-informed priors on gene flow parameters, so that the model had no *a priori* information about the underlying geography. The model was run four times for at least 50 million steps of Markov chain Monte Carlo (MCMC), and convergence was checked in tracer 1.5. We then combined the logfiles and treefiles after trimming off an appropriate amount of burn in, and constructed a maximum clade credibility tree, from which we extracted the ancestral location probabilities. We established the significance of migration rate parameters if they were supported by a Bayes Factor of 3, which also corresponded with the migration rate parameter being required to explain >50% of sampled topologies.

Because MRCA location probability might be biased toward spatial clusters with a large sample size, we evaluated the prior distribution for root location by randomly swapping the location state among sequences during the MCMC chain such that sample size from each spatial cluster remained the same, but the sequences assigned to each location becomes random. This randomization, performed with the tip state operator in BEAST 1.7.5, as described by Edwards et al. (2011), removed location information from the dataset, allowing us to observe the prior expectation for root location.

Given that single-locus inference is still common for Indo-Pacific species (see discussion in Bowen et al. 2014), the genealogical approaches used here make much more effective use of the high information content in the mitochondrial locus than do estimates of genetic structure. Moreover, F_{ST} methods as well as more explicit population genetic models such as Migrate (Beerli and Felsenstein

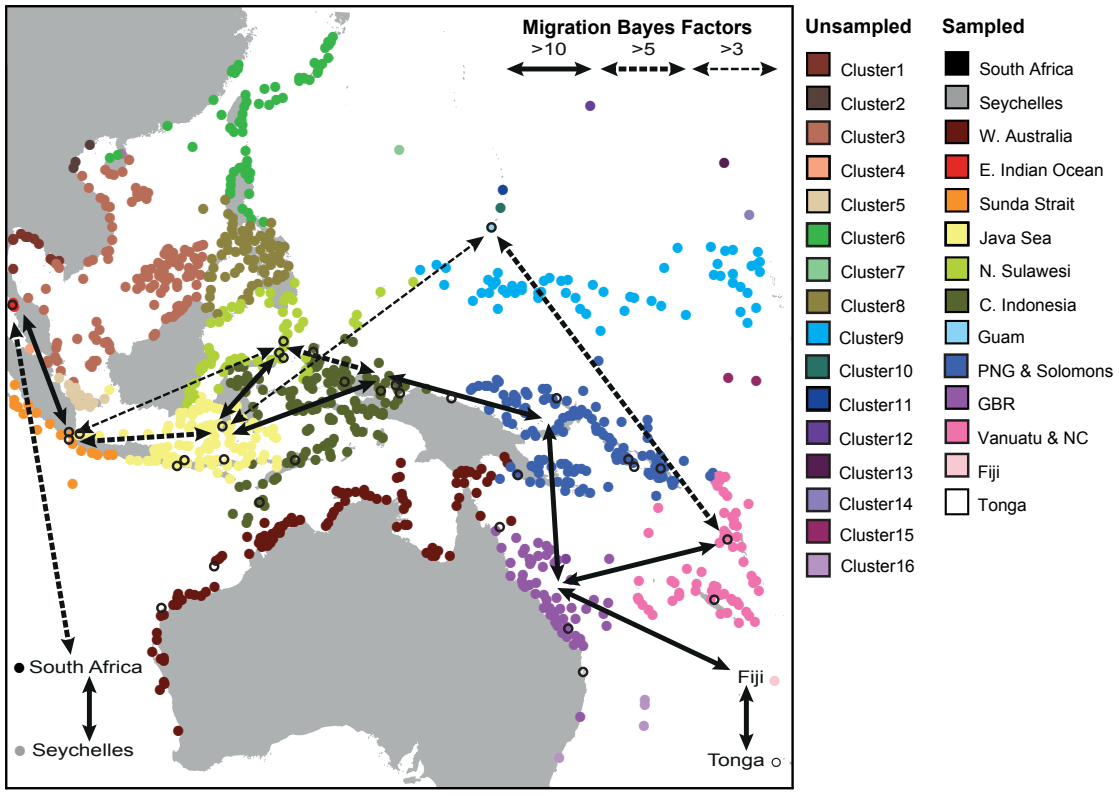


Figure 3. Best supported migration routes (Bayes Factor > 3) among modular population clusters delineated by the biophysical model. Optimal clusters were identified using the network modularity algorithm, and are shown in different colors. Each colored point represents the geographic centroid of reef patches used in the biophysical model. Open circles show genetic sample locations, as in Fig. 1.

2001) make an implicit assumption of genetic equilibrium (i.e., haplotypes have maintained the observed distribution for a long time about 1/2 N_E generations) that is violated by the dynamic changes in the marine habitats of the Coral Triangle over the past hundreds of millennia. The spatial diffusion approach is not completely parameterized as a population genetic model (although Θ is still estimated as part of the skyline model) so it cannot make estimates for the amount of gene flow among populations as is done in Migrate and IMa. However, the absence of population genetic parameters allows a more flexible model that can reconstruct the most likely location of the mitochondrial common ancestor, together with the avenues of colonization and gene flow among multiple populations without prior knowledge of population history (unlike IMa2, Hey and Nielsen 2007; see also Bloomquist et al. 2010). As with all coalescent methods, the addition of sequence data from nuclear loci will eventually allow us broaden our inference to the demographic history of the species, rather than just that of the maternal lineage.

Results

Study locations and COI sequencing

We obtained mtDNA sequences from a total of 791 *L. laevigata* individuals including 274 new sequences that greatly expanded the geographic scope of population sampling, especially from the western Pacific (Table 1, Fig. 1). To avoid using too much missing data by nucleotide position (<5%), we trimmed our alignment to 392 bp in length, containing a total of 110 substitution sites, and this portion of COI was used for subsequent analysis. This resulted in 209 unique haplotypes. New sequences have been deposited in GenBank (Accession Numbers KF834572–KF834833) and a complete Fasta-formatted file of the 791 individuals is deposited in http://www.datadryad.org.

Genetic diversity and differentiation

Thirty equally parsimonious haplotype networks were recovered (one network chosen at random is shown in Fig. 2). The haplotype network highlights the high diversity of haplotypes and presence of many unique or private haplotypes. consistent with previous mtDNA sequencing surveys (i.e., Williams 2000, Crandall et al. 2008b, Kochzius et al. 2009), there were two large emergent clusters. The haplotypes of sampled locations in the West and central Pacific were largely restricted to one cluster (i.e., Vanuatu, New Caledonia, Tonga and parts of the Solomons, Fiji, and Papua New Guinea; top cluster, Fig. 2.) but not entirely (e.g., the sample from Taveuni in Fiji comprised individuals from both clusters). In contrast, locations sampled in the Indian Ocean had haplotypes that fell within the other cluster (bottom cluster, Fig. 2) or were intermediary to both clusters. Many haplotypes were shared across many regions, and especially within the central Indo-Pacific locations, haplotypes from both clusters were common.

Individual populations varied in observed mtDNA diversity (Table 1) with haplotype diversity ranging from 0.42 to 0.80 in populations with n > 10. Several populations showed deviations from

neutral equilibrium conditions especially as evaluated by Fu's (1997) F_s statistic. Among the 26 populations where 10 or more individuals were sampled, there was significant genetic differentiation among many population pairs regardless of the statistic used (Online Table S1). Non-metric multidimensional scaling (NMDS) based on F_{ST} values with Tamura-Nei distances among haplotypes recovered relationships among populations that roughly approximated geography (Fig. 3; non-metric stress = 0.045). Dimension 1 of the NMDS was very strongly correlated with longitude ($R^2 = 0.81$, $P < 1 \times 10{-}9$), while there was no correlation of NMDS dimension 2 to latitude ($R^2 = 0.01$, $P = 0.58$).

All population groupings in hierarchical AMOVAs resulted in significant Φ_{CT} and F'_{CT} values, indicative of substantial regional population structure. The population grouping that consistently returned the highest Φ_{CT} and F'_{CT} values was a two-regional grouping whereby Pacific populations (all populations east of Cenderawasih Bay) were delineated from the remaining populations including those from Cenderawasih, the central Indo-Pacific, and the Indian Ocean populations ($\Phi_{CT} \leq 0.282$, $P < 0.001$; $F'_{CT} \leq 0.637$, $P < 0.038$). For all groupings, Φ_{SC} values were also significantly greater than zero ($P < 0.001$) and F'_{SC} were marginally significant ($P \leq 0.06$), indicative of genetic differentiation among populations within regions (Online Table S2).

Dispersal distance proxies

The bivariate correlation between dispersal distance matrices, evaluated with the simple Mantel test, revealed significant relationships for 20 localities within the Coral Triangle model domain. For the 190 site-pairs, the correlations between all distance-based dispersal proxies were high. The correlation coefficient for Euclidean distance and overwater distance was $r_M = 0.996$ (10,000 permutations, $P < 0.0001$). For Euclidean distance and mean dispersal distance, $r_M = 0.896$ (10,000 permutations, $P < 0.0001$). For overwater distance and mean dispersal distance, $r_M = 0.902$ (10,000 permutations, $P < 0.0001$).

The leading eigenvector community structure algorithm revealed 25 groups across the Coral Triangle (highlighted by different colors in Fig. 3), with strong consistency among dispersal distance metrics, thresholds, and community detection algorithms (not shown). Nine of these groups contained sampling sites for which we had more than 10 samples.

Distance as a predictor of gene flow and differentiation

Using the multiple regression on distance matrices (MRDM) approach to evaluate predictors of genetic structure showed strong positive relationships between both Euclidean and overwater distances and either Φ_{ST} or D_{est} (R_M^2 ranged from 0.46 to 0.68, for the 26 population comparisons), consistent with an isolation-by-distance pattern. The relationship between F_{ST} and distance was substantively lower ($R_M^2 < 0.03$). Due to the collinearity of Euclidean distance and overwater distance ($R_M^2 = 0.99$ for 26 localities) and overwater distance and dispersal distance ($R_M^2 = 0.90$ for the 20 localities) optimal full models retained only one of these distances. Both forward and backward model selection converged on the same linear model for Φ_{ST} or D_{est}. For Φ_{ST}, the best model contained both

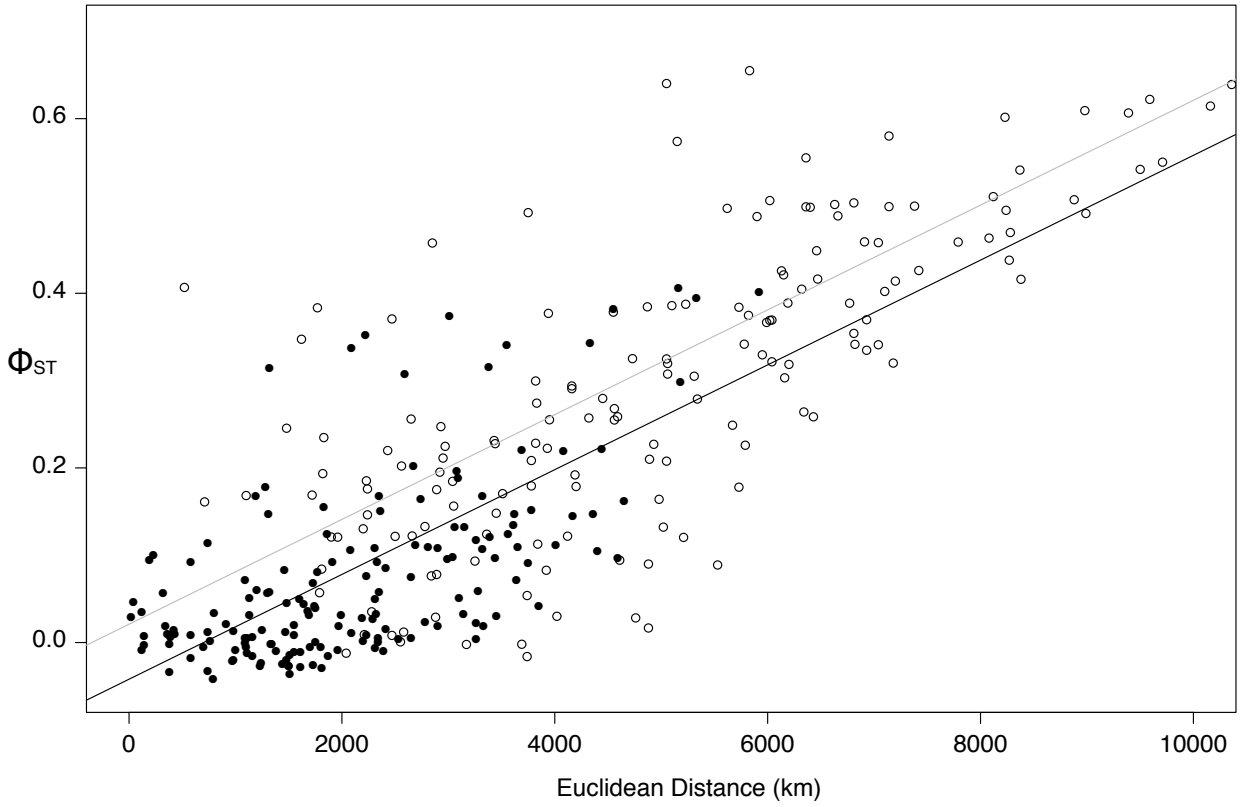


Figure 4. Best linear model of genetic differentiation. Pairwise Φ_{ST} values by Euclidean distance and showing the effect of the division between Indian and Pacific Oceans east of Cenderawasih (R^2 = 0.69, P < 0.001). Grey points represent population pairs including both Indian and Pacific Ocean populations whereas black points represent population pairs within either ocean. Gray and black lines represent the regression lines for between and within ocean comparisons.

Euclidean distance and a barrier to the east of Cenderawasih Bay (Fig. 4: $R^2 = 0.69$, $P < 0.001$), and for D_{est}, the best model contained overwater distance, the barrier to the east of Cenderawasih Bay, and the West Sumatra delineation ($R^2 = 0.50$, $P < 0.001$). For F_{ST}, the best model contained only Euclidean distance but did not explain much variance ($R^2 = 0.03$, $P = 0.014$). With the more restricted set of 20 populations for which we had mean dispersal distance predictions, overwater distance was better predictor of genetic structure than any of the other predictors for Φ_{ST} and D_{est} (Φ_{ST}: $R_{OW}^2 = 0.48$, $R_{Euc}^2 = 0.46$, $R_{BP}^2 = 0.39$, D_{est}: $R_{OW}^2 = 0.30$, $R_{EUC}^2 = 0.30$, $R_{BP}^2 = 0.21$) and F_{ST} was not well predicted regardless of distance metric ($R^2 < 0.12$).

Bayesian selection of important migration parameters

Four independent replicate BEAST runs converged to the same likelihood distribution after the removal of approximately 10–30 million burn-in steps from each run. The combined logfiles each contained about 107 million total steps, with high effective sample size (ESS > 200) values further indicating convergence. The analysis highlighted 14 migration rates as having a Bayes Factor (BF) of >3 (posterior odds of > 3:1). Most of these (11/14) were between adjacent spatial clusters, even though no prior information on location was given in the analysis (Fig. 3). Regions in the central portion of the Coral Triangle had multiple connections among them. All population clusters were connected to the network by at least one significant migration parameter with the exception of Western Australia (here comprising only 7 samples from Ningaloo Reef). The best-supported migration parameter to this region came from the GBR with a BF of 2.54. Randomization of tip locations yielded a migration matrix that was also apparently random, with only five of the 22 well-supported migration parameters occurring between neighboring spatial clusters.

The probability distribution for the location of the most common recent ancestor shows a fairly flat surface across the Coral Triangle, but these probabilities were generally higher than for peripheral populations (Online Fig. 2A). However, randomization of tip locations showed that the prior expectation for the location of the root was highly correlated with sample size (Online Fig. 2A,b; $R^2 = 0.995$).

Discussion

Most phylogeographic studies of Indo-Pacific species to date have relied on estimates of mitochondrial genetic structure to then make inferences about gene flow (e.g., Crandall et al. 2008b, Ackiss et al. 2013, Raynal et al. 2014, and see Keyse et al. 2014 for a review of the geographic and genetic scope of 108 such studies). Because most such studies assume that patterns of gene flow have been held at a static equilibrium over a long period of time, it is impossible to determine from them whether this gene flow (or lack thereof) is historical or contemporary (but inferences are often made about the latter). However, when we invert our view to consider explicit models of gene flow (IBD and spatial diffusion) rather than genetic structure, we are able to see that the data contain a good deal of spatial

information, even in a species with relatively good planktonic dispersal potential and low genetic structure. The fact that spatial distances explain the data better than do oceanographic distances from a biophysical model indicates that historical colonizations and subsequent gene flow events are more strongly reflected in the mitochondrial genome of *L. laevigata* than is contemporary gene flow. These results complement and extend upon earlier observations that genetic data sets from this diverse region are often haunted by the "ghosts of dispersal past" (Benzie 1999).

These novel insights into the population structure and evolution of *L. laevigata* result directly from population sampling that includes localities from the periphery of the Indian and Pacific Oceans as well as from the Coral Triangle. The vast species ranges of many Indo-Pacific marine animals, such as that of *L. laevigata*, make comprehensive population genetic and phylogeographic studies difficult. Previous studies of Indo-Pacific taxa have either sampled in the Pacific and Indian Oceans with limited sampling in the Coral Triangle (e.g., Lavery et al. 1996, Benzie 1999, Williams and Benzie 1998), or extensively within the Coral Triangle without a broader context (e.g., Barber et al. 2006, DeBoer et al. 2008, Crandall et al. 2008b), albeit with some notable exceptions (e.g., Crandall 2008a, Vogler et al. 2012, 2013). Here, we are able to bring together one of the most spatially comprehensive surveys of genetic variation for any single Indo-Pacific species to date, resulting in the broad geographic context necessary for strong inference. Although our inference is confined to the history of a single locus, the extensive geographic scope of sampled populations combined with state-of-the-field analyses (reviewed in Liggins et al. 2013 – Chapter Three) as well as results from a biophysical model allow us to understand the phylogeography of this species at greater resolution than ever before.

Signal from the seascape

The first line of evidence for a strong geographic signal in this mitochondrial data set is provided by the result from non-metric multidimensional scaling of Φ_{ST}, which shows an extremely strong correlation with longitude (R^2 = 0.81, Fig. 5). This NMDS result is reminiscent of classic results from human population genetics (Cavalli-Sforza et al. 1994, Novembre et al. 2008), in which the first two dimensions of variation in the genetic data show a good fit to geography. However, this analysis is based on population genetic distances from a single information-rich locus, rather than PCA scores among individual genotypes (although the latter can be a special case of the former, Wang et al. 2010). Those studies demonstrated that a species (e.g., humans) that has experienced recent expansions and/or high equilibrium gene flow among neighboring populations (and thus has relatively few truly genetically distinct populations) will contain a good deal of spatial information in the two best explanatory dimensions of a multivariate analysis. Although it may be difficult to differentiate between historical and contemporary processes with this non-parametric method, our NMDS approach demonstrates that it is more fruitful to think about Indo-Pacific species in terms of gradational differentiation reflecting historical or contemporary gene flow rather than in terms of distinctly structured populations (see also Patterson et al. 2006; and Novembre and Stephens 2008 for in-depth discussion of underlying theory).

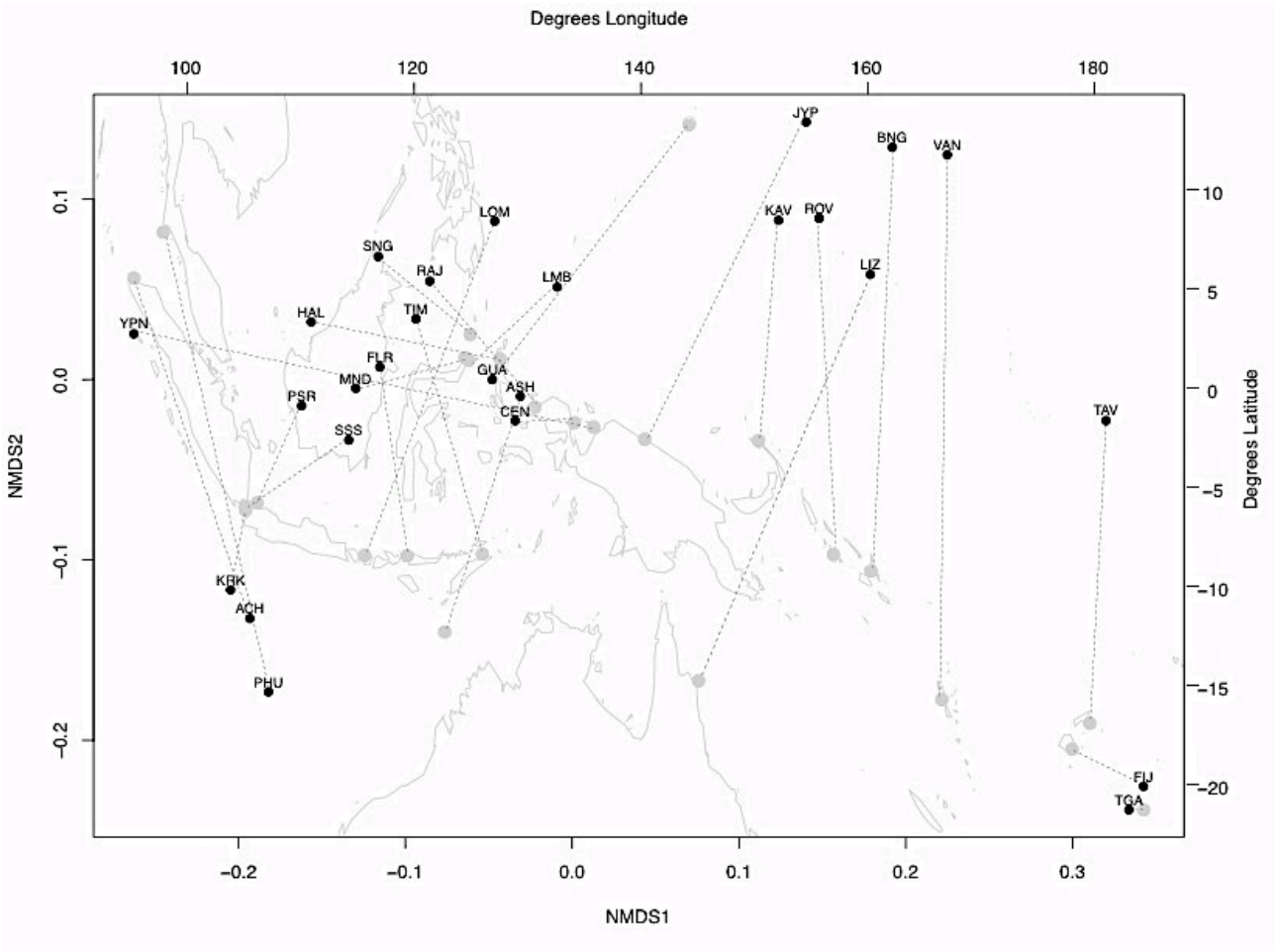


Figure 5. NMDS plots for Φ_{ST} values from populations with sample size > 10. See Table 1 for abbreviations. The first NMDS dimension has a strong correlation with longitude ($R^2 = 0.81$, $P < 1 \times 10^{-9}$), so a map of geography has been underlaid for reference, with the geographical positions of each sample noted.

A second line of evidence for geographic signal in this data set is found in strong and significant MRDM correlations of population genetic distances (Φ_{ST} or D_{est}) with three different proxies for dispersal distance (Fig. 4). This isolation-by-distance result is consistent with a stepping-stone model of gene flow or colonization among neighboring populations. Interestingly, the two geographic distance proxies were consistently better than the modeled larval dispersal distance for explaining the contemporary pattern of genetic distances in mtDNA. This result appears to run counter to that of White et al. (2010) who found that larval dispersal distance from a biophysical model provided a much stronger explanation than geographic distance for patterns of genetic structure in microsatellites. The stronger correlation of genetic distance to geographic distance as compared to oceanographic distance in our dataset likely arises from a mismatch between the timescale over which mtDNA integrates (1,000–100,000 yrs) and the timescale of contemporary larval dispersal among the geographic clusters (approximately 1–100 yrs). The genetic distances calculated from our mitochondrial data set are therefore probably capturing information about historical gene flow and colonization events rather than contemporary gene flow. Thus, mtDNA provides a valuable historical contrast to multi-locus genotyping (notably microsatellites and SNPs), which probably provide a closer fit to contemporary processes (Selkoe and Toonen 2006).

A final line of evidence comes from our phylogeographic diffusion model. by directly reconstructing the mitochondrial history in a spatial context, we are able to consider the role of colonization history and gene flow in the phylogeographic distribution of haplotypes, while simultaneously integrating over the uncertainty in the genealogy. Although we provided no prior information to the diffusion model about the relative spatial locations of each geographic cluster, it selected gene flow parameters that make intuitive sense: gene flow occurs for the most part among clusters that are geographic neighbors (Fig. 3). For example, our model found the Great Barrier Reef to be the most likely source for colonization of Western Australia (although this result did not rise above our threshold for significance), consistent with previous findings in another sea star (Vogler et al. 2013).

By starting with known sample locations and using ancestral state reconstruction to trace these locations back to the root over all possible genealogies, we were also able to arrive at a posterior probability distribution for the location of the MRCA (Online Fig. 2). While it is tempting to interpret from this distribution that the birthplace of the *L. laevigata* mitochondrial matriline was somewhere in the Coral Triangle, it does not depart significantly from a prior expectation based solely on sample size. We therefore can conclude that there is not much information about root location in this data set.

Population structure of L. laevigata in a broader geographic context

Extensive sampling of populations of *L. laevigata* across the majority of its Indo-Pacific range revealed the presence of pronounced genetic structure, but in ways that are materially different from previous studies. Early studies of *L. laevigata* (Williams and Benzie 1996, 1997, 1998) are frequently

cited as classic examples of divergence among populations of Pacific and Indian Ocean marine species across the Indo-Malay-Philippine Archipelago. However, the present study shows a more nuanced picture. First, although the haplotype network delineates two large clusters of haplotypes, these are not highly divergent, regionally distinct clades. Enhanced sample size in our study has filled in previously missing haplotypes. The frequency of each cluster follows a longitudinal cline, resulting in a distinctive pattern of isolation-by-distance (Fig. 4). The multiple-regression approach implemented in MRDM considers the effects of putative barriers to dispersal simultaneously with those of isolation-by-distance (i.e., historical or contemporary stepping-stone gene flow). Out of four possible barriers considered in our MRDM analyses, we found that the putative barrier to the east of Cenderawasih Bay was the only barrier that, when considered together with the effects of geographic distance, provided a consistently good fit to the genetic distances. This result was also captured by our AMOVA analyses, which showed that a simple partition segregating populations to the west (Cenderawasih Bay plus central Indonesia and Indian Ocean localities) and east (western and central Pacific localities) was a better descriptor of geographic differentiation than were partitions based on lines of disjunction associated with western Sumatra, the Sunda Shelf, the Halmahera current, or any combination of the four (Online Table 2).

Our recovery of only a single potential barrier to gene flow is distinctly different from previous work, which suggested barriers in the Sunda Strait and to the west of Cenderawasih Bay (Crandall et al. 2008b, Kochzius et al. 2009). It reflects the change in perspective provided by a larger study area and explicit consideration of the effects of stepping-stone gene flow (Meirmans 2012). While our other AMOVA partitions were significant, and our NMDS plot shows some clustering of Indian Ocean, central Indonesian, Papua + Great Barrier Reef, and South Pacific localities, these distinctions may be more an artifact of sampling design than of any discrete barrier to gene flow.

The proximal explanation for the significance of the genetic disjunction to the east of Cenderawasih Bay is a change in relative frequency of the Indian and Pacific clusters, with the Pacific cluster becoming dominant to the east of this barrier (see Fig. 4a in Crandall et al. 2008). In a way then, the disjunction is an artifact of phylogenetic distance between the two major clusters. However, it has been shown that phylogeographic breaks such as this often come to rest in regions of low contemporary migration (Barton and Hewitt 1985). Therefore we suggest that this particular disjunction could ultimately be due to the lack of reef habitat – moving east from Cenderawasih Bay there is almost 700 km of coastline with very sparse and minimal reef habitat (Spalding et al. 2001). Whereas gene flow likely occurs occasionally across Northern Papua, there are few stepping-stone populations to facilitate the exchange of migrants across generations. The stomatopod *Haptosquilla pulchella* Miers, 1880 shows a sharp genetic discontinuity between Cenderawasih Bay and populations in Papua New Guinea, suggesting that this pattern occurs in other Indo-Pacific taxa, but in general this region is surprisingly unknown as few other studies have included samples from Cenderawasih Bay and the western Pacific (Keyse et al. 2014, and see Liu et al. 2012).

Conclusions

The high dispersal capacity and wide range of many shallow reef Indo-Pacific marine organisms necessitate both dense and comprehensive sampling to provide the fullest phylogeographic context for each species. The present mitochondrial data set provides an example of the benefits of a spatially broadened perspective. Previously inferred barriers to gene flow turn out to be artifacts of an isolation-by-distance signature that was invisible at a smaller scale. The fact that this spatial signal is better explained by geographic distance than by oceanographic distances travelled by larvae indicates that the temporal resolution in the current dataset is low: our view of the present is obscured by the ghosts of dispersal past (*sensu* Benzie 1999). Future phylogeographic studies should strive to broaden their genetic perspective as well as their geographic perspective to increase temporal resolution. Nevertheless, when we consider models of gene flow rather than genetic structure we realize the importance of genetic exchange among neighboring demes as the primary reason for connectivity across a marine species' range (Crandall et al. 2012).

Detailed and broad genetic surveys such as this one usually have been beyond the scope of an individual study, or any single research group. As with most fields of science, increased collaboration and equitable sharing of data and expertise can provide a way forward (see Barber et al. 2014). We hope that the present study, which brings together data from numerous different laboratories and research efforts, will herald even greater collaborative endeavors for this region.

Acknowledgements

Funding for this work was provided by the Australian Research Council (DP0878306, to CR), the US National Science Foundation (OCE-0349177 and DEB-0338566 to PB, and DEB- 0508788 to EDC), the World Wildlife Fund (Kathryn Fuller Post-doctoral Research Fellowship, to EAT) and an Explorer's Club Exploration Fund (to LL). Sampling in Timor-Leste was supported by the Coral Triangle Support Partnership and the Ministério da Aquicultura e Pescas, Direcção Nacional de Pescas e Aquicultura (authorized by A Fernandes, L Fontes, J Freitas; guia de marssa: 502/DNPA/VIII/10 and 452/DNPA/VII/11). Export of samples was authorized by the Departmento de Quarentena das Pescas (export permit: 162/FQ006/EXP./DNQB/ VII/2011). Sampling in the Solomon Islands was via the Australian Government's Pacific strategy Assistance Program and with the assistance of the Roviana Conservation Foundation (Solomon Islands Government Ministry of Education and Human Resource Development and Ministry of Fisheries and Marine Resources Research Permit to S Albert). Sampling in Papua New Guinea was in coordination with the National Research Institute, the Department of Foreign Affairs and Immigration (Research Visa: 10350008304), and the Department of Environment and Conservation (Permit to Export Wildlife: 011318). Sampling in Indonesia was permitted by the Indonesian Institute of Sciences (#1187/SU/KS/2006 and #04239/SU.3/ KS/2006). Sampling in Fiji was permitted to EC by the Ministry of Fisheries and Forests, as was export permission (Permit

#C300/3006). The Vanuatu Environment Unit provided sampling and export permission to EC in Vanuatu. Authority to sample in Tonga was provided by the Ministry of Agriculture and Food, Forests and Fisheries (by U Fa'anunu to J Drew). Authority to sample at Ashmore Reef was provided by the Department of the Environment, Water, Heritage and the Arts (Access to Biological Resources in a Commonwealth Area for Non-Commercial Purposes permit number: AU-COM2010068) and with logistic support from Australian Customs and Border Control. Sampling in the Coral Sea was supported by the Marine Division of the Department of Sustainability, Environment, Water, Population and Communities (Access to Biological Resources in a Commonwealth Area for Non-Commercial Purposes permit number: AU-COM2008042). Sampling at Ningaloo Reef was under the authority of the Western Australia Department of Environment and Conservation (Licence to Take Fauna for Scientific Purposes: SF007126, SF006619; Authority to Enter Calm land/or Waters: CE002227, CE002627). We are grateful to the staff of the Australian Museum Lizard Island research station and Heron Island research station for their facilities and support (Great Barrier Reef Marine Park Authority and Queensland Parks and Wildlife Marine Parks Permit: G08/28114.1, G09/31678.1, G10/33597.1, G11/34640.1; Queensland Government Department of Primary Industries General Fisheries Permit: 118636, 150981; Australian Quarantine Inspection Service Permit to Import Quarantine Material: IP10017966). Collection of older samples from Thailand, Indonesia, Guam, and New Caledonia used in this study were permitted to J Benzie, and are now in the collections of G Wörheide. J Benzie is gratefully acknowledged for ceding his collection to G Wörheide. We especially thank the crew of the Ashmore Guardian, JD Aguirre, J Aini (Ailan Awareness), S Albert, J Barber-Choi, JF Bertrand, K Davis, J Drew, M Erdmann, A Jackson, M Jimuru, J Keyse, J Kinch (National Fisheries College, Papua New Guinea), E Lewis, W Lovell (Freeflow Dive, Dili), I McLeod, A Mirams, M Muñoz, S Pardede, S Penny, R Pinto (and staff of the Coral Triangle Support Partnership), A Smith (Tiki2 Adventure tours), T Sinclair-Taylor, C Starger, M Subia, C Taylor, A Turner, P Waldie, C Yusuf, Ambariyanto, Stephen, Lavud, and Takenda for logistical support, field assistance and help in the laboratory. Finally, we thank C Bird, S Williams, J Pearse, and an anonymous reviewer for insightful critiques of a previous draft of this paper.

Literature cited

Ackiss AS, Pardede S, Crandall ED, Ablan-Lagman M, Ambariyanto, Romena N, Barber PH, Carpenter KE. 2013. Pronounced genetic structure in a highly mobile coral reef fish, *Caesio cuning*, in the Coral Triangle. Mar Ecol Prog Ser. 480:185–197. http://dx.doi.org/10.3354/meps10199

Alberto F, Raimondi P, Reed DC, Watson JR, Siegel DA, Mitarai S, Coelho N, Serrao EA. 2011. Isolation by oceanographic distance explains genetic structure for *Macrocystis pyrifera* in the Santa Barbara channel. Mol Ecol. 20:2543–2554. PMid:21535280. http://dx.doi. org/10.1111/j.1365-294X.2011.05117.x

Aljanabi SM, Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA from PCR-based techniques. Nucleic Acids res. 25:4692–4693. http://dx.doi.org/10.1093/nar/25.22.4692

Balkenhol N, Waits LP, Dezzani RJ. 2009. Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. Ecography. 32:818–830. http:// dx.doi.org/10.1111/j.1600-0587.2009.05807.x

Bandelt H-J, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16:37–48. http://dx.doi.org/10.1093/oxfordjournals.molbev. a026036

Barber PH, Cheng SH, Erdmann MV, Tenggardjaja K, Ambariyanto. 2011. Evolution and conservation of marine biodiversity in the Coral Triangle: insights from stomatopod Crustacea. In: Held C, Koenemann S, Schubart C, editors. Phylogeography and population genetics in Crustacea. Vol. 19. Crustacean Issues. CRC Press. p 1–31.

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. PMid:17089967.

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. 35:235–253. PMid:15737594. http://dx.doi.org/10.1016/j.ympev.2004.10.004

Barber PH, Palumbi SR, Erdmann MV, Kasim Moosa M. 2000. A marine Wallace's line? Nature. 406:692–693. PMid:10963585. http://dx.doi.org/10.1038/35021135

Barber PH, Palumbi SR, Erdmann MV, Moosa MK. 2002. Sharp genetic breaks among populations of *Haptosquilla pulchella* (stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. Mol Ecol. 11:659–674. PMid:11972755. http://dx.doi. org/10.1046/j.1365-294X.2002.01468.x

Barton NH, Hewitt G. 1985. Analysis of hybrid zones. Annu Rev Ecol Syst. 16:113–148. http:// dx.doi.org/10.1146/annurev.es.16.110185.000553

Beaumont MA. 2010. Approximate Bayesian computation in evolution and ecology. Annu Rev Ecol Evol Syst. 41(1):379–406. http://dx.doi.org/10.1146/annurev-ecolsys-102209-144621

Beerli P, Felsenstein J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proc Natl Acad Sci USA. 98(8):4563–4568. PMid:11287657. PMcid:PMc31874. http://dx.doi.org/10.1073/ pnas.081068098

Beerli P, Palczewski M. 2010. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. Genetics. 185(1):313–326. PMid:20176979. PMcid:PMc2870966. http://dx.doi.org/10.1534/genetics.109.112532

Bellwood DR, Hughes T. 2001. Regional-scale assembly rules and biodiversity of coral reefs. Science. 292:1532–1534. PMid:11375488. http://dx.doi.org/10.1126/science.1058635

Benzie J. 1999. Genetic structure of coral reef organisms: ghosts of dispersal past. Am Zool. 39:131–145.

Bird CE, Karl SA, Smouse PE, Toonen RJ. 2011. Detecting and measuring genetic differentiation. In: Held C, Koenemann S, Schubart C, editors. Phylogeography and population genetics in Crustacea. Vol. 19. Crustacean Issues. CRC Press. p. 31–73.

Bloomquist EW, Lemey P, Suchard MA. 2010. Three roads diverged? Routes to phylogeographic inference. Trends Ecol Evol. 25(11):626–632. PMid:20863591. PMcid:PMc2956787. http://dx.doi.org/10.1016/j.tree.2010.08.010

Bowen BW, Rocha LA, Toonen RJ, Karl SA. 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

Buston PM, Jones GP, Planes S, Thorrold SR. 2012. Probability of successful larval dispersal declines fivefold over 1 km in a coral reef fish. Proc R Soc Lond Ser B-Biol Sci. 279(1735):1883– 1888. PMid:22158958. PMcid:PMc3311883. http://dx.doi.org/10.1098/rspb.2011.2041

Cann RL, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. Nature. 325(6099):31–36. PMid:3025745. http://dx.doi.org/10.1038/325031a0

Carpenter KE, Barber PH, Crandall ED, Ablan-Lagman MCA, Ambariyanto, Mahardika GN, Manjaji-Matsumoto BM, Juinio-Meñez MA, Santos MD, Starger CJ, et al. 2011. Comparative phylogeography of the Coral Triangle and implications for marine management. J Mar biol. 2011:1–14. http://dx.doi.org/10.1155/2011/396982

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

Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Princeton, New Jersey, UsA: Princeton University Press.

Chappell J, Shackleton NJ. 1986. Oxygen isotopes and sea level. Nature. 324:137–140. http:// dx.doi.org/10.1038/324137a0

Chenoweth SF, Hughes JM, Keenan CP, Lavery S. 1998. When oceans meet: a teleost shows secondary intergradation at an Indian-Pacific interface. Proc R Soc Lond Ser B-Biol Sci. 265(1394):415–420. PMcid:PMc1688902. http://dx.doi.org/10.1098/rspb.1998.0310

Cowen RK, Lwiza KM, Sponaugle S, Paris CB, Olson DB. 2000. Connectivity of marine populations: open or closed? Science. 287(5454):857–859. PMid:10657300. http://dx.doi. org/10.1126/science.287.5454.857

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, Muñoz MM, Akinbronbi B, Erdmann MV, Barber PH. 2008b. Comparative phylogeography of two sea stars and their ectosymbionts within the Coral Triangle. Mol Ecol. 17:5276–5290. PMid:19067797. http://dx.doi.org/10.1111/j.1365-294X.2008.03995.x

Crandall ED, Sbrocco EJ, DeBoer TS, Barber PH, Carpenter KE. 2012a. Expansion dating: calibrating molecular clocks in marine species from expansions onto the Sunda Shelf following the last glacial maximum. Mol Biol Evol. 29:707–719. PMid:21926069. http://dx.doi. org/10.1093/molbev/msr227

Crandall ED, Treml EA, Barber PH. 2012b. Coalescent and biophysical models of stepping-stone gene flow in neritid snails. Mol Ecol. 21(22):5579–5598. PMid:23050562. http:// dx.doi.org/10.1111/mec.12031

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

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:e88. PMid:16683862. PMcid:PMc1395354. http://dx.doi. org/10.1371/journal.pbio.0040088

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 7:214. PMid:17996036. PMcid:PMc2247476. http://dx.doi. org/10.1186/1471-2148-7-214

Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 22:1185–1192. PMid:15703244. http://dx.doi.org/10.1093/molbev/msi103

Excoffier L, Lischer HEL. 2010. Arlequin suite v3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Res. 10(3):564–567. PMid:21565059. http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 131:479–491. PMid:1644282. PMcid:PMc1205020.

Faurby S, Barber PH. 2012. Theoretical limits to the correlation between pelagic larval duration and population genetic structure. Mol Ecol. 21:3419–3432. PMid:22574811. http://dx.doi. org/10.1111/j.1365-294X.2012.05609.x

Fievet V, Touzet P, Arnaud J-F, Cuguen J. 2007. Spatial analysis of nuclear and cytoplasmic DNA diversity in wild sea beet (*Beta vulgaris* ssp. *maritima*) populations: do marine currents shape the genetic structure? Mol Ecol. 16:1847–1864. PMid:17444897. http://dx.doi. org/10.1111/j.1365-294X.2006.03208.x

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol. 3(5):294–299. PMid:7881515.

Foster NL, Paris CB, Kool JT, Baums IB, Stevens JR, Sanchez JA, Bastidas C, Agudelo C, Bush P, Day O, Ferrari R, et al. 2012. Connectivity of Caribbean coral populations: complementary insights from empirical and modeled gene flow. Mol Ecol. 21:1143–1157. PMid:22276913. http://dx.doi.org/10.1111/j.1365-294X.2012.05455.x

Fu Y-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 147:915–925. PMid:9335623. PMcid:PMc1208208.

Gaither MR, Bowen BW, Bordenave T-R, Rocha LA, Newman SJ, Gomez JA, van Herwerden L, Craig MT. 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 Evol Biol. 11:189. PMid:21722383. PMcid:PMc3145601. http:// dx.doi.org/10.1186/1471-2148-11-189

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. 37:133–147. http://dx.doi. org/10.1111/j.1365-2699.2009.02188.x

Galindo HM, Palumbi SR. 2006. Seascape genetics: a coupled oceanographic-genetic model predicts population structure of Caribbean corals. Curr Biol. 16:1622–1626. PMid:16920623. http://dx.doi.org/10.1016/j.cub.2006.06.052

Gerlach G, Atema J, Kingsford MJ, Black KP, Miller-Sims V. 2007. Smelling home can prevent dispersal of reef fish larvae. Proc Natl Acad Sci USA. 104(3): 858–863. PMid:17213323. PMcid:PMc1783404. http://dx.doi.org/10.1073/pnas.0606777104

Hall R. 2002. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. J Asian Earth Sci. 20:353–431. http://dx.doi.org/10.1016/s1367-9120(01)00069-4

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

Hey J, Nielsen r. 2007. Integration within the Felsenstein equation for improved Markov Chain Monte Carlo methods in population genetics. Proc Natl Acad Sci USA. 104(8):2785–2790. PMid:17301231. PMcid:PMc1815259. http://dx.doi.org/10.1073/pnas.0611164104

Jost L. 2008. G_{ST} and its relatives do not measure differentiation. Mol Ecol. 17:4015–4026. PMid:19238703. http://dx.doi.org/10.1111/j.1365-294X.2008.03887.x

Keyse J, Crandall ED, Toonen RJ, Meyer CP, Treml EA, Riginos CR. 2014. The scope of published population genetic data for Indo-Pacific marine fauna, and future research opportunities. Bull Mar Sci. 90:47–78. http://dx.doi.org/10.5343/bms.2012.1107

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

Kool JT, Paris CB, Barber PH, Cowen RK. 2011. Connectivity and the development of population genetic structure in Indo-West Pacific coral reef communities. Global Ecol Biogeogr. 20(5):695–706. http://dx.doi.org/10.1111/j.1466-8238.2010.00637.x

Legendre P, Lapointe F-J, Casgrain P. 1994. Modeling brain evolution from behavior: a permutational regression approach. Evolution. 48:1487–1499. http://dx.doi.org/10.2307/2410243

Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. PLoS Comput Biol. 5:e1000520. PMid:19779555. PMcid:PMc2740835. http://dx.doi.org/10.1371/journal.pcbi.1000520

Lessios H, Robertson D. 2006. Crossing the impassable: genetic connections in 20 reef fishes across the Eastern Pacific Barrier. Proc R Soc Lond Ser B-Biol Sci. 273:2201–2208. PMid:16901840. PMcid:PMc1635520. http://dx.doi.org/10.1098/rspb.2006.3543

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, et al. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science. 319(5866):1100–1104. PMid:18292342. http:// dx.doi.org/10.1126/science.1153717

Liggins L, Treml EA, Riginos C. 2013. Taking the plunge: an introduction to undertaking seascape genetic studies and using biophysical models. Geogr Compass. 7(3):173–196. http://dx.doi.org/10.1111/gec3.12031

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, Dai CF, Allen GR, Erdmann MV. 2012. Phylogeography of the neon damselfish *Pomacentrus coelestis* indicates a cryptic species and different species origins in the west Pacific Ocean. Mar Ecol Prog Ser. 458:155–167. http://dx.doi.org/10.3354/meps09648

Manel S, Holderegger R. 2013. Ten years of landscape genetics. Trends Ecol Evol. 28(10):614– 621. http://dx.doi.org/10.1016/j.tree.2013.05.012

McMillan WO, Weigt LA, Palumbi SR. 1999. Color pattern evolution, assortative mating, and genetic differentiation in brightly colored butterflyfishes (Chaetodontidae). Evolution. 53:247–260. http://dx.doi.org/10.2307/2640937

Meirmans PG. 2012. The trouble with isolation by distance. Mol Ecol. 21(12):2839–2846. PMid:22574758. http://dx.doi.org/10.1111/j.1365-294X.2012.05578.x

Meirmans PG, Van Tienderen PH. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Mol Ecol Notes. 4:792–794. http://dx.doi. org/10.1111/j.1471-8286.2004.00770.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. 30:937–949. http://dx.doi.org/10.1007/s00338-011-0767-x

Newman MEJ. 2006. Finding community structure in networks using the eigenvectors of matrices. Phys Rev E. 74:036104. PMid:17025705. http://dx.doi.org/10.1103/PhysrevE.74.036104

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, et al. 2008. Genes mirror geography within Europe. Nature. 456(7218):98–101. PMid:18758442. PMcid:PMc2735096. http://dx.doi.org/10.1038/nature07331

Novembre J, Stephens M. 2008. Interpreting principal component analyses of spatial population genetic variation. Nature Genet. 40(5):646–649. PMid:18425127. http://dx.doi.org/10.1038/ng.139

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2012. vegan: community ecology package, R package version 2.0-5.

Oppenheimer S. 2012. Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. Philos Trans R Soc Lond Ser B-Biol Sci. 367(1590):770–784. PMid:22312044. PMcid:PMc3267120. http://dx.doi.org/10.1098/rstb.2011.0306

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. PLoS Genet. 2(12):e190. PMid:17194218. PMcid:PMc1713260. http://dx.doi.org/10.1371/journal. pgen.0020190

Paulay G. 1990. Effects of late Cenozoic sea-level fluctuations on the bivalve faunas of tropical oceanic islands. Paleobiology. 16:415–434.

Polzin T, Daneshmand SV. 2003. On Steiner trees and minimum spanning trees in hypergraphs. Operations Res Lett. 31:12–20. http://dx.doi.org/10.1016/s0167-6377(02)00185-2

Renema W, Bellwood DR, Braga JC, 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. PMid:18669854. http://dx.doi.org/10.1126/science.1155674

Riginos C, Liggins L. 2013. Seascape genetics: populations, individuals, and genes marooned and adrift. Geogr Compass. 7(3):197–216. http://dx.doi.org/10.1111/gec3.12032

Roberts CM, McClean CJ, Veron JEN, Hawkins JP, Allen GR, McAllister DE, Mittermeier CG, Schueler FW, Spalding M, Wells F, et al. 2002. Marine biodiversity hotspots and conservation priorities for tropical reefs. Science. 295:1280–1284. PMid:11847338. http://dx.doi. org/10.1126/science.1067728

Rocha LA, Craig MT, Bowen BW. 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

Saenz-Agudelo P, Jones GP, Thorrold SR, Planes S. 2012. Patterns and persistence of larval retention and connectivity in a marine fish metapopulation. Mol Ecol. 21:4695–4705. http:// dx.doi.org/10.1111/j.1365-294X.2012.05726.x

Selkoe KA, Henzler CM, Gaines SD. 2008. Seascape genetics and the spatial ecology of marine populations. Fish Fish. 9:363–377. http://dx.doi.org/10.1111/j.1467-2979.2008.00300.x

Spalding M, Ravilious C, Green EP. 2001. World atlas of coral reefs. Berkeley: University of California Press.

Spalding MD, Fox HE, Halpern BS, McManus MA, Molnar J, Allen GR, Davidson N, Jorge ZA, Lombana AL, Lourie SA, et al. 2007. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. Bioscience. 57(7):573–583. http://dx.doi.org/10.1641/b570707

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595. PMid:2513255. PMcid:PMc1203831.

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

Toonen RJ, Andrews KR, Baums IB, Bird CE, Concepcion GT, Daly-Engel TS, Eble JA, Faucci A, Gaither MR, Lacchei M, et al. 2011. Defining boundaries for ecosystem-based management: a multispecies case study of marine connectivity across the Hawaiian Archipelago. J Mar Biol. 2011:1–13. http://dx.doi.org/10.1155/2011/460173

Treml EA, Halpin PN. 2012. Marine population connectivity identifies ecological neighbors for conservation planning in the Coral Triangle. Conserv Lett. 5:441–449. http://dx.doi. org/10.1111/j.1755-263X.2012.00260.x

Treml EA, Roberts JJ, Chao Y, Halpin PN, Possingham HP, Riginos C. 2012. Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. Integrative Comp Biol. 52:525–537. PMid:22821585. http://dx.doi. org/10.1093/icb/ics101

Vogler C, Benzie J, Barber PH, Erdmann MV, Ambariyanto, Sheppard C, Tenggardjaja K, Gérard K, Wörheide G. 2012. Phylogeography of the crown-of-thorns starfish in the Indian Ocean. PLoS ONE. 7:e43499. PMid:22927975. PMcid:PMc3424128. http://dx.doi. org/10.1371/journal.pone.0043499

Vogler C, Benzie J, Tenggardjaja K, Barber PH, Wörheide G. 2013. Phylogeography of the crown-of-thorns starfish: genetic structure within the Pacific species. Coral Reefs. 32:515– 525. http://dx.doi.org/10.1007/s00338-012-1003-z

Voris HK. 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. J Biogeogr. 27:1153–1167. http://dx.doi.org/10.1046/j.1365-2699.2000.00489.x

Wainscoat JS, Hill AV, Boyce AL, Flint J, Hernandez M, Thein SL, Old JM, Lynch JR, Falusi AG, Weatherall DJ. 1986. Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. Nature. 319(6053):491–493. PMid:3003580. http://dx.doi. org/10.1038/319491a0

Wang C, Szpiech ZA, Degnan JH, Jakobsson M, Pemberton TJ, Hardy JA, Singleton AB, Rosenberg NA. 2010. Comparing spatial maps of human population-genetic variation using Procrustes analysis. Stat Appl Genet Mol Biol. 9:Article 13. http://dx.doi. org/10.2202/1544-6115.1493

Wang P, Song YT, Chao Y, Zhang HC. 2005. Parallel computation of the regional ocean modeling system. Int J High Performance Comput Appl. 19:375–385. http://dx.doi. org/10.1177/1094342005059115

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 7:256–276. http://dx.doi.org/10.1016/0040-5809(75)90020-9

Wheeler B. 2010. lmPerm: permutation tests for linear models.

White C, Selkoe KA, Watson J, Siegel DA, Zacherl DC, Toonen RJ. 2010. Ocean currents help explain population genetic structure. Proc R Soc Lond Ser B-Biol Sci. 277(1688):1685–1694. PMid:20133354. PMcid:PMc2871860. http://dx.doi.org/10.1098/rspb.2009.2214

Whitlock M, McCauley D. 1999. Indirect measures of gene flow and migration: F_{ST} does not equal 1/(4Nm+1). Heredity. 82:117–125. http://dx.doi.org/10.1038/sj.hdy.6884960

Whitlock MC. 2011. G'_{ST} and D do not replace F_{ST}. Mol Ecol. 20(6):1083–1091. PMid:21375616. http://dx.doi.org/10.1111/j.1365-294X.2010.04996.x

Williams ST. 2000. Species boundaries in the starfish genus *Linckia*. Mar Biol. 136:137–148. http://dx.doi.org/10.1007/s002270050016

Williams ST, Benzie JAH. 1993. Genetic consequences of long larval life in the starfish, *Linckia laevigata* (Echinodermata; Asteroidea), on the Great Barrier Reef, Australia. Mar Biol. 117:71–77. 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. 126:99–107. http://dx.doi.org/10.1007/bF00571381

Williams ST, Benzie JAH. 1997. Indo-West Pacific patterns of genetic differentiation in the high-dispersal starfish *Linckia laevigata*. Mol Ecol. 6:559–573. http://dx.doi.org/10.1046/j.1365-294X.1997.00221.x

Williams ST, Benzie JAH. 1998. Evidence of a biogeographic break between populations of a high dispersal starfish: congruent regions within the Indo-West Pacific defined by color morphs, mtDNA, and allozyme data. Evolution. 52:87–99. http://dx.doi.org/10.2307/2410923

Williams ST, Duda TF. 2008. Did tectonic activity stimulate Oligo-Miocene speciation in the Indo-West Pacific? Evolution. 62(7):1618–1634. PMid:18410535. http://dx.doi.org/10.1111/j.1558-5646.2008.00399.x

Williams ST, Jara J, Gomez E, Knowlton N. 2002. The marine Indo-West Pacific break: contrasting the resolving power of mitochondrial and nuclear genes. Integrative Comp Biol. 42:941–952. PMid:21680374. http://dx.doi.org/10.1093/icb/42.5.941

Wright S. 1950. The genetical structure of populations. Ann Eugenics. 15(1):323–354. http://dx.doi.org/10.1111/j.1469-1809.1949.tb02451.x

Yamaguchi M. 1973. Early life history of coral reef asteroids, with special reference to *Acanthaster planci* (L.). In: Jones OA, Endean R, editors. Biol Geol coral reefs. New York and London: Academic Press, p. 369–389. PMid:4584395. http://dx.doi.org/10.1016/ b978-0-12-395526-5.50019-5

APPENDIX THREE. A novel widespread cryptic species and phylogeographic patterns within several giant clam species (Cardiidae: *Tridacna*) from the Indo-Pacific Ocean

Published in PLoS ONE 8: e80858. 2013

Thomas Huelsken[1], Jude Keyse[1], **Libby Liggins**[1], Shane Penny[2], Eric A. Treml[1,3] and Cynthia Riginos[1]

[1] School of Biological Sciences, The University of Queensland

[2] Research Institute for Environment and Livelihoods, Charles Darwin University

[3] Department of Zoology, University of Melbourne

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.

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 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.

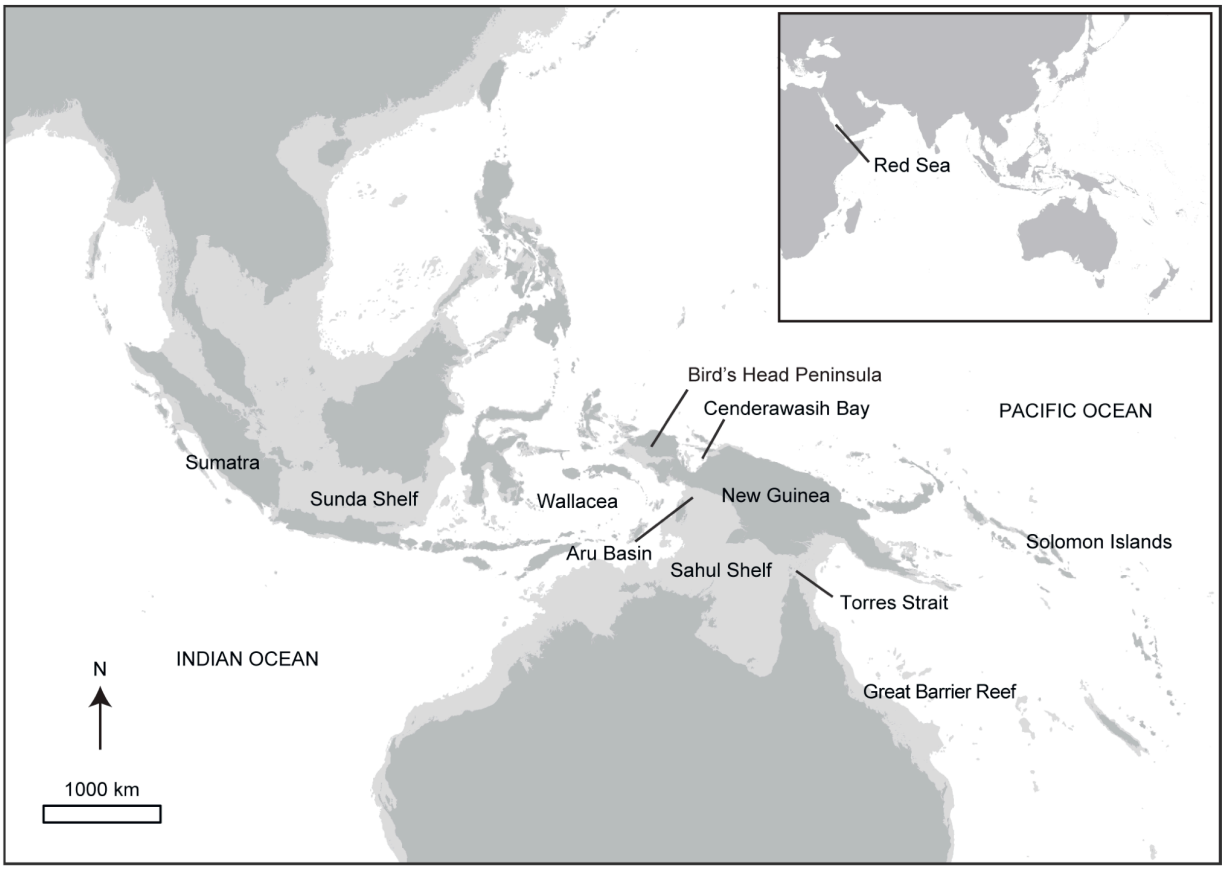


Figure 1. Study region. The light grey outline represents the lowest Pleistocene sea level (120 m depth contour).

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 and 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 1,000 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

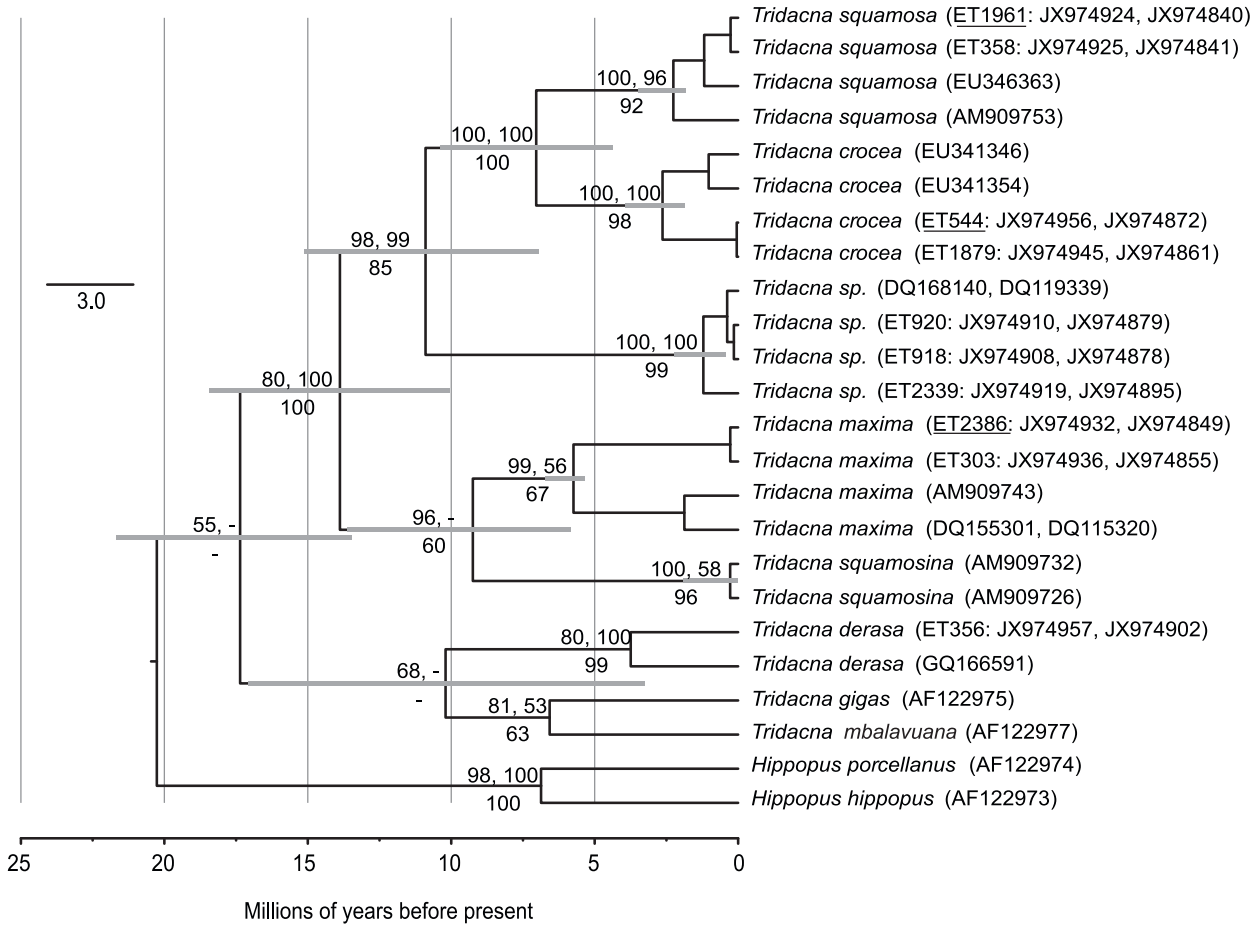


Figure 2. Species relationships within *Tridacna* based on concatenated mitochondrial DNA (COI and 16S) sequences. The topology shown is a time calibrated maximum clade credibility tree inferred with StarBEAST under a birth-death model. Bayesian posterior probabilities from StarBEAST and MrBayes are 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.

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 Islands, the Torres Strait and Lizard Island (and additionally western New Guinea/Cenderawasih 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/Cenderawasih 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

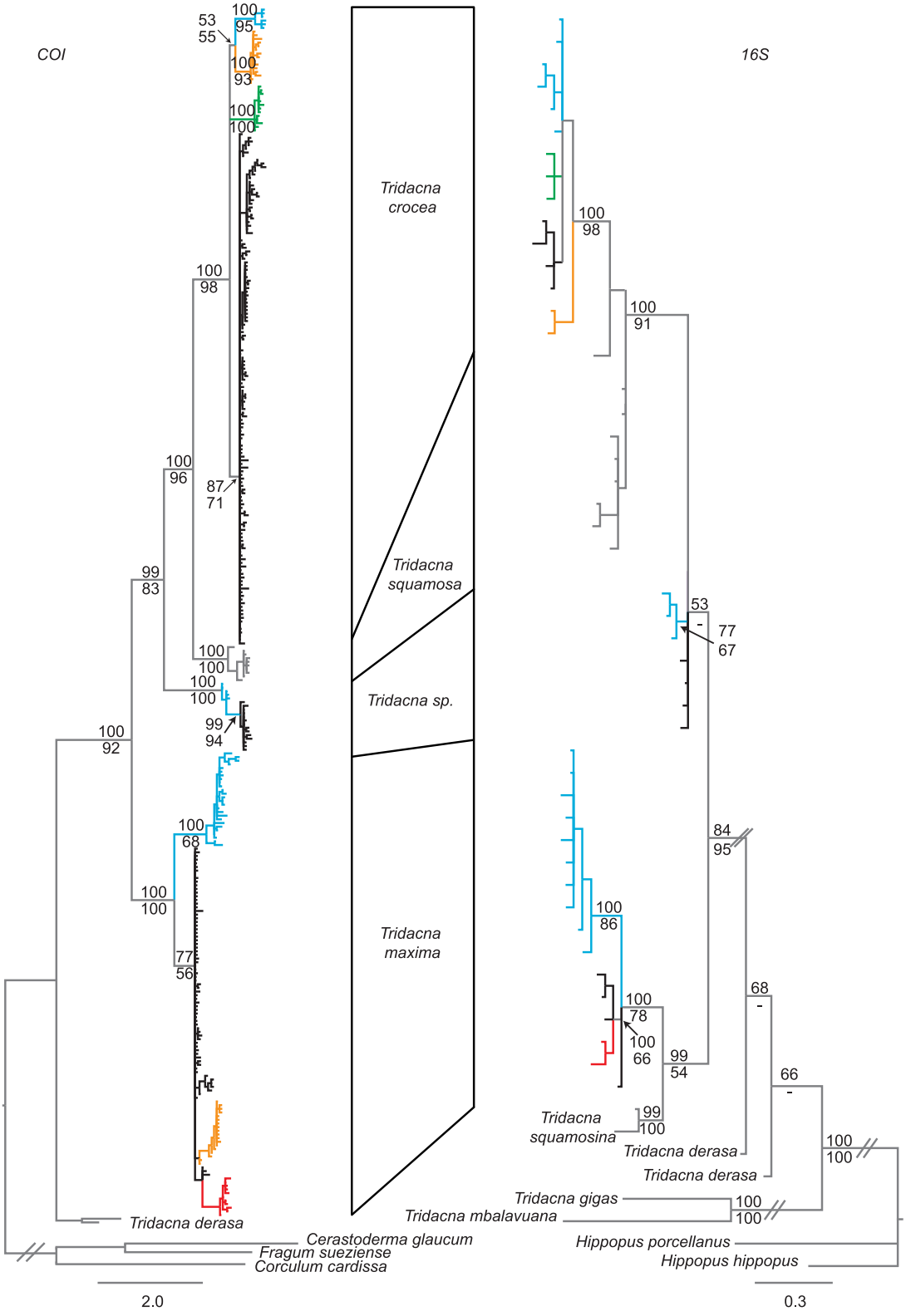


Figure 3. Bayesian phylogenetic trees for mitochondrial CO1 and 16S. MrBayesian consensus trees constructed for each gene region using all available data. Although different species and regions have differential representation, the two gene trees are concordant, as is expected for linked loci. Thus, overall patterns are consistent among research groups. Branch colors correspond to distinct lineages whose geographic distributions are described in Fig. 5.

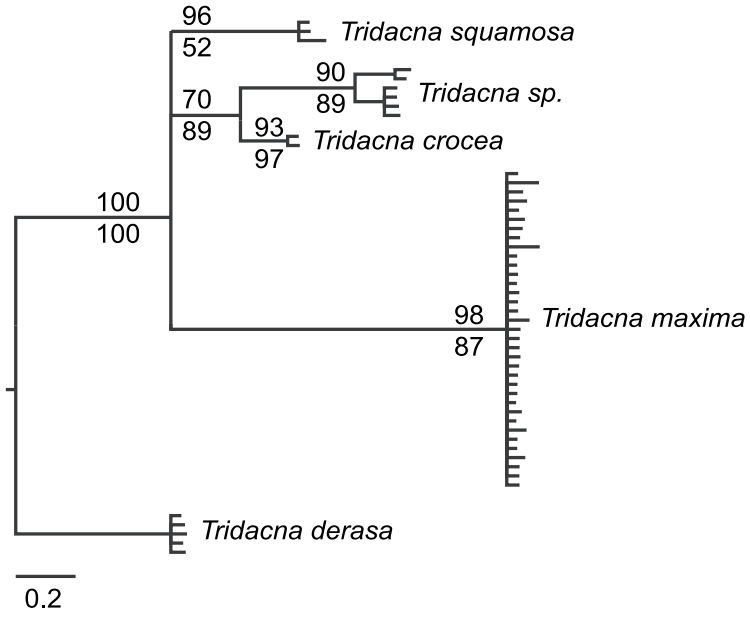


Figure 4. Species relationships within *Tridacna* based an ITS MrBayes consensus tree. Unalignable regions have been excluded. Bayesian posterior probabilities are above branches and RAxML bootstrap support percentages are below branches.

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.

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.*

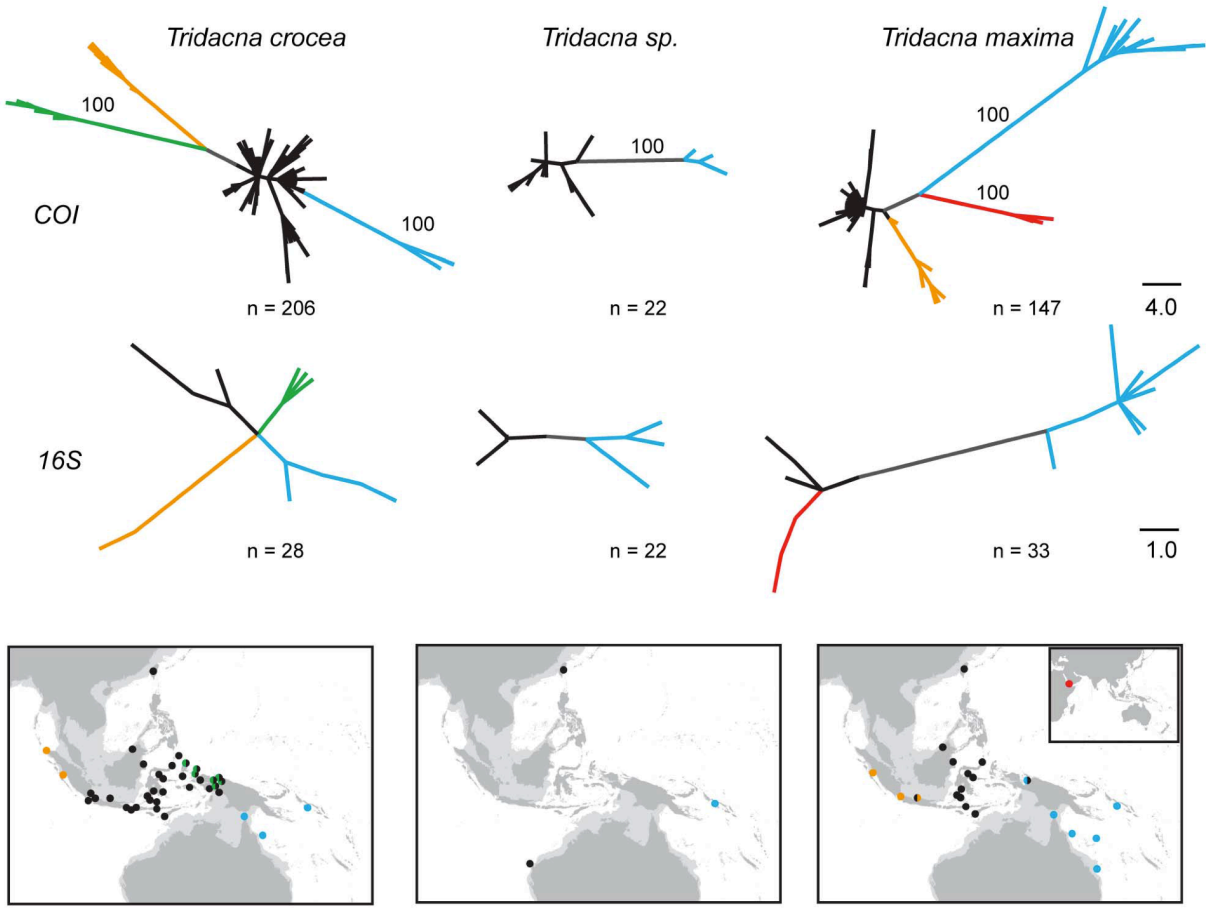


Figure 5. Unrooted parsimony trees and sampling locations for *Tridacna crocea*, *Tridacna* sp., and *Tridacna maxima*. Major lineages on networks are colored and the geographic extent of each lineage is indicated on the map. Relative frequencies of each haplotype are not depicted; each haplotype is 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.

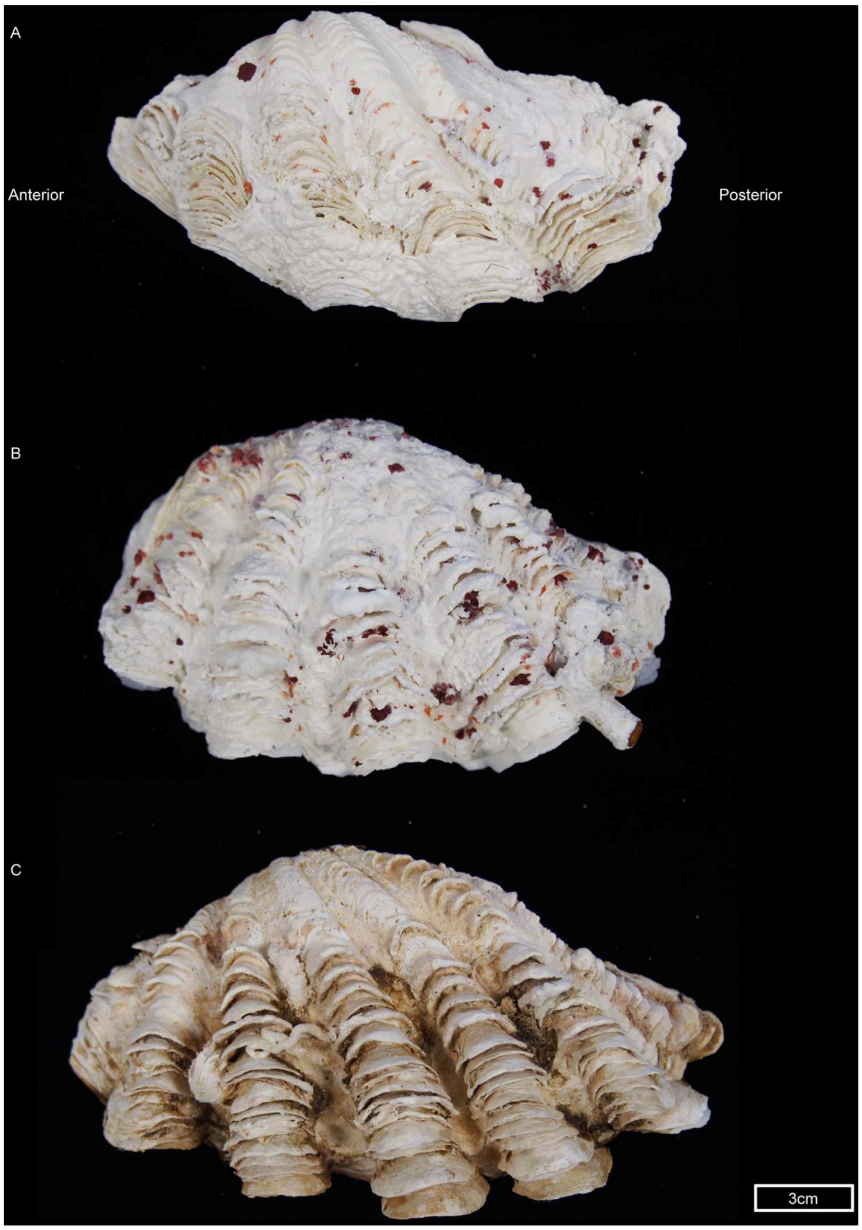


Figure 6. An individual with *Tridacna* sp. mtDNA demonstrating valve morphology consistent with *Tridacna maxima*. A) *Tridacna maxima* from Hibernia Reef, WA, Australia. Accession No# P.52722 (Museum Art Gallery Northern Territory (MAGNT)), original identification based on morphology, B) *Tridacna* 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.

(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.

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 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].

Acknowledgments

Sampling in the Coral Sea was supported by the Marine Division of the Australian Government Department of Sustainability, Environment, Water, Population and 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. 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 Postdoctoral 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) and an Explorers Club Exploration Fund to LL. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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, Lucas JS, 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 Malako-Zoologischen 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 and 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.

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.

APPENDIX FOUR. Hermatypic corals and crown-of-thorns starfish of the Kermadec Islands

To be published in Kermadec Biodiscovery Expedition 2011 (Trnski T, Schlumpf H, eds.). Auckland: Bulletin of the Auckland Museum 20: in press

Zoe T. Richards[1] and **Libby Liggins**[2]

[1] Australian Museum

[2] School of Biological Sciences, The University of Queensland

Main text

The Kermadec Islands (29- 31.5°S and 178-179° W) are an isolated volcanic archipelago situated midway between tropical Tonga and temperate New Zealand. Despite their isolation (~800km from the nearest landmass) and subtropical position (just 1°N of the southern-most coral reefs of the world, Lord Howe Island), previous studies have shown a mix of reef-building (hermatypic) tropical, subtropical and temperate corals are present (Cairns, 1995; Kosmynin, 1994; Brook, 1999; Wicks et al., 2010). While coral reefs typical of tropical locations in the Pacific Ocean are not formed, recent surveys have substantiated that considerable populations of not only reef-building corals, but corallivores are present.

The first accounts of Kermadec Island coral fauna were published by Vaughan (1917) and these records were used by Wells (1954), and Stehli and Wells (1971) in their analysis of coral biogeographic patterns. The next significant coral diversity study was Kosmynin (1994) who raised the number of scleractinian species known from the group from 6 to 14 and Cairns (1995) who documented the azooxanthellate (i.e. lacking algal symbionts) species. In 1999, Brook revised this number to 24 species {17 zooxanthellate species from 15 genera and 7 azooxanthellate species from 6 genera}.

From material collected on the RV Braveheart expedition (2011), 11 hard coral species from 10 genera were identified (8 zooxanthellate, 3 axoozanthellate) (*Hydnophora exesa**; *Pocillopora damicornis**; *Goniastrea favulus**; *Coscinaraea columna*; *Montastrea curta*; *Turbinaria frondens*; *Montipora spongodes*; *Montipora capricornis*; *Tubastrea diaphana*; *Balanophyllia cf. chnous*[#]; *Culicia rubeola*). Three of these species (*) are not recorded from this location in Veron (2000), and one species is a deep-water specialist (#) considered to be endemic to the Southwest Pacific Ocean (Cairns 1995).

The *Balanophyllia* species tentatively identified here as cf. *B. chonus*, is only previously known from the 140-549 m depth range (Cairns, 1995), so the finding of this species in shallow water is

atypical. It is recommended further confirmation of this identification could be should be sought from an azooxanthellate coral taxonomic expert such as Dr Stephen Cairns or Dr Marcello Kitahara. Another of the azooxanthellate specimens identified in the 2011 (*Culicia rubeola*) was recorded at the Kermadec Islands by Brook (1999); however that record was considered dubious {denoted with a (?)}. *Culicia rubeola* is a putative endemic to the New Zealand region, recorded depths of 0-82m (Cairns, 1995) hence records from the Kermadec Islands are regionally significant. However, a revision of the genus is needed and no attempt has been made here to rigorously compare this *C. rubeola* sample to the other 12-14 species known from the genus worldwide.

Furthermore, taxonomic uncertainty surrounds two of the putative zooxanthellate species collected. The first relates to the *Hydnophora* spp. The 2011 samples of *Hydnophora* sp. show close affinity to *Hydnophora exesa* (which was recorded by Kosmynin, 1994 at the Kermadec Islands); however Kosmynin's identification was revised by Brook (1999) as *Hydnophora pilosa*. Further examination of the type material of these samples in relation to the 2011 material is advisable.

Three *Montipora* samples identified from the present collection resemble samples identified by Kosmynin (1994) as *Montipora caliculata*; however Brook (1999) revised Kosmynin's identification to M. capricornis. The 2011 *Montipora* samples show affinity to *Montipora capricornis*. Again, further examination of the Kosmynin and Brook skeletal material {deposited at the Museum of New Zealand, National Museum of Natural History (Smithsonian Institution) and the Auckland War Memorial Museum}, in relation to the 2011 material (deposited in the Australian Museum, Sydney) is advisable.

Despite not currently being classified as a coral reef, tropical hermatypic coral species are present and the benthic cover of hard corals has been documented to reach 40% at Raoul Island (Brook, 1999). Furthermore, for at least three decades, the Kermadec coral community has supported a population of crown-of-thorns starfish (*Acanthaster planci*; McKnight 1978). Observations in 2012 indicate there is currently a high density of adults around Raoul and its satellite islands (i.e. > 0.008 individuals per 1 m^2 on the reef to the west of North Meyer Island), and the presence of juveniles suggests recent recruitment. Previous records suggest high abundance has been sustained for an extended period at this location; up to 0.008 individuals per 1 m^2 were observed on the western side of North Meyer Island in 1995 (Brook 1999), and in 2002 Gardner et al. (2006) observed 0.25 individuals per 1 m^2 on this same reef (6-9 m).

Further systematic study is required to resolve the identity of the *Balanophyllia*, *Culicia*, *Hydnophora* and *Montipora* species of the Kermadec Islands. Overall, this marginal and pristine community represents a valuable opportunity to examine coral and coral predator transitions into high latitude locations and for that reason, further quantitative data is needed to document the current abundance and extent of coral and coral predators around the Kermadec Islands.

Acknowledgements

Thanks to Tom Trnski and Stephen Keable and the Sir Peter Blake Trust, the Commanding Officer and Ship's Company of HMNZS Canterbury (2012 expedition, www.youngblakeexpeditions.org).

References

Brook FJ (1999) The coastal scleractinian coral fauna of the Kermadec Islands, southwestern Pacific Ocean. *Journal of the Royal Society of New Zealand* 29:435-460

Cairns SD (1995) The marine fauna of New Zealand. Scleractinia (Cnidaria: Anthozoa). *Oceanographic Institute Memoir* 103.

Gardner JPA, Curwen MJ, Long J, Williamson RJ, Wood AR (2006) Benthic community structure and water column characteristics at two sites in the Kermadec Islands Marine Reserve, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 40:1, 179-194.

Kosmynin, VN (1994) Shallow-water scleractinian corals from Kermadec Islands. *Atoll Research Bulletin* 418.

McKnight DG (1978) *Acanthaster planci* (Linneaus) (Asteroidea: Echinodermata) at the Kermadec Islands. *New Zealand Oceanographic Records* 4:17-20.

Stehli, FG, Wells, JW (1971) Diversity and age patterns in hermatypic corals. *Systematic Zoology* 20: 115-126.Vaughan, TW (1917) Some corals from Kermadec Islands. *Transactions and Proceedings of the New Zealand Institute* 49: 275-279.

Veron JEN (2000) Corals of the World, Volumes 2 and 3. Australian Institute of Marine Science.

Wells, JW (1954) Recent corals of the Marshall Islands. *Professional Paper of the Unites States Geological Survey* 260: 385-486.

Wicks LC, Gardner JPA, Davy SK (2010) Spatial patterns and regional affinities of coral communities at the Kermadec Islands Marine Reserve, New Zealand – a marginal high-latitude site. *Marine Ecology Progress Series* 400:101-113.