Genetic structure, population connectivity and taxonomic identity of the black-lip pearl oyster *Pinctada margaritifera* (Bivalvia: Pteriidae), across its Indo-Pacific distribution

Thesis submitted by

Monal M. Lal BScMS, MScMS

in November 2016

for the Degree of Doctor of Philosophy in Aquaculture

College of Science and Engineering

James Cook University

Townsville, Queensland, Australia

FRONTISPIECE

Pinctada margaritifera (L.) specimens from Savusavu, Vanua Levu, Fiji Islands. Clockwise from top left: an adult oyster cleaned of epifaunal growth, broodstock oysters feeding while being conditioned for spawning, and examples of the orange and black tissue colour morphotypes characteristic of this species. The scale bars each represent 20 mm.

DECLARATION

I declare that this thesis is my own original work, and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is provided.

Monal M. Lal
25 November 2016

STATEMENT OF CONTRIBUTIONS BY OTHERS

My PhD candidature was supported by an Australian Centre for International Agricultural Research (ACIAR) John Allwright Fellowship (JAF) award, which provided for living costs in Australia, as well as enrolment and tuition fees at James Cook University (JCU). Further living expenses were supported by a $2,500.00 AUD grant award from the JCU College of Science and Engineering Higher Degree Research Enhancement Scheme Funding round in 2016. All primary research costs were supported by the ACIAR research project "Pearl Industry Development in the Western Pacific" (FIS/2009/057)", led by Paul Southgate at the University of the Sunshine Coast. Supplementary research costs were supported by a grant award of $2,000.00 AUD by the then JCU College of Marine and Environmental Sciences (CMES) Graduate Research Scheme Funding round in 2014, as well as CMES Internal Research Allocation (IRA) entitlements of $3,000.00 AUD. A further grant of $3,000.00 AUD was awarded in 2015 from the Crawford Fund, following a successful application during the inaugural round of the Crawford-in-Queensland Student Awards scheme.

Logistic support for fieldwork in the Fiji Islands was provided by ACIAR project partners the Secretariat of the Pacific Community (SPC) and the University of the South Pacific (USP). All maps included in this thesis were generated with the assistance of Litia Gaunavou at SPC, and for either providing or assisting with collection of oyster tissue samples, I acknowledge the contributions of the following individuals (in no particular order): Gustaf Mamangkey, Naomi Gardiner, Ismail Saidi, Samantha Nowland, Rowan McIntyre, Steve Warden, Jo Buckee, Tina Weier, Georgia Langdon, Tevainui Frogier-Ellis, Hoc Tan Dao, Hứa Thái Tuyến, Samad Jahangard, Hossein Rameshi, Mehdi Doroudi, Max Wingfield, Laura Simmons, Gregory Bennett, Philippa Cohen, Waghon Lalao, Yu Wen Chiu, Eric Gan, Laisiasa Cavakiqali, Cherie Morris, Shirleen Bala, Justin Hunter, Pranesh Kishore, Adi Dionani Salaivanua, Kelly Brown, Jerome Taoi, Epeli Loganimoce, Albert Whippy, Bai Whippy, Toga Whippy, Isimeli Loganimoce, Marilyn Vilisoni, Babitu Rarawa, Ilitomasi Nuku, Samisoni Rakai, Patrick Fong, Nepoci Raleve and Claude Prévost.

I was involved in the conceptualisation and experimental design of all work presented in this thesis, including laboratory based experiments and fieldwork for collection of oyster tissue

samples. Laboratory based experiments included gDNA extraction, gDNA and amplicon product quality controls, gDNA quantification and standardisation, ddRADseq library preparation, as well as sample preparation for submission for commercial genotyping (DArTseq^{TM}). I was also primarily responsible for data management, integrity, analysis and interpretation. This included SNP/PAV genotyping and filtering (chapters 2-5), population genomic analyses (chapters 2-4), hydrodynamic particle dispersal simulations (chapters 3 and 4) and phylogenomic analyses (chapter 5). I was the sole author for all written work throughout the six chapters of this thesis, and am the lead author in all peer-reviewed journal article manuscripts derived from this work, both published, and in preparation. Specific co-author contributions for this thesis are outlined by chapter in the table below.

Chapter	Publishable unit generated	Author intellectual inputs and affiliations
1	**Lal, M.M.**, Southgate P.C., Jerry D.R., Zenger K.R. (2016) Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. **Marine Genomics** 25, 57-68.	**Lal MM**[1] Study execution, project design, writing, analysis and editing **Southgate PC**[1,2] Project design, conception, funding, supervision and editing **Jerry DR** [1] Supervision, technical advice on population genomic analysis and editing **Zenger KR**[1] Project design, conception, supervision, technical advice on population genomic analysis and editing
2	**Lal, M.M.**, Southgate P.C., Jerry D.R., Bosserelle C., Zenger K.R. (2016) A parallel population genomic and hydrodynamic approach to fishery management of highly-dispersive marine invertebrates: the case of the Fijian black-lip pearl oyster *Pinctada margaritifera*. *PloS ONE* **11**, e0161390.	**Lal MM**[1] Study execution, project design, writing, analysis and editing **Southgate PC**[1,2] Project design, conception, funding, supervision and editing **Jerry DR** [1] Supervision, technical advice on population genomic analysis and editing **Bosserelle, C**[3] Design, conception and execution of hydrodynamic particle dispersal model, and editing **Zenger KR**[1] Project design, conception, supervision, technical advice on population genomic analysis and editing
3	**Lal, M.M.**, Southgate P.C., Jerry D.R., Bosserelle C., Zenger K.R. (unpubl.) Swept away: ocean currents and seascape features	**Lal MM**[1] Study execution, project design, writing, analysis and editing

	influence genetic structure across the 18,000 Km Indo-Pacific distribution of a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. Submitted to **BMC Biology**. **Lal M.M.**, Southgate P.C., Jerry D.R., Bosserelle C., Zenger K.R. (unpubl.) Range-wide larval connectivity of the black-lip pearl oyster *Pinctada margaritifera* using hydrodynamic dispersal simulation. In preparation for **PloS ONE**.	**Southgate PC**[1,2] Project design, conception, funding, supervision and editing **Jerry DR** [1] Supervision, technical advice on population genomic analysis and editing **Bosserelle, C**[3] Design, conception and modification of hydrodynamic particle dispersal model, and editing **Zenger KR**[1] Project design, conception, supervision, technical advice on population genomic analysis and editing
4	**Lal, M.M.**, Southgate P.C., Jerry D.R., Zenger K.R. (unpubl.) Fresh evidence for the presence of a species complex in the black-lip pearl oyster *Pinctada margaritifera* (Bivalvia: Pteriidae). In preparation for **Scientific Reports**.	**Lal MM**[1] Study execution, project design, writing, analysis and editing **Southgate PC**[1,2] Project design, conception, funding, supervision and editing **Jerry DR** [1] Supervision, technical advice on phylogenomic analysis and editing **Zenger KR**[1] Project design, conception, supervision, technical advice on population genomic analysis and editing

Author affiliations

[1] Centre for Sustainable Tropical Fisheries and Aquaculture, and College of Science and Engineering, James Cook University, Townsville, QLD 4811, Queensland, Australia

[2] Australian Centre for Pacific Islands Research, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Maroochydore, QLD 4558, Queensland, Australia

[3] Geoscience Division, Secretariat of the Pacific Community, 241 Mead Road, Nabua, Suva, Fiji Islands

ACKNOWLEDGEMENTS

This PhD thesis represents the culmination of four very challenging, but rewarding years of work, and would not have been possible without the guidance and providence of the Almighty, and the support, encouragement and inputs of a great many individuals, to whom I am very thankful.

Firstly, I would like to express my sincere gratitude and appreciation to my primary advisor Kyall Zenger, for his support, guidance, encouragement and mentorship throughout my candidature. I arrived at JCU without a background in genetic analysis, and Kyall, I will always be grateful for your patience and friendship, which helped navigate the very steep learning curve that came with completing my PhD. Your ready availability, particularly during the early years of my candidature, have greatly contributed to my formative years as a researcher.

To Paul Southgate, who recommended me for an ACIAR JAF award and incorporated my research into the larger ACIAR project this work formed a part of, I owe you a large debt of gratitude for making my PhD studies possible. I greatly appreciate you helping me find my feet at JCU at the start of my candidature, and your inputs as the "pearl oyster guru". Your mentorship, advice, guidance and support on various aspects of this journey have helped me achieve a lifelong goal.

To Dean Jerry, I thank you very much for your sage advice, mentorship, encouragement, support and constructive inputs. Your special brand of humour kept even the more serious aspects of this PhD light-hearted, and made it an enjoyable experience. I still laugh when I need to consult my progress meeting recordings, as the often present early minutes of banter remind me of the great pleasure it was to work with you, and the very supportive environment of the aquaculture genetics research group.

I also acknowledge the valuable contribution of Cyprien Bosserelle, without whom the particle dispersal modelling would not have been possible. I will always be grateful for your agreement to collaborate, and availability to help troubleshoot the very complex code written

for the software. I look forward to working with you in the future, to further develop this very useful tool.

I am also grateful for the support of ACIAR, which made studying for a PhD in Australia possible. In particular, I would like to thank Tim Pickering for presenting me as a suitable candidate for a JAF award, Chris Barlow for his valuable advice, guidance and interest in how the research has progressed, as well as Sharon Harvey, for her help with administrative support. I am also grateful to Sharon for the opportunity to talk about my experiences as a JAF fellow in Canberra to commencing awardees in 2016. At JCU, I extend my gratitude to Alex Salvador and Katherine Elliot from the International Student Support office, Elizabeth Tynan and Jodie Wilson at the Graduate Research School, as well as Debbie Berry and Tammy Walsh at the College of Science and Engineering.

Studying at JCU has been a life-changing experience, and during my time here, I have had the opportunity to meet and work with a wide range of amazing people. To colleagues at the Molecular Ecology and Evolution Laboratory (MEEL), thank you all for the stimulating discussions, help with various analyses, and Friday beers when milestones were achieved. I am particularly grateful to Shannon Kjeldsen, whose advice, support, friendship and assistance were invaluable. In no particular order, I would like to acknowledge Maria Nayfa, Eike Steinig, Roger Huerlimann, David Jones, Donald McKnight, Tansyn Noble, Diana Pazmino-Jaramillo, Lynne van Herwerden, Gregory Maes, Carolyn Smith-Keune, Heather Welladsen, Georgia McDougall, Hoc Tan Dao, Basiita Komugisha, Giana Gomes, José Domingos, Floriaan Devloo-Delva, Sandra Infante Villamil, Kyle Allen, Ali Razaghi, Jarrod Guppy, Julie Goldsbury, Alyssa Budd and Geoffrey Collins.

I would also like to acknowledge the encouraging and constructive feedback obtained from anonymous reviewers of journal articles produced during my candidature, as your inputs and insights have enhanced the quality of the work. Embarking on a project to study an organism as widely distributed as *P. margaritifera* was an ambitious goal, which would not have been realised without the assistance and contributions of numerous individuals who were involved in the fieldwork for sample collection. These people are acknowledged in the statement of contributions of others section of this thesis.

Finally, I would like to thank my family and friends, for their support and encouragement. In particular, *vinaka vakalevu* to my parents and brother, Joeli Veitayaki, the late Johnson Seeto, Viliame Waqalevu, Andreas Frey, Kelly Brown, the late Makereta Brown, Adi Siteri Tikoca, Jerome Taoi, Jake Taoi, Samisoni Rakai, Wame Tokalau, Cherie Morris, Shirleen Bala, Laura Williams, Marilyn Vilisoni, Smita Singh, Prerna Chand, Salote Waqairatu, Sione Kaitu'u, Epeli Loganimoce and Pranesh Kishore.

My heartfelt thanks to you all,
Monal

SUMMARY

The black-lip pearl oyster *Pinctada margaritifera* (L.) is a bivalve mollusc highly valued for cultured pearl and pearlshell production throughout its extensive Indo-Pacific natural distribution, where it makes substantial contributions to local economies and supports coastal community livelihoods. Despite its commercial importance, substantial knowledge gaps exist for this species, particularly regarding genetic structure and population connectivity at both local and regional scales, as well as its taxonomic identity. This information is required for the development of sustainable fishery management strategies, as well as responsible aquaculture practices, to ensure the persistence of healthy wild populations, and continued commercial production.

The overarching goal of the research undertaken for this thesis was to investigate the stock structure, connectivity and taxonomy of *P. margaritifera*, to inform fishery management and aquaculture practices across the extent of its Indo-Pacific distribution, with a particular focus on the Fiji Islands. Specifically, over four separate investigations, I develop novel genome-wide single nucleotide polymorphism (SNP) markers for this species, and use them to investigate population genetic structure, diversity, connectivity and local adaptation of Fijian oysters, as well as for populations sampled from across the broader ~18,000 km species distribution. I also compare estimates of population connectivity derived from genomic analyses with an independent hydrodynamic particle dispersal model, to corroborate patterns of larval transport between study sites. Finally, I utilise phylogenomic analyses to assess the evolutionary relationships of the black-lip pearl oyster across its natural distribution, and also to ascertain its taxonomic identity among other members of the family Pteriidae.

The first investigation developed 5,243 novel genome-wide SNP markers for *P. margaritifera*, and tested their utility by assessing population structure, genetic diversity, as well as detecting regions of the genome underlying functional differences among populations. It involved 156 Fijian oysters sampled from three wild, and one hatchery produced population. Shallow but significant genetic structure was revealed among all wild populations (average pairwise F_{st} = 0.046), with clear evidence of a genetic bottleneck in the hatchery population (N_{e}LD = 6.1), compared to wild populations (N_{e}LD >192.5). F_{st} outlier detection to

differentiate individuals between the orange and black tissue colour morphotypes characteristic of this species, revealed 42-62 highly differentiated SNPs ($p<0.02$), while case-control association discovered up to 152 SNPs ($p<0.001$). Database searches revealed that five of these SNPs were associated with a melanin biosynthesis pathway, demonstrating their biological relevance. This investigation demonstrated the utility of genome-wide SNP data for assessment of genetic structure and diversity in *P. margaritifera*, with transferability to other highly-dispersive marine taxa for their conservation and management.

The second investigation utilised 4,123 genome-wide SNPs, together with an independent hydrodynamic particle dispersal model to assess genetic structure, diversity, local adaptation and population connectivity at 6 farm and 5 wild Fijian sites. Weak fine-scale patterns of population structure indicative of broad-scale admixture were observed among wild oysters, while a hatchery-sourced farmed population exhibited a higher degree of genetic divergence (hatchery oysters cf. all other populations F_{st}=0.085–0.102). This hatchery-produced population had also experienced a bottleneck (N_{eLD}=5.1; 95% C.I.=[5.1-5.3]); compared to infinite N_{eLD} estimates for all wild oysters. Simulation of larval transport pathways confirmed the existence of broad-scale admixture by surface ocean currents, correlating well with fine-scale patterns of population structure discovered. F_{st} outlier tests failed to detect genetic signatures supportive of selection, with only 2-5 directional outlier SNPs identified (average F_{st}=0.116). The lack of biologically significant population genetic structure, absence of evidence for local adaptation and larval dispersal simulation, all indicated the existence of a single genetic stock of *P. margaritifera* in the Fiji Islands for management purposes. The combined use of independent high resolution genomic and oceanographic data as demonstrated here is a novel approach that can be applied to other broadcast spawning taxa.

The third investigation examined the microevolutionary forces influencing genetic structure, connectivity and adaptive variation across the ~18,000 km Indo-Pacific distribution of *P. margaritifera*. Concordance with a theoretical population model known as the Core-Periphery Hypothesis (CPH), was used as a framework for this assessment. The CPH predicts that genetic diversity is expected to be highest at the centre of a species' distribution, progressively decreasing with increased differentiation towards outer range limits, as populations become increasingly isolated, fragmented and locally adapted. Analyses utilising 9,624 genome-wide SNPs and 580 oysters sampled from 14 sites, discovered differing patterns of significant and substantial broad-scale genetic structure between the Indian and

Pacific Ocean basins. Indian Ocean populations were markedly divergent (F_{st}=0.253-0.418, p<0.001), compared to Pacific Ocean oysters, where basin-wide gene flow was much higher (F_{st}=0.001-0.109, p<0.001). Visualisation of population structure at selectively neutral loci resolved three and five discrete genetic clusters for the Indian and Pacific Oceans respectively, while evaluation of genetic structure at adaptive loci for Pacific populations (89 SNPs under directional selection; F_{st}=0.101-0.437, FDR=0.05), revealed five clusters identical to those detected at neutral SNPs, suggesting environmental heterogeneity within the Pacific. Patterns of structure and connectivity were supported by Mantel tests of isolation by distance (IBD) and independent hydrodynamic particle dispersal simulations. These findings have revealed that population organisation in this species is highly complex and far more elaborate than generalised CPH predictions, with structuring being produced by the interaction of ocean currents, IBD and seascape features at a broad scale, together with habitat geomorphology and local adaptation at regional levels.

The fourth and final investigation examined evolutionary relationships and the taxonomic identity of *P. margaritifera*. This study was required as the current species classification is not supported by molecular data, and includes a total of six subspecies that are described exclusively using morphological characters. Here, 69 oysters were sampled from 14 populations in both the Indian and Pacific Oceans. Samples were also collected from the congeneric taxa *P. maxima* and *P. mazatlanica* (n=29 and n=10, respectively), and phylogenetic reconstruction carried out using both 8,308 genome-wide SNPs and 10,000 dominant loci. Reconstructions using neighbour-joining (Nei's 1972 unbiased distances), maximum likelihood and Bayesian approaches all indicate that the taxonomy of *P. margaritifera* is more complex than previously indicated, with distinct evolutionary significant units (ESUs) identified within Tanzanian and Iranian populations, correlating with type localities for two Indian Ocean morphological subspecies descriptions. Contrastingly, phylogenies generated for Pacific Ocean *P. margaritifera* resolved a large monophyletic clade, suggesting little support for two of three morphological subspecies classifications reported from this ocean basin. Furthermore, *P. mazatlanica* specimens all formed a basal clade closest to French Polynesian *P. margaritifera*, suggesting it may not constitute a separate species. Collectively, these findings provide evidence to support a suggestion by previous studies that *P. margaritifera* comprises a species complex; however, further investigation involving finer-scale sampling with higher sample densities is required to resolve regional ESU boundaries.

Collectively, this thesis presents the most comprehensive evaluation of genetic structure, population connectivity and evolutionary relationships for *P. margaritifera* to date. The data generated have permitted fundamental insights into the stock and taxonomic structure of this species, which are invaluable for its sustainable fishery management and aquaculture, with extension to other taxa possessing similar biological attributes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF PLATES

LIST OF ADDITIONAL FILES

Please note that all additional files are presented as .gif format animations, and need to be opened in a web browser to display correctly.

Additional file 1. Animation of full particle dispersal model simulation run for 2009 over 100 days. Particle seed location colour codes are identical to those described in **Fig 3.4**.

Additional file 2a. Animation of particle dispersal model simulation using 2014 HYCOM data for spawning season 1. Particle seed location colour codes for 11 populations are identical to those described in **Figure 4.1**.

Additional file 2b. Animation of particle dispersal model simulation using 2014 HYCOM data for spawning season 2. Particle seed location colour codes for 10 populations are identical to those described in **Figure 4.1**.

Additional file 2c. Animation of particle dispersal model simulation using 2015 HYCOM data for spawning season 1. Particle seed location colour codes for 11 populations are identical to those described in **Figure 4.1**.

Additional file 2d. Animation of particle dispersal model simulation using 2015 HYCOM data for spawning season 2. Particle seed location colour codes for 10 populations are identical to those described in **Figure 4.1**.

APPENDICES

ABBREVIATIONS

ACIAR	Australian Centre for International Agricultural Research
ACT	Australian Capital Territory
ADO	Allelic dropout
AFLP	Amplified fragment length polymorphism
AGRF	Australian Genome Research Facility
AMOVA	Analysis of Molecular Variance
BIC	Bayesian Information Criterion
BLAST	Basic Local Alignment and Search Tool
cBot™	Illumina robotic cluster generation system
COI	Cytochrome c oxidase subunit I
CPH	Core-periphery hypothesis
CU	Conservation Unit
DArT PL	Diversity Arrays Technology Ltd
DAPC	Discriminant Analysis of Principal Components
ddRADseq	double-digest Restriction site-Associated DNA sequencing
DNA	Deoxyribonucleic acid
DVM	Dorsoventral measurement
DMSO	Dimethyl sulfoxide (C_2H_6OS)
dsDNA	Double stranded deoxyribonucleic acid
ENSO	El Niño Southern Oscillation
EtBr	Ethidium bromide ($C_{21}H_{20}BrN_3$)
ESU	Evolutionary significant unit
FAO	Food and Agriculture Organisation of the United Nations
FDR	False discovery rate
GBR	Great Barrier Reef
GBS	Genotyping by sequencing
gDNA	genomic deoxyribonucleic acid
GO	Gene ontology
GWAS	Genome-Wide Association Study
HL	Homozygosity by locus
HWE	Hardy-Weinberg equilibrium
HYCOM	Hybrid Coordinate Ocean Model
IBD	Isolation by distance
IBS	Identity-by-similarity
IR	Internal relatedness
ITS	Internal transcribed spacer DNA region
IUPAC	International Union of Pure and Applied Chemistry
JCU	James Cook University, Australia
k-NN	Number of nearest neighbours k-threshold
LD	Linkage disequilibrium
MAF	Minor allele frequency
MAS	Marker-assisted selection
MEEL	Molecular Ecology and Evolution Laboratory, JCU
MCMC	Markov Chain Monte Carlo algorithm
MLH	Multi-locus heterozygosity

mRNA	Messenger ribonucleic acid
mtDNA	Mitochondrial deoxyribonucleic acid
MU	Management unit
MUSCLE	MUltiple Sequence Comparison by Log-Expectation
NCBI	National Centre for Biotechnology Information (U.S.A.)
NGS	Next-generation sequencing
NJ	Neighbour-joining
PAV	Presence-absence variant
PCR	Polymerase chain reaction
PE	Paired end
PhD	Doctor of Philosophy
PIC	Polymorphic Information Content
PLD	Postlarval duration
PNG	Papua New Guinea
PSRF	Potential Scale Reduction Factor
QQ-plot	Quantile-Quantile plot
QTL	Quantitative Trait Locus
RADseq	Restriction site-Associated DNA sequencing
RE	Restriction enzyme
RRL	Reduced-representation locus
rRNA	Ribosomal ribonucleic acid
SD	Standard deviation
SH	Standardised heterozygosity
SNP	Single nucleotide polymorphism
SPC	Secretariat of the Pacific Community
U	Unit measurement of restriction enzyme

CHAPTER 1: General introduction

1.1 Stock assessment, fishery management and aquaculture of marine species

Fisheries around the world that have been commercially exploited for decades or centuries are in a state of peril, with several stocks either on the brink of collapse, or harvested beyond the point of recovery (Pauly *et al.* 2002; Reiss *et al.* 2009). Taxa such as the Atlantic bluefin tuna (*Thunnus thynnus*) are overfished to the point where the species is threatened with regional extirpation (Fromentin *et al.* 2014), while many others remain heavily data deficient (Lewison *et al.* 2009). This is especially true for many invertebrates and by-caught organisms that may have low commercial, but high ecosystem values; and unfortunately for many taxa, their conservation status remains unknown (Lewison *et al.* 2009; Sims *et al.* 2008).

Given the finite nature of global marine resources, continued catch declines are projected to have dire consequences for food security and continued livelihoods for much of the world's population (FAO 2015; Pauly & Zeller 2016). In order to arrest the overexploitation of vulnerable taxa, and better manage those still supported by healthy populations, much emphasis in recent years has been placed on sustainable fishery exploitation (Pauly *et al.* 2002). However, the long term viability of any fishery depends to a large extent on how well the biological, ecological and socio-political aspects of the system in question are understood, and appropriately managed (Patterson *et al.* 2001; Pons *et al.* 2016).

The success or failure of many fishery management programmes depends upon sound scientific information assembled from multiple sources on the status of the species concerned, and its translation into appropriate management action and monitoring efforts (Cotter *et al.* 2004). Of particular importance is information on the spatial boundaries of stocks, their demography, ecology, as well as evolutionary characteristics (Laugen *et al.* 2014; Ovenden *et al.* 2015). A stock in this context is defined as "a group of individuals that sustains itself over time, and that responds in a similar way to environmental changes within a discrete geographic area" (Gosling 2015). While traditional stock assessment tools are highly capable of providing demographic information on populations (e.g. growth and mortality rates), evolutionary data such as genetic structure, effective population size and

local adaptation to specific environments also warrant careful consideration (Carvalho & Hauser 1994; Reiss *et al.* 2009).

Genetic tools have great potential to assist fishing industries in maintaining productive and sustainable harvests, by providing evolutionary information. Molecular data can be applied to a broad range of fishery management and monitoring tasks, from determining basic stock structure for policy development and performing ecosystem monitoring, to disease surveillance and gauging resilience against anthropogenic and environmental pressures (Ovenden *et al.* 2015). For many marine species which possess pelagic larvae with high dispersal capabilities, quite often a practical understanding of population connectivity for stock management may only be realised through genetic analyses (see Chapter 2). Additionally, for species that are data deficient, or in situations where implementation of traditional fishery assessment methods are logistically or financially challenging, molecular surveying techniques can provide answers to questions of stock boundaries, or local adaptation, while using comparatively fewer resources (Ovenden *et al.* 2015; Reiss *et al.* 2009; Ward 2000).

Many considerations for management of wild fisheries also apply to aquaculture, as natural populations are often relied upon as sources of broodstock and/or seed. While several important aquaculture species (e.g. salmonids) are truly domesticated (complete closure of life cycle in captivity), many more are only a generation or two removed from their wild counterparts (Domingos *et al.* 2013; Stickney 2005; Yue 2013). Escapes of farmed individuals grown in close proximity to natural populations, and consequent "genetic pollution" of wild individuals through interbreeding and introgression are among further concerns raised (Noble *et al.* 2014). Investigation and development of solutions to address these issues all either require, or can benefit from, insights provided by molecular tools, and a multidisciplinary approach involving both genetic and traditional fishery assessment and monitoring techniques (e.g. mark-recapture studies or physical surveys), is warranted in many cases.

1.2 Knowledge gaps and challenges for fishery assessment in marine invertebrates

Marine invertebrates constitute a considerable proportion of world capture fisheries, totalling over 21% by value of all fishery products traded in 2006 (Anderson *et al.* 2011; Thorpe *et al.*

2000). Unfortunately, many invertebrate fisheries are typically not assessed, remain unmonitored and are often unregulated (Anderson *et al.* 2011); despite their intrinsic, social and high ecosystem values. Aquaculture of several invertebrate taxa is also substantial, totalling 36% by value in 2014, of which crustaceans (primarily prawns and shrimp) as well as molluscs, comprised the largest share (FAO 2015).

Aquatic invertebrate taxa present a number of challenges for fishery assessment and management efforts, as they typically possess large population sizes, wide natural distributions and larvae which spend varying amounts of time developing in the plankton (Broquet *et al.* 2013; Strathmann & Strathmann 2007). For many invertebrates, the dispersive phase of the life cycle is often the larva, while adults are either completely sessile (e.g. barnacles, sponges and ascidians), or possess very limited mobility (e.g. sea urchins, bivalves, gastropods). Conversely, several species are migratory, both as adults and larvae (e.g. squid and lobsters), with population characteristics similar to vagile finfish taxa (Reitzel *et al.* 2013; Thorpe *et al.* 2000). For species in which larvae are primarily responsible for transporting genetic material between populations, a major problem for stock assessment arises when the biology of the larval stage is poorly understood, or as is the case for many species, still remains to be identified (Thorpe *et al.* 2000). For low gene flow species, where both larva and adult are not very mobile, dispersal may be restricted, and often localised; resulting in low stock sizes. Taxa with low stock sizes may also be more vulnerable to overfishing than high gene flow species, and require a separate set of management actions (Thorpe *et al.* 2000).

Cryptic speciation is an additional challenge for stock assessment in marine invertebrate taxa, even for species such as squid that are commercially important, and therefore comparatively better studied than some other taxa (Knowlton 1993; Solé-Cava & Thorpe 1991). While this situation may simply reflect difficulties in investigating the taxonomy of generally soft-bodied organisms with few hard parts or obvious distinguishing features, it does underscore the need for genetic data in fishery assessment exercises to correctly identify discrete management units (MUs/stocks), as well as evolutionary significant units (ESUs) (Thorpe *et al.* 2000).

Stock assessments of bivalve molluscs are affected by many of the issues described above, but can also present additional problems, among which is the patchy distribution of

individuals. This can arise from highly variable patterns of larval recruitment, predation and competition, as well as the influences of natural and anthropogenic disturbances (Gosling 2015). The effect of heterogenous individual distribution often means that traditional bivalve stock assessment surveys typically require extensive long-term sampling programmes to determine distribution and abundance, which in most situations can be costly and impractical (Cotter *et al.* 2004; Gosling 1982, 2015; Gosling & Wilkins 1985).

It is clear from previous research that the biological and physical mechanisms driving population structure and connectivity in bivalves and other marine invertebrates are exceedingly complex, and require systematic investigations to correctly identify MUs and ESUs (Funk *et al.* 2012). The incorporation of high resolution genetic data here is essential for fishery assessment and aquaculture practices, as apart from stock and species identification, it also permits assessment of adaptive variation, and evaluation of demographic parameters (e.g. growth and survival rates). The latter has particular application for aquaculture, through the improvement of commercial traits of interest by selective breeding as well as disease investigation (Thorpe *et al.* 2000; Yue 2013; Zhong *et al.* 2014).

1.3 Aquaculture and fishery management of the black-lip pearl oyster

The black-lip pearl oyster *Pinctada margaritifera*, is a marine bivalve mollusc that has a broad Indo-Pacific distribution (**Figure 1.1**), and is highly valued for cultured pearl and pearl shell production (Southgate *et al.* 2008; Wada & Tëmkin 2008). Aquaculture of this species constitutes a valuable industry and is an important source of coastal community livelihood across almost the entire extent of its distribution (Southgate *et al.* 2008; SPC 2003). Despite its value, very little information is available on the stock structure of the species outside of the central and western Pacific (Wada & Jerry 2008).

Previous studies examinining contemporary population structure and connectivity in *P. margaritifera*, have indicated mixed findings using a range of different marker types (allozymes, mtDNA and microsatellites), and have largely been limited to the Cook Islands and French Polynesia (Arnaud-Haond *et al.* 2003a; Arnaud-Haond *et al.* 2004; Arnaud-Haond *et al.* 2008; Benzie & Ballment 1994; Durand & Blanc 1988; Durand *et al.* 1993; Lemer & Planes 2012, 2014). Interestingly, genetic heterogeneity at small spatial scales in the presence of high gene flow, sometimes referred to as "genetic patchiness" has been described

in French Polynesia, as well as both weak and strong population structuring from restricted gene flow between "open" and "closed" atoll lagoon hydromorphologies, respectively (Lemer & Planes 2014).

Very little is known about contemporary stock structure and connectivity, or evolutionary relationships across the broader distribution of the species. The current species description is exclusively on the basis of morphological characters (Jameson 1901), and recognises a total of six sub-species (Gervis & Sims 1992). In the Pacific basin, Hawaiian populations are known as *P. margaritifera galstoffi* (Bartsch), Cook Islands and French Polynesian individuals *P. m. cummingi* (Reeve), and all central and western Pacific specimens *P. m. typica* (Linnaeus). Indian Ocean populations are represented by *P. m. persica* (Jameson; Persian Gulf), *P. m. erythraensis* (Jameson; Red Sea) and *P. m. zanzibarensis* (Jameson; East Africa, Madagascar and Seychelle Islands). A seventh sub-species (*P. margaritifera mazatlanica*), was subsequently designated a separate species; *P. mazatlanica* (Hanley, 1856), following the original description (Cunha *et al.* 2011; Wada & Tëmkin 2008).

The six sub-species of *P. margaritifera* are closely associated with regional geographic locations, and require confirmation using molecular techniques to clarify the taxonomy of the species across its broader natural range. Given the phenotypic plasticity and morphological diversity present in Pteriid pearl oysters (Masaoka & Kobayashi 2005a, b), and that some shell shape and size differences are apparent between populations of *P. margaritifera* (Allan 1959; Wada & Tëmkin 2008), insights gained from molecular data are expected to shed light on the number of ESUs present. Because of the aforementioned intraspecific variation that has been documented, several studies have suggested that *P. margaritifera* might in fact constitute a species complex (Cunha *et al.* 2011; Ranjbar *et al.* 2016; Tëmkin 2010; Wada & Tëmkin 2008).

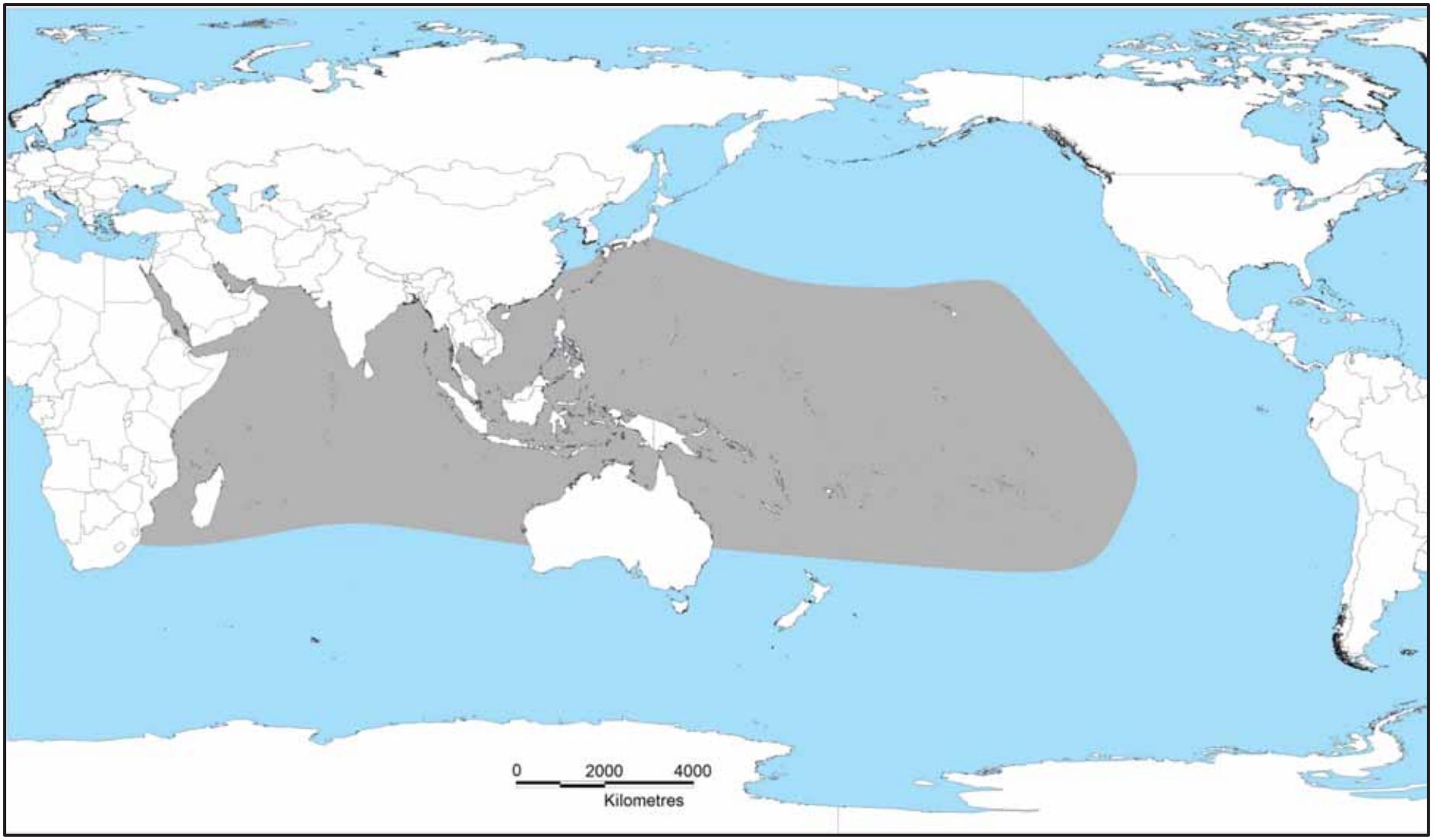


Figure 1.1 Map of the approximate known natural distribution of *P. margaritifera*. The range limits and extent of distribution are presented in grey, and adapted from Wada and Tëmkin (2008).

For the Fiji Islands, cultured pearls and pearl shell production from *P. margaritifera* is a valuable industry and substantial source of coastal community livelihoods. It produces a high-value, low-volume and non-perishable product with a comparatively smaller environmental footprint than most other forms of aquaculture, making it an ideal export commodity for developing Pacific island countries (Southgate *et al.* 2008; SPC 2003, 2007). The industry is almost exclusively dependent on wild oysters for which there are currently no comprehensive fishery management guidelines, and therefore no information is available on the number of discrete populations present, their levels of genetic fitness and relatedness, or if domestic translocation of animals is suitable for the establishment of new pearl farms.

Preliminary stock assessment using transect surveys reported low abundances of *P. margaritifera* at all locations examined, and recommended immediate conservation efforts to increase population densities of wild oysters (Friedman *et al.* 2010; Passfield 1995). If the Fijian pearling industry is to grow and sustain itself, genetic stock assessment is required to gauge the status of populations, and provide biological data upon which responsible aquaculture and fishery management policy can be based.

1.4 Thesis overview

The overarching goal of the research undertaken was to investigate the genetic structure, population connectivity and taxonomy of the black-lip pearl oyster, to inform fishery management and aquaculture practice across the extent of its Indo-Pacific distribution, with a particular focus on the Fiji Islands. This goal was achieved through targeting four primary objectives, which were to:

i. Develop novel genome-wide Single Nucleotide Polymorphism (SNP) molecular markers for *P. margaritifera*, and utilise them for performing a wide range of population genomic and phylogenomic analyses, as detailed in objectives (ii), (iii) and (iv) below.
ii. Investigate the genetic structure, diversity, connectivity and adaptive variation of populations in the Fiji Islands, for informing fishery management and aquaculture development of the Fijian pearling industry.

iii. Examine the genetic structure, diversity and connectivity of *P. margaritifera* across its Indo-Pacific distribution, to gain insights into the processes organising and regulating populations at both the global and regional scales, and
iv. Evaluate the taxonomic identity of the species from distribution-wide samples, using phylogenomic reconstruction methods.

Each of these objectives translate into separate investigations, which are presented as independent data chapters in this thesis.

Chapter 2 assesses the utility of ddRADseq genotyping for detecting genome-wide SNPs in *P. margaritifera*, and describes the development of a set of SNP markers for a range of population genomic and association analyses. It also outlines an optimised ddRADseq laboratory and analysis workflow for this species, and tests it through investigation of population structure, genetic diversity, F_{st} outlier detection and a genome-wide association study (GWAS) in four closely-related Fijian populations.

Chapter 3 utilises the genome-wide SNP resource developed in Chapter 2 to assess the stock structure, genetic diversity, adaptive variation and connectivity of *P. margaritifera* in the Fiji Islands, for fishery and aquaculture management. A molecular and biophysical approach is adopted, using population genomic analyses, and an independent hydrodynamic particle dispersal simulation to infer larval connectivity between sampling sites. The findings of this study and its implications for fishery management and aquaculture of Fijian *P. margaritifera* are explored and discussed, as well as the utility of the analytical framework for stock assessment in other broadcast-spawning marine taxa.

Chapter 4 assesses the genetic structure, connectivity and adaptive variation of *P. margaritifera* across the extent of its Indo-Pacific distribution, spanning over 18,000 km. Here, population genomic data and hydrodynamic dispersal simulation are used to elucidate stock structure, connectivity and local adaptation within the Indian and Pacific Ocean basins. This study sheds light on the links between genetic structure, ecology and oceanography in this species, to reveal how populations of a highly-dispersive marine invertebrate occupying expansive, heterogenous habitats are organised and maintained.

Chapter 5 investigates evolutionary relationships within *P. margaritifera*, and provides a benchmark for review of its taxonomic identity. This investigation is the first to assess samples collected over the species' distributional range, and results are discussed in the contexts of taxonomic identity, and regional spatial marine management strategies for conservation and aquaculture efforts.

1.5 Publications arising from this thesis

Peer reviewed scientific journal articles produced over the course of this PhD candidature which are either published[1], undergoing peer review or in preparation at the time of writing are as follows:

Lal, M.M., Southgate P.C., Jerry D.R., Zenger K.R. (2016) Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. *Marine Genomics* **25**, 57-68.

Lal, M.M., Southgate P.C., Jerry D.R., Bosserelle C., Zenger K.R. (2016) A parallel population genomic and hydrodynamic approach to fishery management of highly-dispersive marine invertebrates: the case of the Fijian black-lip pearl oyster *Pinctada margaritifera*. *PloS ONE* **11**, e0161390.

Lal, M.M., Southgate P.C., Jerry D.R., Bosserelle C., Zenger K.R. Swept away: ocean currents and seascape features influence genetic structure across the 18,000 Km Indo-Pacific distribution of a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. Submitted to *BMC Genomics*.

Lal, M.M., Southgate P.C., Jerry D.R., Zenger K.R. (unpubl.) Fresh evidence for the presence of a species complex in the black-lip pearl oyster *Pinctada margaritifera* (Bivalvia: Pteriidae). Submitted to *Scientific Reports*.

[1] Reprints of all published literature have been included as Appendix 5.1

Lal M.M., Southgate P.C., Jerry D.R., Bosserelle C., Zenger K.R. (unpubl.) Range-wide larval connectivity of the black-lip pearl oyster *Pinctada margaritifera* using hydrodynamic dispersal simulation. In preparation for *PloS ONE*.

CHAPTER 2: Development and evaluation of genome-wide SNP markers for population genomic analyses in *P. margaritifera*

2.1 Introduction

Broadcast spawning marine organisms that undergo prolonged planktonic larval development can achieve high rates of connectivity over large spatial scales. As a result, levels of population genetic differentiation in these taxa are often low, but can also be biologically relevant (Limborg *et al.* 2012). Furthermore, selective forces can still impact local populations even with rates of high gene flow (Nielsen *et al.* 2009). In order to detect fine-scale variability and the presence of local adaptation which may be overlooked using traditional molecular markers, new genome-wide genetic resources incorporating thousands of SNP loci are typically required (Krück *et al.* 2013; Limborg *et al.* 2012; Nielsen *et al.* 2009; Palumbi 2003; Pujolar *et al.* 2014). The detection of fine-scale structure and signatures of selection are important for delineating conservation and management units (CUs and MUs; Funk *et al.* 2012) for both wild and captive species, as populations (see Waples & Gaggiotti, 2006) may otherwise appear homogenous in their organisation and distribution.

To reveal these underlying differences, marker sets with sufficient sensitivity and resolving power are required to accurately inform conservation and fishery management efforts, but are also necessary in aquaculture applications. Genome-wide SNPs are powerful and highly versatile markers capable of addressing this problem, and are gaining broader use in population genomic and phylogenetic investigations, as well as for understanding genome structure and genome-wide association studies (GWAS; Davey *et al.* 2011; Helyar *et al.* 2011; Kai *et al.* 2014; Pool *et al.* 2010; Rasic *et al.* 2014). Given the availability of such marker sets, genome mapping and association studies are now real possibilities for a number of established species in aquaculture, including quantitative trait locus (QTL) identification and marker assisted selection (MAS; (Allendorf *et al.* 2010; Angeloni *et al.* 2012; McAndrew & Napier 2011; Yue 2013).

The advent of Restriction-site Associated DNA (RADseq) genotyping methods combined with next generation sequencing (NGS) technologies, has enabled the delivery of genome-

wide SNPs for both model and non-model organisms, at considerably lower costs and less hands-on time investment than traditional techniques (Davey & Blaxter 2011; Davey *et al.* 2011; Nielsen *et al.* 2011). A recent RADseq-based genotyping method is double digest RADseq (ddRADseq), which permits precise selection of the proportion of the genome required for sequencing and maximises the number of sequence reads incorporated in the analysis, without the need for a reference genome (Hohenlohe *et al.* 2011; Kai *et al.* 2014; Peterson *et al.* 2012). This is particularly important for non-model marine organisms which include many species used for aquaculture (McAndrew & Napier 2011), where genomic resources are often lacking, or absent altogether.

Despite its popularity, ddRADseq along with other RADseq methodologies have seen limited application to marine invertebrate taxa due to challenges such as a lack of *a priori* knowledge of genome size, variability and the frequency of restriction enzyme cut sites (Toonen *et al.* 2013). This is especially true for marine bivalves, as their genomes are characterised by a high density of repetitive non-coding regions and increased levels of polymorphism (Suárez-Ulloa *et al.* 2013). To date, draft genomes for only two bivalve species are available; which include the Pacific oyster *Crassostrea gigas* and the Akoya pearl oyster *Pinctada fucata* (Takeuchi *et al.* 2012; Zhang *et al.* 2012).

Given the increased occurrence of genome-wide polymorphism rates characteristic of marine bivalves (e.g. up to 1 SNP per 40 base pairs reported for *C. gigas*; (Harrang *et al.* 2013; Hedgecock *et al.* 2005; Zhang *et al.* 2012), variability at restriction enzyme cut sites may erode the genotyping power of RADseq approaches by altering fragment distributions and causing allelic drop outs (ADO; (Davey & Blaxter 2011; Huang & Knowles 2014). For genomic investigations involving bivalves and other highly polymorphic taxa, determination of the extent to which genotyping efficiency is affected and thorough analysis to maintain data integrity are required.

The black-lip pearl oyster *Pinctada margaritifera* is the basis of a valuable aquaculture industry in the Fiji Islands that is almost exclusively dependent on oysters collected from the wild. However, there are currently no comprehensive fishery management guidelines nor genomic resources available for this species, as a country-wide stock assessment has never been undertaken. This chapter examines the suitability of ddRADseq genotyping for detecting genome-wide SNPs in *P. margaritifera*, and assesses the utility of the marker set

for a range of population genomic and association analyses. In addition, it provides an optimised ddRADseq methodology for this species, including analysis recommendations for high-quality SNP data recovery in this taxon.

To demonstrate the power and versatility of the genome-wide SNPs discovered, population structure and variability were investigated in four closely-related Fijian populations. Furthermore, genome-wide association testing and F_{st} outlier detection were used to distinguish between individual oysters belonging to the two different tissue colour morphotypes characteristic of this species, within separate populations. Finally, the suitability of the genome-wide SNPs discovered was assessed for their application to further population genomic investigations involving this highly important bivalve mollusc, and to illustrate the applicability of ddRADseq for similar analyses in other highly dispersive marine taxa.

2.2 Methods and materials

2.2.1 Specimen collection, tissue sampling and DNA extraction

Adult *P. margaritifera* (n=156) sized between 7-15 cm in dorso-ventral measurement (DVM) were collected from four Fijian sites representing one hatchery-produced population (n=25), and three populations of wild oysters (n=50, 32 and 49 at Namarai, Raviravi and Savusavu respectively). All wild oysters were sampled either directly from natural coral reef habitats or spat collectors (Southgate 2008); see **Figure 2.1** for a site map). Proximal mantle and adductor muscle tissues (3 and 6 cm respectively) were removed and transferred to tubes containing 20% salt saturated dimethyl sulfoxide (DMSO-salt) preservative (Dawson *et al.* 1998). Genomic DNA was extracted using a modified cetyl trimethyl ammonium bromide (CTAB, Amresco, cat. #0833-500G) chloroform/isoamyl alcohol protocol, with a warm (30 °C) isopropanol precipitation (Adamkewicz & Harasewych 1996). To clean up all DNA extractions, a Sephadex G50 (GE 2007) spin column protocol was used prior to quantification with a Nanodrop 1000 Spectrophotometer (Thermo Scientific)

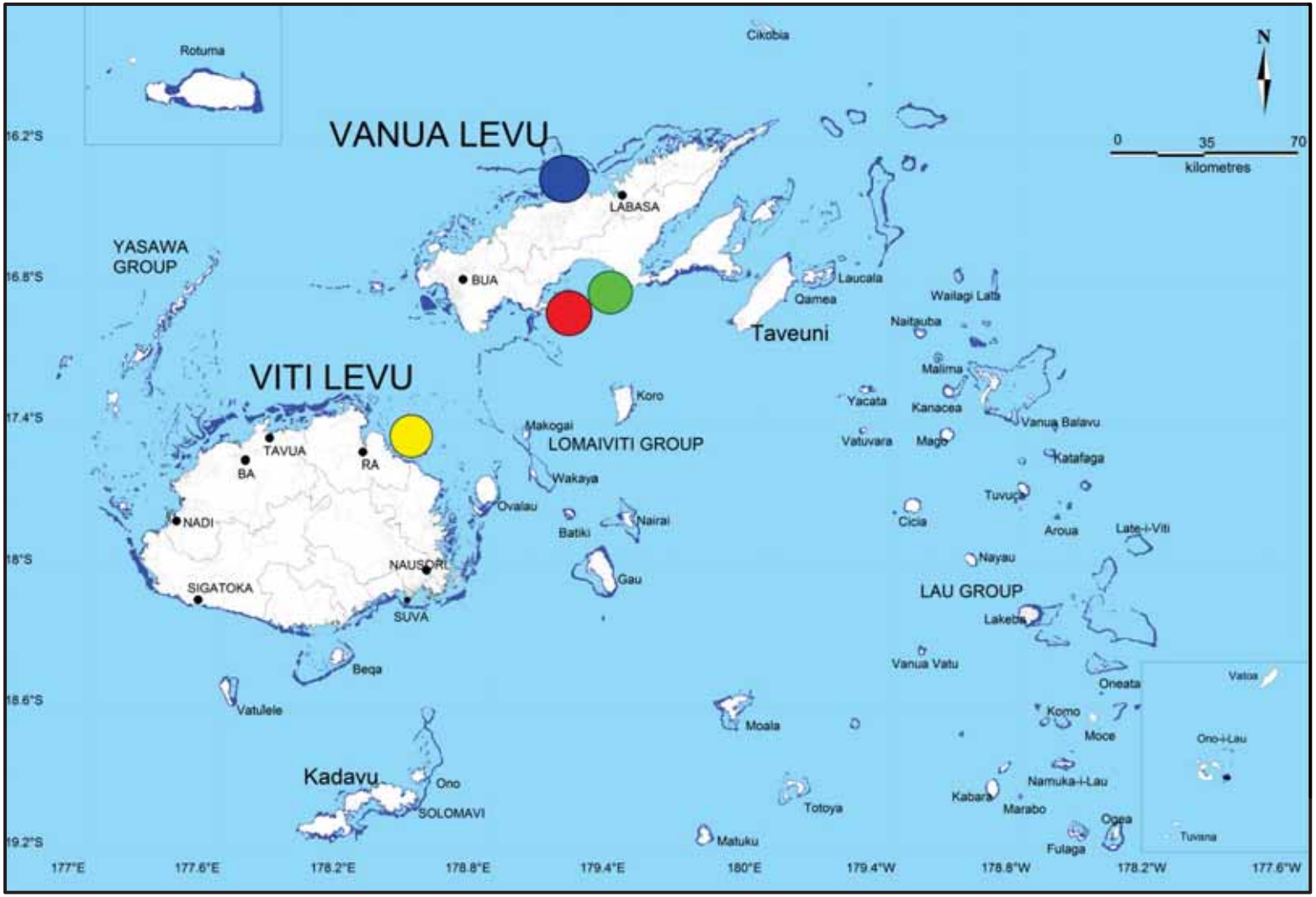


Figure 2.1 Map showing locations and numbers of individuals in the Fiji Islands where populations of *P. margaritifera* were sampled. Reef outlines are presented in dark blue, and the sampling locations are as follows: Raviravi farm wild stock (yellow, n=32), Savusavu farm wild stock (green, n=49), Savusavu farm hatchery stock (red, n=25) and Namarai farm wild stock (blue, n=50).

2.2.2 ddRADseq library preparation, *in silico* size selection and sequencing

Double digest restriction site-associated (ddRAD) libraries were prepared following Peterson *et al.* (2012), with the following modifications. Briefly, after being restriction digested for 18 hr using 10 U each of MspI and EcoRI-HF endonucleases (New England Biolabs, cat. #R0106S and #R3101S) and ~1.5 µg gDNA at 37°C, all samples were cleaned up using Sera-Mag SpeedBeads Carboxylate-modified Microparticles (Thermofisher Scientific cat. # 4515-2105-050250) and quantified with an AccuBlue High Sensitivity dsDNA kit (Biotium, cat. #31006). Custom barcoded adapters (P1 and P2; refer to Peterson *et al.*, 2012 for sequences) were then ligated to 500 ng of digested DNA fragments using a T4 DNA ligase (New England Biolabs, cat. #M0202S).

To determine the optimum fragment size selection window, *in silico* restriction digest simulations were performed on the Pacific oyster *C. gigas* (NCBI GenBank ID 10758) and Akoya pearl oyster *Pinctada fucata* (http://marinegenomics.oist.jp/pinctada_fucata) genomes respectively. Simulations were carried out in the *R* package *SimRAD* (Lepais & Weir 2014) at varying size selection windows, and the results extrapolated to the estimated size of the *P. margaritifera* genome (0.824 Gb). Results indicated 16,141 and 16,938 fragments to be expected at a target window of 490 ± 40 bp, based on the *C. gigas* and *P. fucata* genomes respectively. Selection of this fragment size window was based on sufficient DNA recovery during the size selection step. These results agree well with simulations by Peterson *et al.* (2012) using the same enzyme combination (~17,000 fragments at >7× coverage).

Samples were pooled into sets of 48 according to their P1 adapter barcodes, quantified using a Nanodrop 1000 Spectrophotometer (Thermo Scientific) and size selected at 490 ± 40 bp using a Pippin Prep automated size selector (PIP0001), and 2% agarose gel Pippin Prep cassette with ethidium bromide (Sage Science, cat. #CSD2010). Pooled samples were cleaned with Sera-Mag Streptavidin Coated Microparticles (Thermofisher Scientific) and amplified to generate Illumina sequencing libraries using Phusion high fidelity DNA polymerase (New England Biolabs, cat. #M0530S). The following cycling conditions on a Biorad C1000 thermal cycler were used: initial denaturation at 98°C for 30s, then 14 cycles of 95°C for 15s, 66°C for 30s and 72°C for 45s, followed by a final extension step at 72°C for 10 min. Fragment size distribution and molarity of the amplified libraries were quantified on an Agilent 2100 Bioanalyser with a DNA 12000 chip set (cat. #5067-1508). Five libraries

(48 barcoded and pooled individuals × 5 unique Illumina TruSeq indices) were combined at equimolar ratios for sequencing in a single lane. After cluster generation and amplification (HiSeq PE Cluster Kit V4 cBot™), 100 bp paired-end sequencing was performed on an Illumina HiSeq 2000 platform at the Australian Genome Research Facility (AGRF) in Melbourne, Victoria. Each lane was spiked with a PhiX control at 1% of the total library and all sequence reads demultiplexed on the basis of the Illumina TruSeq indices incorporated on the PCR2 primer.

2.2.3 Sequence quality control, marker filtering and genotype calling

Raw reads obtained following sequencing were processed using FastQC v.0.10.1 (Babraham Bioinformatics 2007), for assessment of sequence quality scores (Phred33), AT/GC content and duplicated or overrepresented sequences. All read filtering and SNP genotyping was carried out using STACKs v.1.20 software (Catchen *et al.* 2013; Catchen *et al.* 2011). The module 'process_radtags' was used to clean the data to remove any reads with an uncalled base, discard reads with a Phred33 quality score >10, as well as rescue barcodes and RAD-tags. The filtered read files were then renamed to reflect sample names, barcodes trimmed and the forward and reverse reads for each individual concatenated into single 195 bp reads using a custom Perl shell script.

The STACKs pipeline core modules ('ustacks', 'cstacks', 'sstacks' 'rxstacks' and 'populations') were used to process all reads, with final genotypes called at a Minor Allele Frequency (MAF) of 2% and a minimum stack depth of 10. The term "stack" here is used to denote a set of sequence reads anchored at a particular genomic locus by restriction enzyme cut sites, after Catchen *et al.* (2011). Parameters for each module were left at their default values except for the maximum base pair distance allowed to align secondary reads to primary stacks (-N=6) in 'ustacks' and the number of mismatches allowed between sample tags when generating the catalogue (-n=5) in 'cstacks'. Primary reads align exactly to create primary stacks in the 'ustacks' module while secondary reads contain mismatches, and are set aside for later alignment to increase stack depth. By increasing the -N and -n parameters from their default settings (-N=3 and -n=3), SNP calling confidence was increased and missing data minimised. The minimum proportions of loci allowed across individuals was set at 20% and across populations at 50% (-r and -p options respectively). In addition, the most informative SNP at each locus was selected, and these were genotyped in at least 10 individuals within a

population to produce robust population differentiation estimates (see Willing *et al.* 2012), and represented in a minimum of 2 populations across the whole dataset (Huang & Knowles 2014).

Furthermore, all loci were screened using Arlequin v.3.5.1.3 (Excoffier *et al.* 2005) across all populations for departure from Hardy-Weinberg Equilibrium (HWE), and removed if deviations were significant after FDR correction ($p<0.00001$), or loci were monomorphic across all populations (Zenger *et al.* 2007a; Zenger *et al.* 2007b). Additionally, sequences associated with all loci were compared against NCBI viral and bacterial sequence databases using Basic Local Alignment Search Tool (BLAST) searches (Johnson *et al.* 2008), to detect contamination which may have occurred during library preparation, and all matching loci excluded from the final dataset.

2.2.4 Assessments of genomic diversity, inbreeding and population differentiation

To estimate genomic diversity within and among populations, allelic diversity indices including average observed (H_o) and average expected heterozygosities corrected for population sample size ($H_{n.b.}$) were computed, together with inbreeding coefficient (F_{is}) values using Genetix v.4.05 (Belkhir *et al.* 1996). The effective population size based on the linkage disequilibrium method (N_{eLD}) was also calculated for each population with NeEstimator v.2.01 (Do *et al.* 2014). Individual multi-locus heterozygosity was also examined, with the average homozygosity by locus (HL), standardised heterozygosity (SH) and internal relatedness (IR) computed for each population with the *R* package *Rhh* (Alho *et al.* 2010). In addition, the average multi-locus heterozygosity (Av. MLH) per population was computed following (Slate *et al.* 2004).

Pairwise F_{st} estimates were calculated using Genetix v.4.05 (Belkhir *et al.* 1996). Visualisation of broad-scale population structure was carried out by performing a Discriminant Analysis of Principal Components (DAPC) in the *R* package *adegenet* 1.4.2 (Jombart 2008; Jombart & Ahmed 2011; Jombart *et al.* 2010). The DAPC was carried out for all loci, and an α-score optimisation used to determine the number of principal components to retain.

2.2.5 Resolution of fine-scale population structure

To reveal any fine-scale genomic stratification between and among all populations, network analysis was carried out using the NetView P pipeline v.0.4.2.5 (Neuditschko *et al.* 2012; Steinig *et al.* 2016). This pipeline constructs an identity-by-similarity (IBS) distance matrix using the PLINK v.2.050 toolset (Purcell *et al.* 2007), and the network is visualised in the Cytoscape v.2.8.3 network construction package (Smoot *et al.* 2011). The IBS matrix and corresponding networks were constructed at various thresholds of the maximum number of nearest neighbour (k-NN) values between 5 and 40. A simulation of parentage analysis to determine if any single parents, parent pairs or parent-offspring pairs were present was also carried out using Cervus v.3.0.7(Kalinowski *et al.* 2007).

2.2.6 Investigation of tissue colour morphotype discrimination

Pinctada margaritifera expresses two distinct phenotypic colour morphs in several parts of its soft anatomy, which are either a deep black or orange-red colour (Hwang & Okutani 2003; Shirai & Nakamura 1994; Wada & Tëmkin 2008; see **Figure 2.2**). Based on observed population phenotypic ratios, this trait is most likely under simple genetic control. To assess the power and utility of the SNP loci for genome-wide association studies, a case-control association test was carried out to discriminate between individuals belonging to the two colour morphotypes within each population. In addition, F_{st} outlier tests were performed on all three wild populations to independently detect loci involved in colour morphotype differentiation.

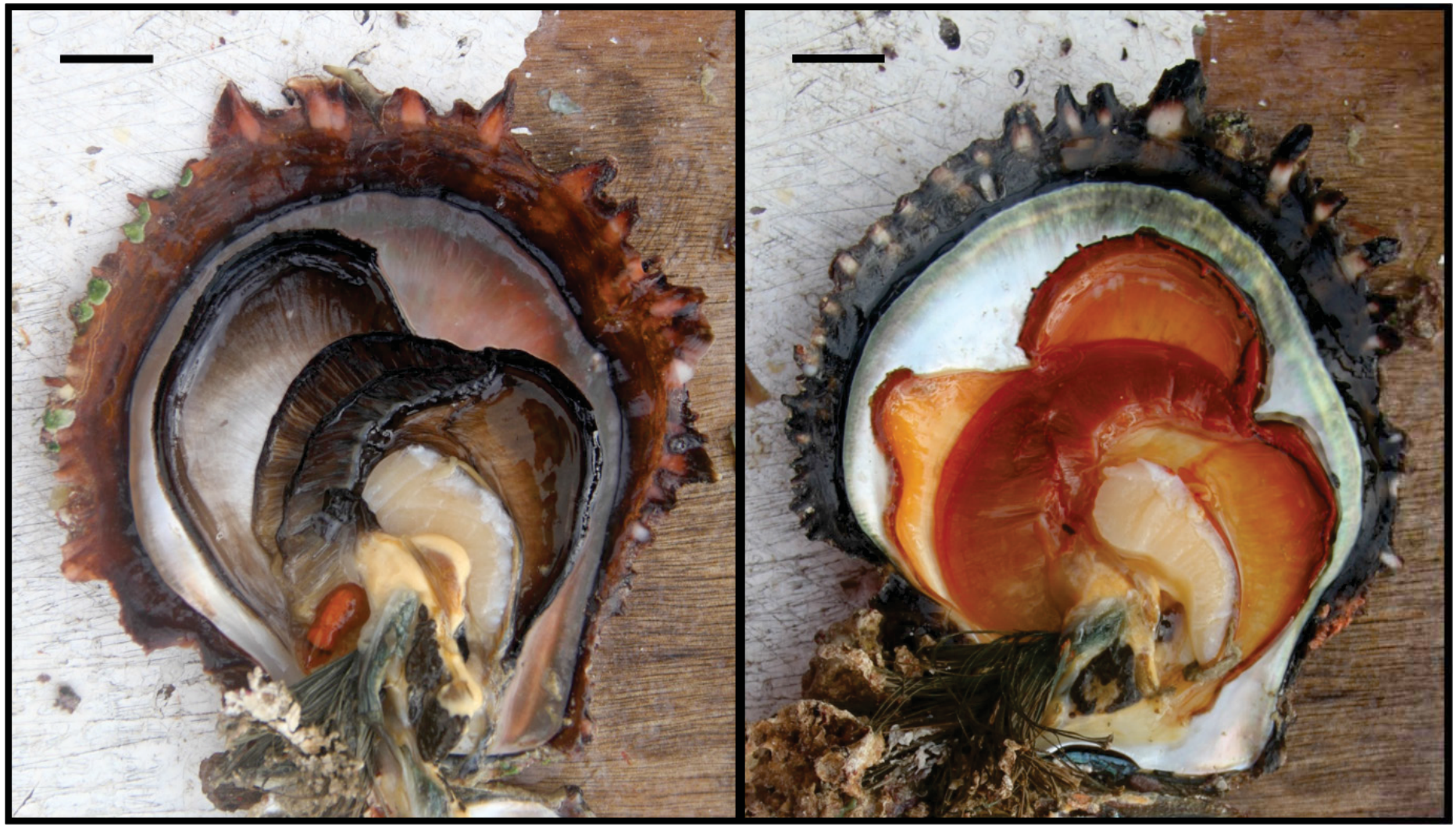


Figure 2.2 Black and orange morphotypes of *P. margaritifera* collected in Savusavu, Vanua Levu, Fiji Islands. Pictured are the left valves of two 20 month old individuals. The scale bars each represent 20 mm.

2.2.7 Case-control association testing, LD testing and gene identity

The genome-wide case-control association test was implemented in the PLINK v.2.050 toolset (Purcell *et al.* 2007) using all four populations separately. Individuals with either orange or black tissue colour phenotypes were assigned as cases or controls respectively. Testing was carried out using the dominant, recessive, allelic and genotypic models within a 95 % confidence interval. Each test was based on a per locus 2 × 3 contingency table and carried out independently, using 100,000 permutations and employing p-value adjustments based on pairwise locus Pearson's Chi-Square tests, to assess departure from the null hypothesis that cases and controls had the same distribution of genotype counts. By default, a stepwise Bonferroni correction was also applied to the test significance values.

To visualise separation of individuals within a population by colour morphotype, unrooted Neighbour-Joining (NJ) distance trees were drawn using MEGA6 (Tamura *et al.* 2013), based on 1-proportion of shared alleles distance matrices constructed in Microsat 1.5d (Minch *et al.* 1995; Zenger *et al.* 2007a). All loci found to have a significant ($p<0.005$) association between the case and control phenotypes were then subjected to an exact test for genotypic Linkage Disequilibrium (LD) in Genepop v. 4.3 (Rousset 2008), using 100,000 dememorisations, 1000 batches and 5000 iterations per batch. The full sequences for all loci determined to be in LD, together with those for all non-linked loci were subsequently retrieved using the 'populations' module of the STACKs pipeline, and then separately searched with BLAST (Johnson *et al.* 2008) queries to determine gene identity terms.

2.2.8 F_{st} outlier detection between colour morphotypes

To independently compare results of the association testing with F_{st} outlier detection, three closely-related wild populations based on geographic distance (Ra, Raviravi and Savusavu wild) were used, and individuals within them separated on the basis of individual tissue colour into "sub-populations" (Johnston *et al.* 2014; Wilding *et al.* 2001). By applying this artificial "selection pressure", any loci demonstrating association with colour would be detected as F_{st} outliers between the "sub-populations". Apart from evaluating suitability of the marker set for outlier detection, these tests also allowed for comparison of overlapping loci discovered between both methods (association testing vs. F_{st} outliers).

Two independent outlier detection methods were used which employ different analytical approaches (Narum & Hess 2011; Pujolar *et al.* 2014; White *et al.* 2010b). The first method was implemented in BayeScan v.2.1 (Foll 2012; Foll & Gaggiotti 2008), and all analyses performed on a 1:10 prior odds probability for the neutral model and commenced with 20 pilot runs consisting of 5,000 iterations each. This was followed by 100,000 iterations with a burn-in length of 50,000 iterations (Foll 2012). Once probabilities had been calculated for each locus, the Bayescan 2.1 function *plot_R* was used in the *R* v.3.2.0 statistical package to identify putative outlier loci at various False Discovery Rates (FDR). A range of FDR values from 0.01 to 0.20 were evaluated based on preliminary testing and recommendations by Ball (2013) and Hayes (2013).

The second method employed the LOSITAN selection detection workbench (Antao *et al.* 2008), which utilises a frequency-based approach to assess relationships between F_{st} and H_e. All LOSITAN outlier detection runs were computed within a 95 % confidence interval under an infinite alleles model, with 50,000 iterations evaluating a range of FDR values from 0.01 to 0.20 to match the BayeScan 2.1 analyses. All other test parameters remained at their default settings, with the exception of the 'Neutral' mean F_{st} and 'Force mean F_{st}' options being enabled.

The results of the Bayescan 2.1 and LOSITAN analyses, together with the construction of Quantile-Quantile plots (QQ-plots) at a 95 % confidence interval were used to assess the suitability of an FDR threshold for outlier detection between the two methods. All QQ-plots were constructed using the *R* package *GWASTools* v.1.14.0 (Gogarten *et al.* 2012), incorporating all SNP loci in one plot and excluding all outliers in another to confirm that the remaining loci conformed to a normal distribution. In order to visualise separation of individuals between colour morphotypes, unrooted NJ trees were redrawn using the methods described earlier (see section **2.2.7**).

2.3 Results

2.3.1 Genotyping and SNP discovery

A total of 295,678,888 PE reads were obtained following sequencing of 156 individuals from the four populations. Following filtering using the STACKs pipeline to remove low quality reads (Phred33 score <30), ambiguous barcodes and overrepresented sequences, 276,918,263 reads remained. These were used to create a catalogue containing 163,106 stacks, for generation of genotypes in all individuals using a median number of 410,675 reads per individual. An average number of 31,122 stacks were assembled for each individual, at an average read depth per stack of 12.2. Further filtering at a minimum read depth of 10 per stack and MAF >0.02 resulted in a total of 21,331 genome-wide SNPs being genotyped (see **Appendix 2.1**). Following final filtering steps (e.g. HWE, individuals/populations with minimum genotyping rates and removal of contaminant sequences), 5,243 high-quality polymorphic genome-wide SNPs remained for population genomic analyses. A total of 2,929 common SNPs were genotyped across all three wild populations. The greatest number of common SNPs was observed between the two most geographically close populations (Namarai-Savusavu - 4,086 SNPs), while the two most geographically distant populations had the least (Raviravi-Namarai - 2,047 SNPs).

2.3.2 Population genomic diversity and differentiation

Inbreeding coefficients (F_{is}) were high in all populations, ranging from 0.3049 to 0.3930 (**Table 2.1**), with low observed heterozygosities (H_o: 0.1076 - 0.1438; $H_{n.b.}$: 0.2568 - 0.3024). Estimates of effective population sizes (N_{eLD}) were robust, and varied from an infinite value for the Raviravi population, to 410.1 and 192.5 for the Namarai and Savusavu wild populations respectively (**Table 2.1**). The Savusavu hatchery population returned the lowest N_{eLD} of 6.1, indicating a genetic bottleneck likely caused by hatchery spawning practices.

To investigate the effects of sequencing depth and missing data on heterozygosity levels, F_{is} and H_o were calculated for each population at a range of missing data thresholds from 10% to 80%, and at read depths ranging from 5 to 15 to ascertain the degree of potential bias. A substantial shift in F_{is} and H_o was seen from a read depth of 7 to 9, with no change beyond a read depth of 10. Furthermore, there were no differences observed in F_{is} and H_o between the

Table 2.1 Genetic diversity indices for the wild and farmed *P. margaritifera* populations sampled. The parameters calculated included the effective population size by the linkage disequilibrium method (N_{eLD}; 95 % confidence intervals indicated within brackets), observed heterozygosity (H_o), average expected heterozygosity corrected for population sample size ($H_{n.b.}$), inbreeding coefficient (F_{is}), average multi-locus heterozygosity (Av. MLH), homozygosity by locus (HL), standardised heterozygosity (SH) and internal relatedness (IR).

Population	**n**	N_{eLD} **[95 % C.I.]**	H_o **(± SD)**	$H_{n.b.}$ **(± SD)**	F_{is} **(p<0.01)****	**Av. MLH (± SD)**	**HL (± SD)**	**SH (± SD)**	**IR (± SD)**
Namarai (wild)	50	410.1 [410.1; 1115.8]	0.1076 (± 0.0992)	0.2568 (± 0.1460)	0.3718	0.0970 (± 0.0176)	0.8075 (± 0.0332)	0.9680 (± 0.1563)	0.3642 (± 0.1011)
Raviravi (wild)	32	∞ [∞; ∞]	0.1096 (± 0.1055)	0.2611 (± 0.1496)	0.3930	0.1038 (± 0.0246)	0.8158 (± 0.0322)	0.9570 (± 0.1674)	0.3792 (± 0.0967)
Savusavu (hatchery)	25	6.1 [5.9; 6.4]	0.1438 (± 0.1541)	0.3024 (± 0.1532)	0.3049	0.1360 (± 0.0164)	0.7836 (± 0.0219)	1.2044 (± 0.1224)	0.2957 (± 0.0723)
Savusavu (wild)	49	192.5 [161.9; 236.5]	0.1084 (± 0.1002)	0.2586 (± 0.1467)	0.3613	0.0957 (± 0.0127)	0.8085 (± 0.0286)	0.9703 (± 0.1349)	0.3637 (± 0.0799)

** Significance threshold for homozygote excesses

Table 2.2 Population pairwise F_{st} estimates for 5,243 SNP loci in individuals of *P. margaritifera* from four Fijian populations using Genetix 4.05 (Belkhir *et al.* 1996).

	Namarai (wild)	Raviravi (wild)	Savusavu (hatchery)
Raviravi (wild)	0.0011	-	-
Savusavu (hatchery)	0.0919*	0.0886*	-
Savusavu (wild)	0.0026	0.0018	0.0929*

* Significantly different values between populations at p<0.05

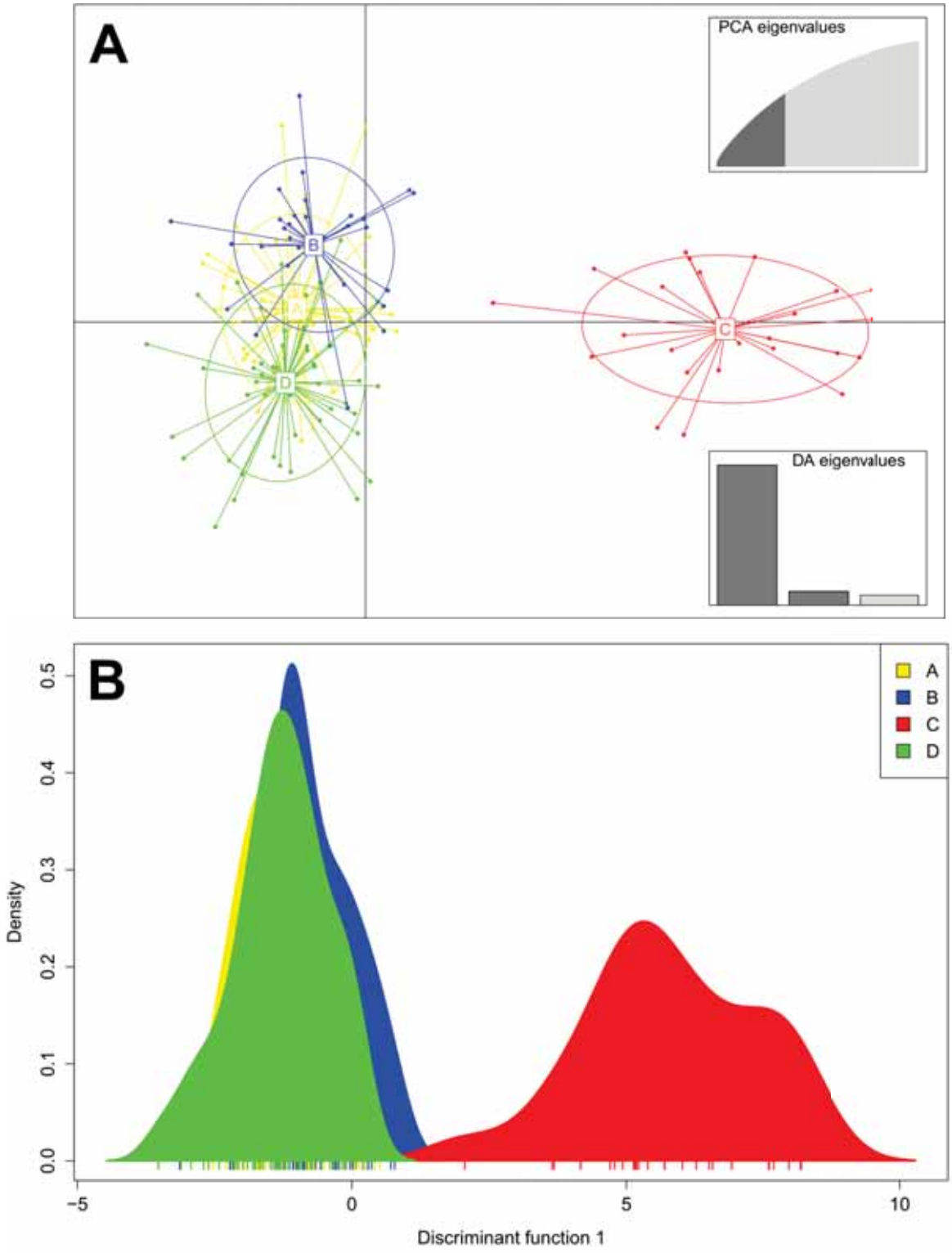


Figure 2.3 **A:** Discriminant Analysis of Principal Components (DAPC) scatterplot drawn using 5,243 SNPs across 156 *P. margaritifera* individuals in the *R* package *adegenet*. Dots represent individuals, with colours denoting sampling origin and inclusion of 95 % inertia ellipses. The Savusavu hatchery population (C) is represented in red, the Savusavu wild population in blue (B), the Namarai wild in yellow (A) and the Raviravi wild in green (D). **B:** A plot of the individual densities against the first discriminant function retained show that the greatest proportion of variation lies with it.

missing data evaluations. To examine the impact of potential restriction site polymorphisms on population differentiation estimates, the final dataset of 5,243 SNPs was filtered to retain only 588 loci which had individual F_{is} values approximating zero. Population pairwise F_{st} estimates calculated as a test using this reduced dataset were not skewed when compared to estimates computed using the full set of markers. Despite the differences in the numbers of SNPs genotyped between populations, the full dataset (5,243 SNPs) remained robust and retained a sufficiently high density of SNPs for downstream population genomic analyses (Andrews & Luikart 2014; Huang & Knowles 2014).

The patterns observed in the average individual multi-locus heterozygosity (MLH) calculations for each population largely matched trends seen in the H_o values (**Table 2.1**), with the Savusavu hatchery population having the highest average MLH (0.1360), while the remaining populations ranged from 0.0970 (Namarai) to 0.1038 (Raviravi). There were no private alleles contained in the final filtered dataset. Similar trends between populations were apparent in the standardised heterozygosity (SH) and average homozygosity by locus (HL) calculations, with the Savusavu hatchery population having the highest SH (1.2044) and lowest HL (0.7836), although both these metrics were generally high with ranges of 0.9570 - 1.2044 for SH and 0.7836 - 0.8158 for HL. Measures of internal relatedness (IR) correlated well with the F_{is} values, ranging from 0.2957 - 0.3792.

2.3.3 Broad and fine-scale population structure

Examination of population pairwise F_{st} values indicated a pattern of broad-scale gene flow, with the average F_{st} among all populations being 0.046 (**Table 2.2**). Despite the low level of population structuring, the Savusavu hatchery population appears to be weakly differentiated by having slightly, but significantly ($p<0.05$) higher F_{st} values of approximately 0.09 compared to the other three populations, which ranged from 0.0011 between Raviravi and Namarai, to 0.0026 between Namarai and Savusavu wild.

Visualisation of broad-scale population structure using a DAPC (**Figure 2.3a**) with all markers revealed two distinct genetic groups, one which included the Namarai, Raviravi and Savusavu wild populations intermixing in a single cluster, and the other which comprised only the Savusavu hatchery population. The separation of the Savusavu hatchery population is particularly evident when the individual density distribution of the first retained

discriminant function (**Figure 2.3b**) is examined. Results of the simulation of parentage analysis revealed that no single parents, parent pairs or parent-offspring pairs were present in the dataset.

Network analysis at k-NN=40 (**Figure 2.4**) to visualise fine-scale population differentiation revealed a very similar pattern to the DAPC and pairwise F_{st} calculations. Individuals belonging to the Namarai, Raviravi and Savusavu wild populations demonstrated a high degree of admixture and formed a single cohesive group, whereas the Savusavu hatchery population remained in a separate cluster. The degree of separation seen between the two groups was not as distinct as depicted in the DAPC plot, as Savusavu hatchery individuals clustered together within the larger grouping formed by the other populations. Identical trends were observed at lower k-NN values ranging from 5-35 (results not presented here), with the overall pattern of separation remaining consistent.

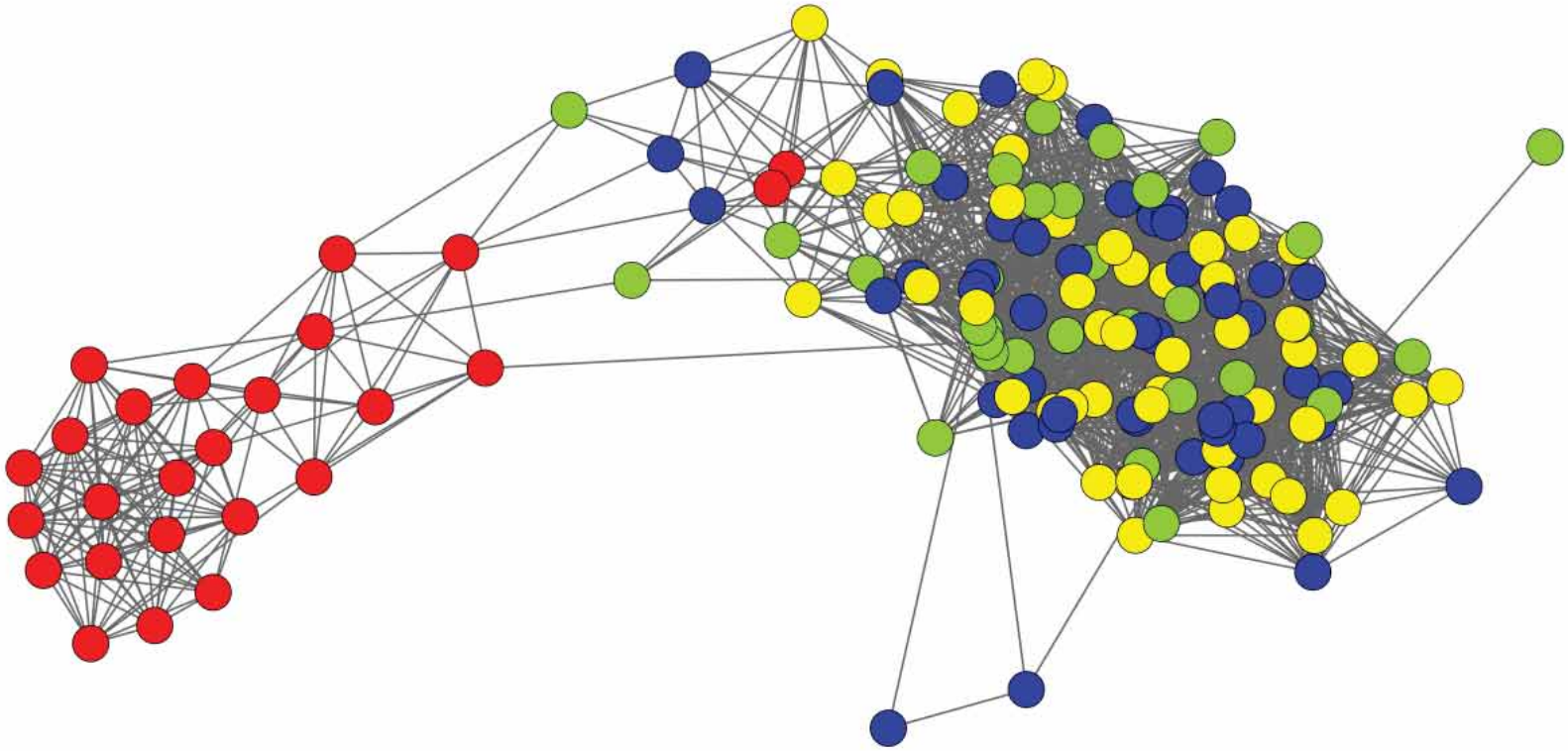


Figure 2.4 Population network of *P. margaritifera* individuals created using the NetView P v.0.4.2.5 pipeline based on an identity-by-similarity (IBS) distance matrix implemented in PLINK after Steinig *et al.* (2016). The network has been visualised at a maximum number of nearest neighbour (k-NN) threshold of 40, using 5,243 SNPs and 156 individuals. The Savusavu hatchery population is represented in red, the Savusavu wild population in blue, the Namarai wild in yellow and the Raviravi wild in green.

2.3.4 Discrimination of tissue colour morphotypes

Case-control association testing across the four populations revealed 64-152 highly-significant SNPs ($p<0.001$) under the allelic model, compared with those under the dominant (4-21 SNPs), recessive (6-18 SNPs) and genotypic (1-13 SNPs) models. Missing data was accounted for by specifying the minimum number of two populations to be considered for each locus test. When evaluating the percentage of common loci across all replicate populations, the highest proportion of significant SNPs was found in the recessive test (83.8 %), while the allelic test had the least (19.5%). Both dominant and genotypic tests had intermediate levels (69.8% and 55% respectively). These results suggest that the recessive model provides the greatest support for associated loci across the biological replicates. Visualisation of populations with unrooted NJ trees using all common loci revealed marked separation between individuals of different colour morphotypes (Namarai example provided in **Figure 2.5**). Despite the separation of morphotypes using these loci, no single SNP was diagnostic of colour morphotype, likely due to recombination effects, missing/incorrect data or the presence of several genes controlling the trait.

Detection of F_{st} outlier loci was carried out at a range of FDRs (0.01, 0.05, 0.10, 0.15 and 0.20) across all wild populations separated into colour morph groupings. For the purpose of differentiating individuals within populations on the basis of colour morphotype, only directional outliers are reported here. The results revealed 1-7 directional loci identified by Bayescan 2.1, to 42-121 detected by LOSITAN (**Appendix 2.2**). All directional outliers detected by Bayescan were also detected by LOSITAN. Following further evaluation using QQ-plots (Namarai example in **Appendix 2.4**), a final stringent FDR of 0.01 was selected, which returned 42-62 outlier loci within the populations. Visualisation of outlier loci with unrooted NJ trees revealed distinct separation of colour morphotypes within all populations tested (Raviravi population example shown in **Figure 2.6**), indicating that the directional loci selected at an FDR of 0.01 are likely to be true outliers, and connected to genomic regions associated with tissue pigmentation.

Comparisons between the significant loci ($p<0.001$) discovered in the association testing and those identified as significant F_{st} outliers by LOSITAN indicated 65% commonality between all populations for both methods. Although a large proportion of loci were co-identified, it is most likely that the disparities observed are due to differences in loci typed between

populations and the inclusion of missing data. A shortlist of all loci from the case-control association analyses was tested for LD, with 73 out of a total of 204 loci determined to be linked. All linked and unlinked loci were then BLASTn searched against the GenBank nucleotide (nt) database for terms related to pigmentation genes or pathways.

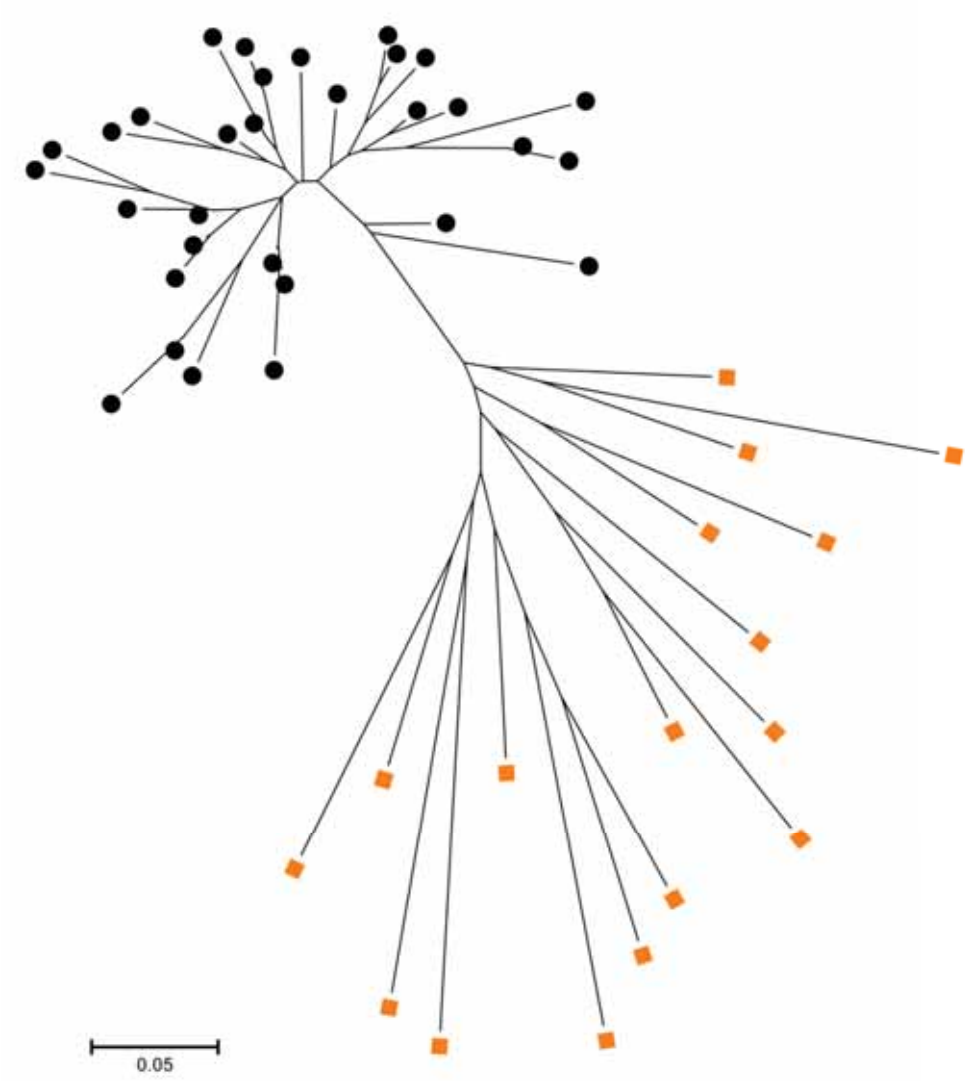


Figure 2.5 Unrooted neighbour-joining tree showing colour morphotype separation in *P. margaritifera* based on 83 significant ($p<0.001$) SNPs detected in case-control association testing in the Namarai population (n=50). Tree drawn in MEGA6 using 1-psa distance among 17 and 31 oysters expressing the orange and black morphotypes respectively. Scale bar indicates genetic distance.

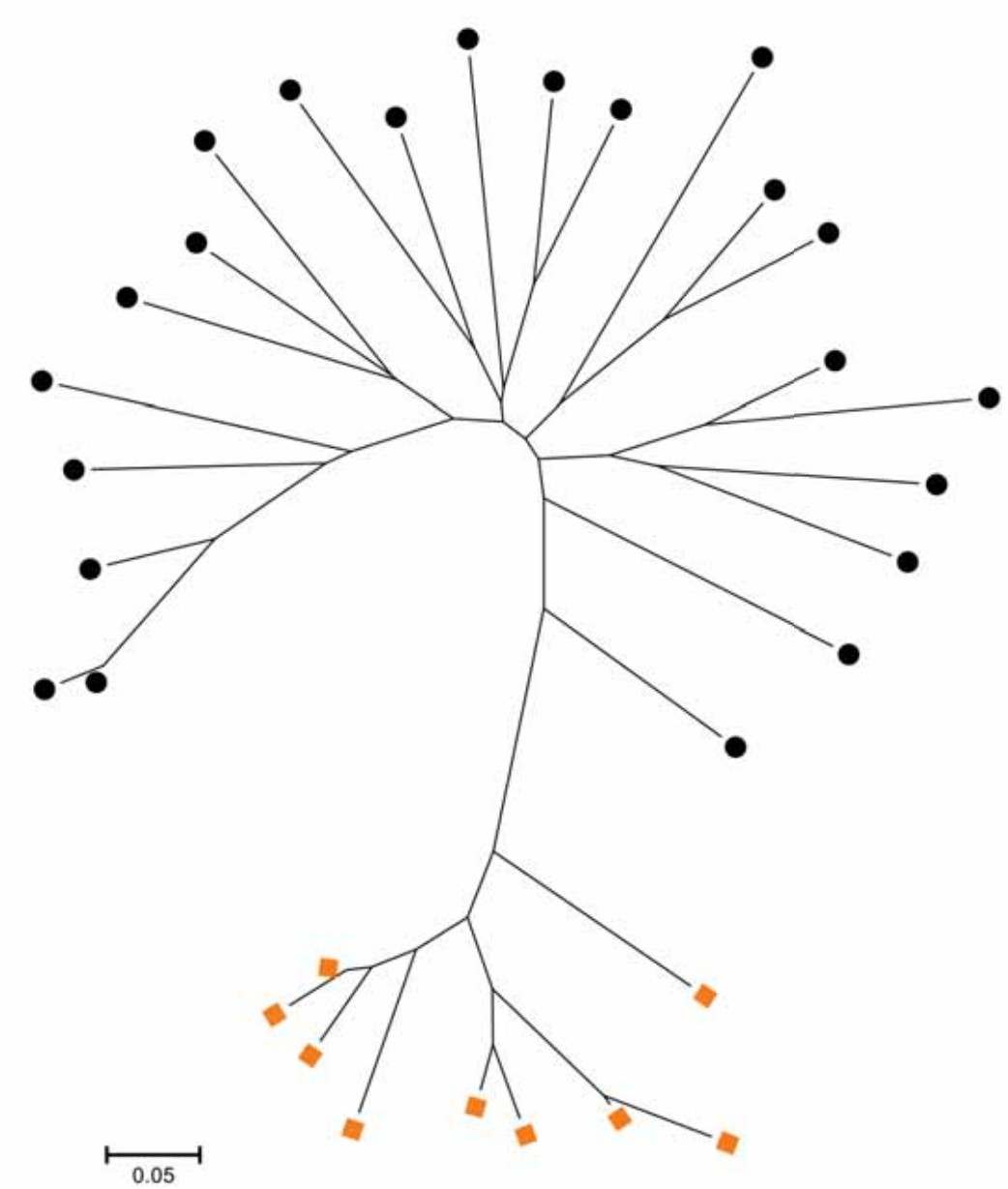


Figure 2.6 Unrooted neighbour-joining tree showing colour morphotype separation in *P. margaritifera* based on 100 Bayescan 2.1 and LOSITAN F_{st} outlier SNPs in the Raviravi population (n=32). Tree drawn in MEGA6 using 1-psa distance among 9 and 22 oysters expressing the orange and black morphotypes respectively. Scale bar indicates genetic distance.

Sequences containing 27 of the 204 loci searched returned information, with 22 of these situated in microsatellite loci typed for *Pinctada* spp., biomineralisation proteins, gonadal development, sex differentiation and cellular processes (see **Appendix 2.3**). Five loci interestingly returned hits for tyrosinase pathways, which is known to play an important role in the formation of the shell matrix and melanin biosynthesis in mollusc tissues and shells (Nagai *et al.* 2007; Takgi and Miyashita 2014). These sequence matches were for (NCBI

Accession numbers included in parentheses): *P. margaritifera* mRNA for tyrosinase 1 [HE610377.1], *P. margaritifera* mRNA for tyrosinase 2 [HE610378.1], *P. maxima* tyrosinase A2 (TyrA2) mRNA [KJ533305.1], *P. maxima* tyrosinase B5 (TyrB5) mRNA [KJ533314.1] and *P. fucata* mantle protein 11 (MG11) mRNA [DQ116438.1].

2.4 Discussion

2.4.1 Utility of ddRADseq

Development of 5,243 genome-wide SNPs for the black-lip pearl oyster using ddRADseq has permitted for the first time evaluation of genomic diversity and fine-scale population structure in this species for the Fiji Islands. The SNP marker set identified a genetic bottleneck in the hatchery produced population, and was also capable of detecting F_{st} outliers and performing an association analysis. Successful implementation of population genomic analyses as described here for *P. margaritifera* given the known complexities of bivalve genomes, holds much promise for application to other bivalve molluscs, as well as various other broadcast spawning non-model marine taxa. Comprehensive conservation and management strategies for these species rely on the recognition of fine-scale patterns of population structure as well as signatures of selection to reveal local adaptation, and the findings presented here validate the use of ddRADseq for providing these insights.

2.4.2 Population genomic diversity and differentiation

The overall levels of observed heterozygosity across all populations were lower than expected (**Table 2.1**). However, heterozygote deficits have been well documented in *P. margaritifera* and *P. maxima* (Arnaud-Haond *et al.* 2008; Durand & Blanc 1988; Durand *et al.* 1993; Lind *et al.* 2009), as well as other marine molluscs (Brownlow *et al.* 2008; Hedgecock *et al.* 2004; Miller *et al.* 2013; Peñaloza *et al.* 2014; Plutchak *et al.* 2006). A number of explanations have been proposed for this observation, including inbreeding effects, null alleles, aneuploidy, Wahlund effects (reduced heterozygosity caused by sub-population structuring) and indirect or direct natural selection acting on marker loci (Beaumont 1991; Peñaloza *et al.* 2014; Zouros & Foltz 1984).

While the inbreeding coefficients reported for *P. margaritifera* here were high (0.3049 - 0.3718 cf. Arnaud-Haond *et al.*, 2008), inbreeding alone is unlikely to account for the absence of heterozygotes, as other signatures such as reductions in N_e and correlated trends in multi-locus metrics (**Table 2.1**; IR, SH, HL) were absent. It is equally unlikely that the inflated F_{is} values result from a Wahlund effect or aneuploidy, given the panmictic population structure (NetView P network) and high levels of gene flow detected, and that to date no natural case of haploidy has been documented in the Mollusca (Plutchak *et al.* 2006).

The remaining possibilities to account for the observed heterozygote deficit are the inclusion of null alleles in the dataset or natural selection acting in favour of homozygotes during the pelagic larval stage (Addison & Hart 2005), after which post-settlement selection favours increased heterozygosity. The occurrence of null alleles is a persistent problem in genome-wide SNP datasets, and is problematic to resolve in order to maintain genotyping accuracy and avoid bias in estimations of population genetic parameters (Carlson *et al.* 2006; Dąbrowski *et al.* 2014). With increasing geographic distance between populations, higher levels of missing data were observed at loci genotyped between them. This disproportionate sampling of the genome has been related to mutations in restriction enzyme cut sites during ddRADseq genotyping, which can also lead to null alleles in populations, particularly for taxa with high rates of genome polymorphism (Andrews & Luikart 2014; Puritz *et al.* 2014). The presence of null alleles here for *P. margaritifera* cannot be ruled out despite the stringent filtering applied to the dataset, however its potential negative effect on downstream analysis was not detected.

Changes in ontogenetic selective pressure have been reported for the blue mussel *Mytilus edulis*, and slipper limpet (Beaumont 1991; Plutchak *et al.* 2006; Zouros & Foltz 1984). For *M. edulis*, differential mortality of homozygous and heterozygous genotypes occurred at most loci at the larval stage, which resulted in a detectable heterozygote deficit in juvenile animals (Beaumont 1991). However, the use of more than one method to estimate heterozygosity (e.g. MLH) and evaluate population structure (e.g. NetView P to rule out a Wahlund Effect), offer a potential solution for taxa in which heterozygote deficits are commonly observed.

Differences in N_e in the current study were also observed among the wild populations and between all wild populations and the hatchery population. The variation in N_e among wild populations may be explained by differential settlement and survival on spat collectors or reef

habitats these individuals were collected from, as has been reported for *P. maxima* (Lind *et al.* 2009; Lind *et al.* 2010) and *P. margaritifera* in French Polynesia (Arnaud-Haond *et al.* 2008). The hatchery produced population returned the lowest N_e (6.1) of all populations, providing clear indication of a genetic bottleneck resulting from a substantial loss of diversity due to differential family contributions and variable survival in a hatchery setting. Between 30 and 50 randomly selected broodstock oysters are typically used for a single spawning run at the hatchery where these animals were produced. However, only a small proportion of these are likely to contribute offspring towards a hatchery-reared cohort. This observation appears to be characteristic of a number of broadcast spawning species in aquaculture, and has previously been reported in *P. maxima*, flounder (*Paralichthys olivaceus*) and Atlantic halibut (*Hippoglossus hippoglossus*), among other taxa (Frost *et al.* 2007; Jackson *et al.* 2003; Lind *et al.* 2009; Lind *et al.* 2010; Sekino *et al.* 2003).

2.4.3 Broad and fine-scale population structure

The discovery of high levels of genetic admixture among the wild populations sampled using both broad-resolution (DAPC) and fine-resolution (NetView P) methods (**Figures 2.3a**, **2.3b**, and **2.4**) is unsurprising, given the relatively short distances (<150 Km) separating populations and the larval dispersal strategy of this species. The pairwise F_{st} comparisons (**Table 2.2**) support the existence of high gene flow between populations, effectively producing a singular genetic group consisting of the Namarai, Raviravi and Savusavu wild sample sets. Low levels of population structure have previously been described in *P. margaritifera* elsewhere, including Japan and French Polynesia at small to medium spatial scales <100 Km (Arnaud-Haond *et al.* 2003a; Arnaud-Haond *et al.* 2004; Durand & Blanc 1988; Durand *et al.* 1993), which stems from the wide larval dispersal capability and associated high levels of gene flow in this species.

Individuals belonging to the Savusavu hatchery sample set were genetically distinct from all other individuals sampled. However, this differentiation has resulted from a genetic bottleneck due to family effects typically observed as a result of hatchery practices as discussed earlier. On the basis of the neutral SNP markers described here, indications are that the populations sampled may be considered a homogenous genetic group. Nonetheless, it would be advantageous to sample oysters collected at additional sites at greater spatial scales, and also examine the adaptive variation potential at all sites, to make determinations on how

many Management Units (MUs; Funk *et al.* 2012) might exist for *P. margaritifera* in the Fiji Islands.

2.4.4 Discrimination of tissue colour morphotypes

Results of the association testing revealed a pattern of separation between morphotypes, as did the F_{st} outlier testing (**Figure 2.6**). Additional loci separating morphotypes were detected using the association test which were not discovered in F_{st} outlier testing, and different putatively associated loci were discovered between populations (32 % locus commonality). This is attributed to population effects, as not all loci were genotyped at 100 % commonality between populations. Despite these confounding effects, given the combination of 65 % commonality in loci detected between the outlier and associating testing across all populations, visualisation of segregation by morphotype in the NJ trees as well as the GO data, there appears to be biological association between the candidate loci and the orange vs. black tissue colour phenotype.

The BLAST information revealed five candidate loci linked to known tyrosinase expression pathways in the genus *Pinctada*, which interestingly reinforces the finding that the genomic regions discriminating between morphotypes are related to those which have biological functions linked to colour. Tyrosinase is an enzyme that catalyses the initial steps for melanin biosynthesis, which is responsible for the colouration of molluscan tissues and shells (Nagai *et al.* 2007; Takgi & Miyashita 2014). Work by Jabbour-Zahab *et al.* (1992) on the mantle histochemistry of *P. margaritifera* found a melanin-like pigment responsible for the black colouration of the shell and mantle tissue, which may account for the colouration of individuals expressing the black morphotype. Given the proportions of orange and black morphotypes observed in Fijian populations of *P. margaritifera*, the inheritance mode of either black or orange phenotypes may involve a simple, qualitative trait.

Phenotypic association investigations and adaptive outlier tests in *P. margaritifera* have not previously been reported, and information on the mechanisms behind mantle tissue colour is valuable for future MAS and identification of QTL in this species (Jones *et al.* 2014a; McGinty *et al.* 2012). The ability to separate colour morphotypes also illustrates the sensitivity of the genome-wide SNP marker set developed here to detect fine-scale intra-population variation, and demonstrates their potential for use in future studies examining

local adaptation in *P. margaritifera*. This will be of great value to future investigations seeking to delineate Management Units (MUs) in the Fiji Islands, given that a panmictic population structure is evident using neutral markers alone, and both neutral and adaptive variation need to be considered together for making informed management recommendations (Funk *et al.* 2012).

2.5 Conclusions

The development of genome-wide SNPs for *P. margaritifera* has provided a valuable tool for characterising populations in the Fiji Islands. The genotyping flexibility and efficiency delivered by ddRADseq for this species which previously lacked genomic resources, has high transferability to other broadcast spawning non-model marine taxa. The SNP markers have demonstrated high utility in detecting a mixed population structure across the four closely-related populations surveyed, and possess sufficient sensitivity to detect the signature of a genetic bottleneck in the hatchery produced population. The ability to distinguish between oysters belonging to different colour morphotypes also indicates that these markers are highly capable of detecting adaptive outlier loci and performing association testing, to provide further information on broad and fine scale stratification, diversity, relatedness, signatures of selection and candidate genes. These findings also highlight the potential value of ddRADseq in other marine taxa, for a range of investigations geared towards their conservation, exploitation and fishery management.

CHAPTER 3: Evaluation of genetic structure and connectivity in Fijian *P. margaritifera* for aquaculture and fishery management

3.1 Introduction

Sustainable management and conservation of marine species is faced with a number of challenges, among which is the wide distribution of taxa across diverse habitats and geopolitical jurisdictions, that make species-specific management plans difficult to design and implement. Many taxa also face high rates of exploitation, that in some cases has led to the collapse or abnormally slow recovery of wild fisheries, bringing into question whether current management strategies are effective or appropriate (Reiss *et al.* 2009; Waples 1998; Ward 2000). The need for informed fishery management has resulted in the development of the stock concept for aquatic species, which can allow for targeted conservation efforts and informed exploitation, once stock boundaries have been defined (Waples 1998; Waples *et al.* 2008). Despite the usefulness and importance of the stock concept, there is currently no clear consensus on what constitutes a stock, and numerous definitions in the literature shift emphasis for defining stock boundaries between the degree of demographic homogeneity within stocks, and their reproductive isolation (Carvalho & Hauser 1994). Since a stock is the fundamental unit used for fishery assessment and administration, it is imperative that the spatial boundaries delineated are also biologically meaningful, to ensure correct management action (Reiss *et al.* 2009; Waples & Gaggiotti 2006).

For assessment of a particular stock, it is important to determine the number and extent of populations being examined. However, the biological concept of a population has either ecological (demographic interactions of individuals), or evolutionary (genetic structuring) aspects (Carvalho & Hauser 1994; Reiss *et al.* 2009). Reiss *et al.* (2009) make the observation that many fishery management and assessment tools focus primarily on the ecological aspects of populations (e.g. population growth and mortality rates), while overall management goals also include many evolutionary criteria, such as the conservation of genetic diversity and maintenance of sustainable spawning stock biomass. This disconnect highlights the need for bridging the gap between fisheries management and population genetics, and particularly for characterising stock boundaries, identifying the level of

divergence required to manage two populations together, or as separate entities (Carvalho & Hauser 1994; Gaggiotti *et al.* 2009; Reiss *et al.* 2009; Waples & Gaggiotti 2006; Waples *et al.* 2008).

A major problem posed for application of the stock concept in the marine environment is the relative absence of barriers to dispersal and migration compared to terrestrial systems, and the highly-dispersive larval stages of many species (Waples 1998). For species which are either highly mobile and/or broadcast spawners with prolonged pelagic larval duration (PLD), the potential for gene dispersal is high, often resulting in weak population differentiation that is evident over large geographic distances (Hauser *et al.* 2011; Palumbi 2003; Waples & Gaggiotti 2006; Weersing & Toonen 2009). Furthermore, despite the presence of weak population structure, selective forces can produce fitness differences between populations through local adaptation (Nayfa & Zenger 2016).

For a large number of species that exhibit high levels of gene flow, low levels of genetic structure may be present, but difficult to detect (Reiss *et al.* 2009; Waples 1998). The importance of detecting low, but biologically significant differentiation for understanding the ecology and evolution of these taxa, and implications for their conservation and management is discussed by André *et al.* (2011), Gaggiotti *et al.* (2009), Hauser and Carvalho (2008), Palumbi (1994, 2003), Waples (1998) and Waples and Gaggiotti (2006). It is clear from these studies that a common solution for delimiting population and stock boundaries in high gene flow species is not possible, but rather assessment on an individual basis is required, taking into consideration the biological, ecological and fishery management issues involved. Additionally, in situations where traditional stock assessment is not possible (e.g. due to logistic or financial reasons), genetic approaches examining fine-scale population structure and functional differences (such as local adaptation), can be important for resolving stock boundaries.

A potential solution in recent years has been the use of genome-wide SNPs, which can reveal fine-scale patterns of population structure to highlight differences between populations, and also detect signatures of selection (Lal *et al.* 2016b; Limborg *et al.* 2012; Pujolar *et al.* 2014); with much higher resolving power than traditional markers (e.g microsatellites and mtDNA). However, while genetic analyses by themselves are a powerful tool for investigating population connectivity and structure, consideration of other data for defining stocks such as

phenotypic information, demographic data, or biophysical modelling should not be overlooked (Cannuel *et al.* 2009; Liggins *et al.* 2013; Reiss *et al.* 2009). For broadcast spawning species with prolonged PLD, investigations considering independent environmental and molecular data together, can provide unrivaled insights into the biological and physical processes that organise and regulate population structure (Berry *et al.* 2012; Waples *et al.* 2008). Hydrodynamic dispersal modelling is an analysis tool that relies on oceanographic data, and can be used for simulation and independent inference of larval dispersal from source to sink locations (Berry *et al.* 2012; Galindo *et al.* 2006). Because many marine species produce large quantities of very small larvae with variable PLD that makes tagging and tracking studies very difficult, highly realistic estimates of population connectivity can be achieved when hydrodynamic dispersal data are combined with genetic analyses (Berry *et al.* 2012; Dao *et al.* 2015; Siegel *et al.* 2003; Siegel *et al.* 2008; Waples *et al.* 2008).

Bivalve molluscs present a number of unique challenges for stock assessment, which include highly variable patterns of larval dispersal and recruitment. Additionally, traditional bivalve stock assessment surveys typically require extensive sampling to determine distribution and abundance, which in most situations can be costly and impractical. Because the adults of many taxa are sedentary and recruitment rates highly variable, a stock may occupy a discrete geographic region as large as an entire reef system, or as small as a single bivalve bed (Gosling 2015). When coupled with the homogenising effects of larval exchange over large distances, accurate stock assessment can quickly become problematic. For many bivalves, and pearl oysters in particular, examination of morphological differences for stock assessment primarily relies on variable shell characters to elucidate differences between individuals, populations and species (Wada & Tëmkin 2008). This can be a difficult exercise, particularly during the early stages of development (Hare *et al.* 2000), as factors including phenotypic plasticity and environmental effects can confound measurements. In recent times, molecular methods have been increasingly relied upon to provide solutions to these problems (Wada & Jerry 2008; Wada & Tëmkin 2008).

In French Polynesia, the black-lip pearl oyster *P. margaritifera* (Pteriidae) displays substantial genetic fragmentation, despite being a broadcast-spawner with an extended PLD of 26-30 days (Alagarswami *et al.* 1989; Doroudi & Southgate 2003). This has been related primarily to habitat heterogeneity, with significant genetic structure detected between open and closed atoll lagoon systems (Arnaud-Haond *et al.* 2008; Lemer & Planes 2014). Here,

detection of both fine-scale and broad-scale patterns of differentiation were identified as being biologically important for fishery and aquaculture management (Arnaud-Haond *et al.* 2003b; Lemer & Planes 2012). For the Fiji Islands, cultured pearls and pearl shell production from *P. margaritifera* is a valuable industry and substantial source of coastal community livelihoods. It produces a high-value, low-volume and non-perishable product with a comparatively smaller environmental footprint than most other forms of aquaculture, making it an ideal export commodity for developing Pacific island countries (Southgate *et al.* 2008; SPC 2003, 2007). The industry is almost exclusively dependent on wild oysters for which there are currently no comprehensive fishery management guidelines, and therefore no information is available on the number of discrete populations present, their levels of genetic fitness and relatedness, or if domestic translocation of animals is suitable for the establishment of new pearl farms.

Two preliminary stock assessment surveys using traditional methods reported low abundances of *P. margaritifera* at all locations examined, and recommended immediate conservation efforts to increase population densities of wild oysters (Friedman *et al.* 2010; Passfield 1995). The study described previously in chapter 2 which examined oysters sampled at four Fijian sites discovered a mixed pattern of population structure, and identified a need for comprehensive evaluation of additional populations to determine country-wide patterns of genetic structure and connectivity (Lal *et al.* 2016b). In this the stock structure of *P. margaritifera* in the Fiji Islands is assessed for fishery and aquaculture management, using independent population genomic and hydrodynamic modelling approaches. This work provides valuable insights for the fishery management and aquaculture of this commercially important bivalve mollusc, and also demonstrates solutions for challenges that apply to stock assessment efforts in other broadcast-spawning marine taxa, that possess similar life history characteristics.

3.2 Methods and materials

3.2.1 Specimen collection, tissue sampling and DNA extraction

Adult and juvenile *P. margaritifera* (n=427) sized between 7-18 cm in DVM were collected from 11 sites in the Fiji Islands, representing both farmed and wild populations country-wide from December 2012 to October 2013 (**Figure 3.1**). Permission to sample wild sites was obtained from Fijian traditional fishing ground (*i qoliqoli*) custodians, while farm site access was permitted by farm owners. The vast majority of farmed oysters are collected as settling wild juveniles or spat, that recruit to dedicated settlement substrates (spat collectors) deployed by farms. Additionally, limited numbers of individuals are propagated in a single hatchery, and are the progeny of wild-sourced broodstock. Broodstock oysters for hatchery production are preferentially sourced from wild populations, and killed following each spawning run, with new oysters collected for subsequent hatchery production cycles. Oysters are grown in pocket panel nets that are suspended in the water column from long lines (Southgate 2008). At all farm sites, wild populations are present in adjacent habitats. Farmed oysters were sampled at Ra (n=50), Raviravi (n=32), Taveuni (n=43) and three locations in Savusavu: Vatubukulaca (n=50); Wailevu (n=49) and a hatchery-produced population also at Wailevu (n=50). Oysters collected from all farms originated either from spat collectors (Southgate 2008), or were gleaned from adjacent coral reef habitats. Wild populations were sampled at two sites on the Island of Kadavu (Galoa Island; n=25 and Ravitaki; n=25), the Yasawa archipelago (Naviti Island; n=35), Udu Point (n=18) and the Lau archipelago (Nayau Island, n=50). Two sites were sampled on Kadavu to detect any differentiation present between adjacent locations due to environmental heterogeneity (e.g. reef effects). Proximal mantle and adductor muscle tissues (1.5 and 1 cm respectively), were collected and handled for gDNA extraction as described earlier in chapter 2. All oysters were handled in accordance with James Cook University's animal ethics requirements and guidelines.

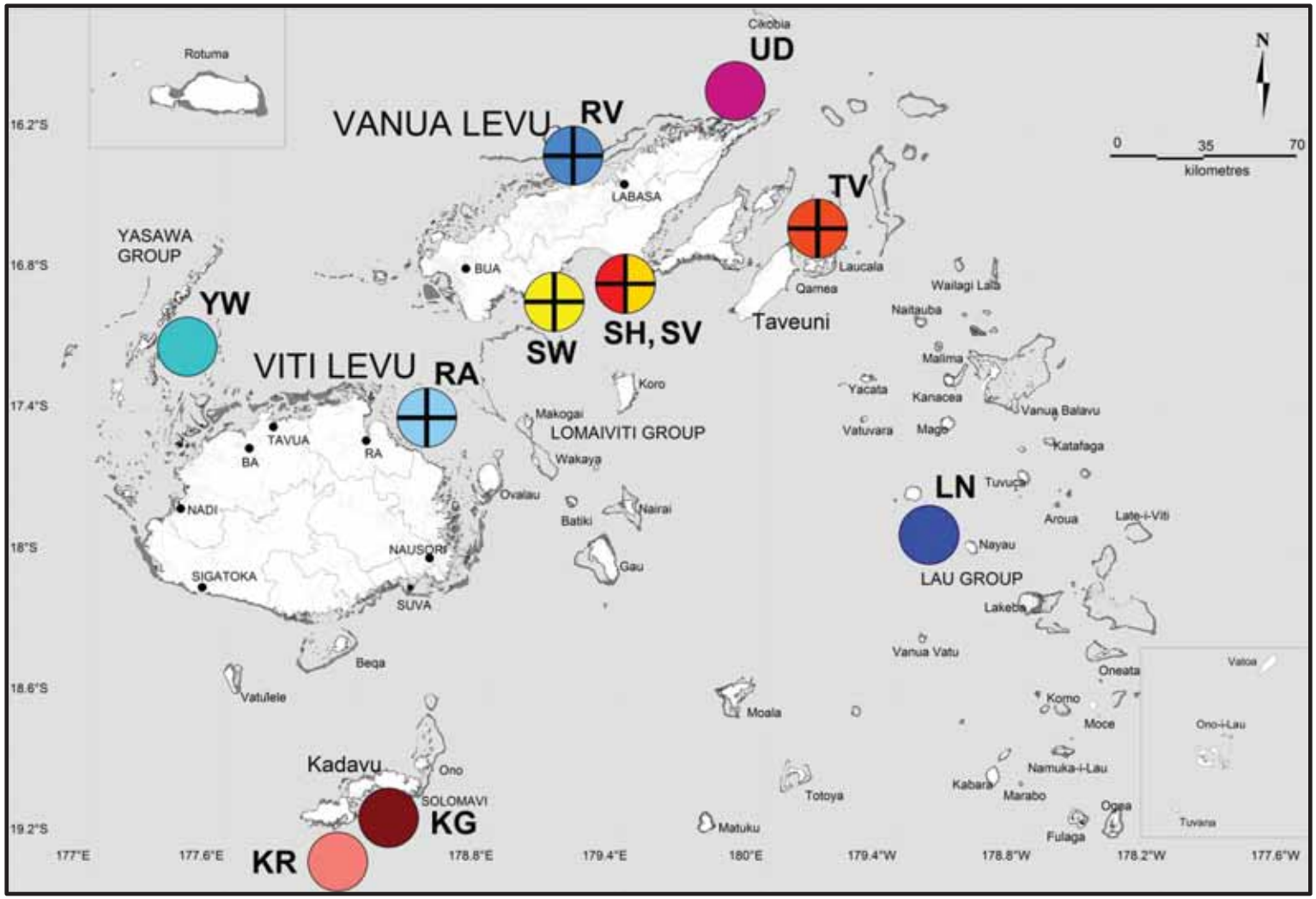


Figure 3.1 Map of sampling locations in the Fiji Islands indicating where wild and farmed *P. margaritifera* were collected. Reef outlines are presented in dark grey. Solid circles represent wild oyster collection sites, while circles superimposed with a cross indicate farm locations. Site codes represent the following locations: YW, Naviti Island in the Yasawa group; RA, farm site at Namarai, Ra; SW, farm site at Wailevu, Savusavu; SH, farm site at Wailevu, Savusavu for hatchery produced oysters; SV, farm site at Vatubukulaca, Savusavu; RV, farm site at Raviravi; UD, Vunikodi, Udu Point; TV, farm site at Wailoa, Taveuni; LN, Nayau Island in the Lau group; KG, Galoa Island off Kadavu Island and KR, Ravitaki on Kadavu Island.

3.2.2 ddRADseq library preparation and sequencing

Double digest restriction site-associated (ddRAD) libraries were prepared following the methods of Peterson *et al.* (2012), with a number of modifications for *P. margaritifera* as described by Lal *et al.* (2016b) and in chapter 2. Briefly, nine libraries in total were prepared (48 barcoded individuals per pool × nine unique Illumina TruSeq indices), from which four libraries were pooled at equimolar ratios for sequencing in one lane, while the remaining five libraries were pooled for a second lane. After cluster generation and amplification (HiSeq PE Cluster Kit V4 cBot^{TM}), 100 bp paired-end sequencing was performed on an Illumina HiSeq 2000 platform at the Australian Genome Research Facility (AGRF) in Melbourne, Victoria.

3.2.3 Sequence quality control, marker filtering and genotype calling

Raw reads obtained following sequencing were processed as described by Lal *et al.* (2016b), with all read filtering and SNP genotyping carried out using STACKs v.1.20 software (Catchen *et al.* 2013; Catchen *et al.* 2011). From all available SNPs, only the most informative SNP per locus was selected for further analysis. Final genotypes were called at a Minor Allele Frequency (MAF) of 2% and minimum stack depth of 10, with the minimum proportions of loci allowed across individuals set at 20%, and across populations at 50% (-r and -p options respectively). In addition, each unique SNP was genotyped in at least 10 individuals within a population, and represented in a minimum of two populations across the whole dataset (Huang & Knowles 2014).

All loci were screened using Arlequin v.3.5.1.3 (Excoffier *et al.* 2005) for departure from Hardy-Weinberg Equilibrium (HWE), and removed if deviations were significant after FDR correction ($p<0.00001$), or loci were monomorphic across all populations (Zenger *et al.* 2007a; Zenger *et al.* 2007b). All loci were also tested for genotypic linkage disequilibrium (LD) in Genepop v.4.3 (Rousset 2008), as per Lal *et al.* (2016b).

3.2.4 Evaluation of genetic diversity, inbreeding and population differentiation

For assessment of genetic diversity within and between populations, allelic diversity indices were computed as described in chapter 2. These metrics included average observed (H_o), and average expected heterozygosities corrected for population sample size ($H_{n.b.}$), inbreeding

coefficients (F_{is}) and effective population size based on the linkage disequilibrium method (N_{eLD}). Furthermore, family relationships among all individuals within sampled populations were assessed with ML-RELATE (Kalinowski *et al.* 2006), which allowed for the identification of any parent-offspring, full-sib or half-sib pairs present. Relationships between individuals from different regions were also evaluated by assessing all populations together, in order to detect migration levels.

High levels of genome-wide polymorphism characterise many bivalves and other marine invertebrates, which may affect RADseq-based genotyping approaches by disproportionately sampling the genome due to mutations in restriction enzyme cut sites (Andrews & Luikart 2014; Puritz *et al.* 2014). As previously outlined in chapter 2, to ascertain the potential degree of bias, F_{is} and heterozygosity were calculated for the dataset during preliminary testing at a range of missing data thresholds from 80 to 20%. These parameters were also calculated at varying read depths per stack from 5 to 15 (in the STACKs 'populations' module), before performing final F_{is} and heterozygosity computations. Heterozygosity and F_{is} changed with increasing read depth per stack from 3 to 6, however, no substantial change occurred beyond a read depth of 7. Based on these results, a final read depth threshold of 10 was selected for generating final genotypes.

To investigate individual genomic levels of diversity, multi-locus heterozygosity was examined, with the standardised heterozygosity (SH) and internal relatedness (IR) computed for each population with the *R* package *Rhh* (Alho *et al.* 2010). Furthermore, the average multi-locus heterozygosity (Av. MLH) per population was computed manually following Slate *et al.* (2004), along with the proportion of rare alleles with a MAF <5%. To investigate levels of population structure between sampling locations, pairwise F_{st} estimates for each population were calculated using Arlequin v.3.5.1 (Excoffier *et al.* 2005) with 10,000 permutations, and broad-scale population structure visualised by performing a Discriminant Analysis of Principal Components (DAPC) in the *R* package *adegenet* 1.4.2 (Jombart 2008; Jombart & Ahmed 2011; Jombart *et al.* 2010). The DAPC was carried out for all loci, and an α-score optimisation used to determine the number of principal components to retain. Additionally, the 'find.clusters' function of *adegenet* was utilised to determine the optimal number of actual clusters using the Bayesian Information Criterion (BIC) method.

3.2.5 Resolution of fine-scale population structure

To reveal any fine-scale stratification between and among all populations, network analysis was carried out using the NetView P pipeline v.0.4.2.5 (Neuditschko *et al.* 2012; Steinig *et al.* 2016), following the methods described in chapter 2. The IBS matrix and corresponding networks were constructed at various thresholds of the maximum number of nearest neighbour (k-NN) values between 5 and 40. Additionally, a hierarchical Analysis of Molecular Variance (AMOVA) was carried out in GenAlEx v.6.5 (Peakall & Smouse 2006), to examine variation between farmed and wild groups of populations.

3.2.6 Examination of adaptive variation

To detect signatures of selection, all pairwise population combinations were considered for F_{st} outlier detection. Testing failed to detect any outlier loci, with the exception of three population pairs (see results section **3.3.4**). Two independent outlier detection methods were used to identify candidate loci under selection, comprising the BayeScan v.2.1 (Foll 2012; Foll & Gaggiotti 2008) and LOSITAN selection detection workbench (Antao *et al.* 2008) packages. All settings used were as described in chapter 2, with the following modifications; BayeScan 2.1 analyses were performed on a 1:10 prior odds probability for the neutral model and commenced with 20 pilot runs consisting of 5,000 iterations each. This was followed by 100,000 iterations with a burn-in length of 50,000 iterations (Foll 2012). A range of FDR values from 0.01 to 0.10 were evaluated based on preliminary testing, and recommendations by Ball (2013) and Hayes (Hayes 2013). All LOSITAN outlier detection was computed within a 95% confidence interval under an infinite allele model, with 50,000 iterations also evaluating a range of FDR values from 0.01 to 0.10 to match the BayeScan 2.1 analyses. All other test parameters remained at their default settings, with the exception of the 'Neutral' mean F_{st} and 'Force mean F_{st}' options being enabled.

3.2.7 Particle dispersal simulation

To independently compare results of the population genomic analyses with environmental data and to simulate larval transport pathways between sampling locations, a particle dispersal model was developed, which is publicly available at https://github.com/CyprienBos

serelle/DisperGPU. Larvae of *P. margaritifera* typically remain in the plankton for 26-30 days prior to settlement (Alagarswami *et al.* 1989; Doroudi & Southgate 2003), and due to very limited motility, are largely dispersed by current advection and turbulent diffusion in the ocean surface (mixed) layer.

3.2.7.1 Hydrodynamic and dispersal numerical models

The particle (numerical) dispersal model was driven by current velocity output from the global HYbrid Coordinate Ocean Model (HYCOM) data (Chassignet *et al.* 2007; Cummings 2005). HYCOM is a global hydrodynamic model that simulates ocean surface heights, currents, salinity and temperature, both at the surface and at depth. The model is driven by meteorological forcing, and constantly constrained by the assimilation of global, remote and in-situ ocean observations. As the model simulates regional and global circulation, it does not include tidal or surface wind waves. HYCOM is highly useful for forecasting and simulation experiments, with public availability at https://hycom.org. The HYCOM model had a resolution of 1/12th of a degree and output every day. Although it simulates current movement in all three dimensions, only the surface layer was used to drive the dispersal model, as this is where larvae remain in the water column (Beer & Southgate 2000). The particle model used a standard Lagrangian formulation (Siegel *et al.* 2003; Siegel *et al.* 2008), where particles have no physical representation, but rather track the displacement of neutrally buoyant small objects such as larvae (relative to the model resolution), at the ocean surface. Particle displacement is expressed as:

$$\Delta x = u_p * \Delta t + K \qquad Eq.(1)$$

Here x represents particle position (latitude and longitude), Δx is particle displacement during a time step Δt (which was set at 1 hour), and u_p is the surface current speed at the location of the particle. K is the eddy diffusivity which takes account of the random displacement of the particle, due to turbulent eddies at a scale smaller than the hydrodynamics model resolution. K is calculated after Viikmäe *et al.* (2013) as follows:

$$K = \sqrt{-4E_h \Delta t \log(1 - R_{NA})} \cos(2\pi R_{NB}) \qquad Eq.(2)$$

Here E_h is a horizontal turbulent diffusion coefficient, and R_{NA} with R_{NB} are normally distributed random numbers. The horizontal turbulent diffusion coefficient is unknown, but

assumed to be 1 m^2s^{-1}. u_p is calculated by interpolating the velocity from the hydrodynamics model, both spatially and temporally. Gridded surface currents are first interpolated to the dispersal step, after which the current velocity at each particle position is calculated using a bi-linear interpolation of the gridded surface currents, where only surface currents are taken into account and vertical movements neglected (Markey *et al.* 2016). The particle age is retained and increases with simulation progression.

3.2.7.2 Model configuration

Particles were seeded in eight locations broadly corresponding to locations from where oysters were sampled for genetic analyses (see **Fig 3.4**). Seeding locations were represented at scales larger than the sampling locations to factor in the extent of surrounding coral reef habitat and farm boundaries. All seed areas were also extended farther offshore to account for the fact that the HYCOM model is not adapted for shallow water environments, and does not resolve fine-scale hydrodynamic patterns <10 km (Halliwell 2004). At each seed location, 25,600 particles were released once at the start of the simulation, which optimised the computational requirements for running the dispersal model.

The simulation was carried out using HYCOM data for February-April 2009 and 2010, based on observations of the peak spawning period for *P. margaritifera* in Fiji (Saucedo & Southgate 2008; Vilisoni 2012), and to test for circulation pattern differences over El Niño Southern Oscillation (ENSO) event years (2009 recorded an El Niño). Selection of this timeframe was also based upon inference of when sampled oysters were likely to be completing larval development and undergoing settlement, using shell size to approximate age (Pouvreau & Prasil 2001; Pouvreau *et al.* 2000). In this way, results of both the genetic and hydrodynamic analyses were restricted to the oysters sampled.

Particle positions were extracted at time intervals of 1, 15, 30 and 60 days post-seeding and no mortality or competency behaviour of the particles was simulated. Explicit, quantitative correlation of the genetic and hydrodynamic analyses was not possible, as this would have required genetic analysis of oysters at all potential source and sink locations with dense sampling coverage, and modelling of substantially more complex particle behaviour than computational resources permitted. Instead, an independent approach was adopted here, to

examine congruency of results produced by the two analyses. Although the model is unsuitable for evaluation of recruitment rates, it does allow insights into possible connectivity between sampling locations.

3.3 Results

3.3.1 Genotyping and SNP discovery

Following sequencing, a total of 765,273,656 PE raw reads were obtained for all nine libraries across both lanes. Read filtering using the STACKs pipeline ('process_radtags' and 'ustacks' modules) to discard low quality reads (Phred33 score <30; 5.25% discarded), ambiguous barcodes and overrepresented sequences, resulted in 725,064,036 high quality reads remaining. These reads were used to generate a locus catalogue in the 'cstacks' module containing 303,650 stacks (**Appendix 3.1**). This catalogue was used to generate all genotypes, using a median number of 555,524 reads to assemble 33,738 stacks for each individual (average read depth per stack of 17.81). Subsequent filtering at a minimum read depth of 10 per stack and MAF>0.02 resulted in a total of 42,341 genome-wide SNPs being genotyped. The primary dataset of 42,341 SNPs was screened to retain only the single most informative SNP per locus, remove those loci significantly deviating from HWE ($p<0.00001$) and under LD ($p<0.0001$) across all populations, retain individuals/populations with maximum genotyping rates, and also remove loci generated from contaminant sequences. These steps generated a final dataset of 4,123 high quality, polymorphic, genome-wide SNPs for further population genomic analyses.

3.3.2 Population genomic diversity and differentiation

Observed heterozygosities were significantly lower ($p<0.05$) than expected heterozygosities for all populations (H_o: 0.0621 - 0.1461; $H_{n.b.}$: 0.2903 - 0.3449, see **Table 3.1)**, and displayed similar trends to the proportions of rare alleles present in each population. The individual average multi-locus heterozygosity (MLH) computations matched the trends in observed heterozygosity, with the Kadavu (Ravitaki, wild) and Udu Point (wild) populations having the lowest (0.0687) and highest (0.1522) values, respectively. Lower MLH values were observed for island archipelago populations, when compared with oysters sampled from

locations neighbouring larger land masses; e.g. Yasawa and the two Kadavu sites (0.0703, 0.0695 and 0.0687 respectively), vs. Ra, Raviravi and Udu Point (0.1407, 0.1465 and 0.1522, respectively). Similar patterns were apparent in the standardised heterozygosity (SH) metrics (**Table 3.1**), with island archipelago population SH values ranging from 0.5361-0.8899 (Kadavu; Galoa to Lau), and mainland populations producing values between 0.8249-1.1609 (Savusavu; Vatubukulaca to Udu Point).

Inbreeding coefficient (F_{is}) values were variable across populations (**Table 3.1**), ranging from 0.4370 for the Savusavu hatchery population, to 0.6876 for the Kadavu (Ravitaki) wild population. Interestingly, the hatchery produced Savusavu oysters demonstrated the lowest F_{is} values, whereas several wild populations, such as Yasawa (0.6423) and Taveuni (0.5513), produced higher values. Generally, slightly higher F_{is} values were observed among populations sourced from island archipelagos, e.g. Taveuni, Yasawa and the two Kadavu sites (0.5513, 0.6423, 0.6407 and 0.6876, respectively). This contrasted with estimates for oysters collected from fringing reef systems connected with the major islands of Viti Levu and Vanua Levu; e.g. Raviravi, Ra, Udu Point and Wailevu at Savusavu (0.4552, 0.4639, 0.4740 and 0.4903, respectively). Internal relatedness (IR) was comparable to the F_{is} values calculated for each respective population. The highest IR values were observed for all island populations, ranging from 0.6189 (Lau) to 0.7907 (Kadavu, Ravitaki). Among the farmed populations, the Raviravi (0.4943), Ra (0.5105), Savusavu (Wailevu; 0.5567) and Savusavu (Wailevu hatchery; 0.5713) oysters exhibited intermediate IR values, while the highest IR was recorded for oysters sampled at Savusavu (Vatubukulaca; 0.6760).

Estimates of effective population sizes were infinite for all populations (**Table 3.1**), with the exception of the Ra (658.4; [95% CI: 534 - 854.9]), Savusavu (Wailevu; 152.4 [95% CI: 142 - 164.3]) and Savusavu hatchery oysters (5.2 [95% CI: 5.1-5.3]). Pearl oysters obtained from these locations were all farmed animals, and sourced from spat collector deployments adjacent to the farm sites. The only farm sites sampled which produced infinite N_{eLD} values were Taveuni and Ra, however, most of these animals had been directly collected from reef systems adjacent to the farms themselves. The Savusavu hatchery population was found to be bottlenecked with the lowest N_{eLD} of 5.2, most likely as a result of variable family survival and broodstock contributions, as discussed in chapter 2.

Table 3.1 Genetic diversity indices for the wild and farmed *P. margaritifera* populations examined. The parameters calculated included proportion of rare alleles (<5%), effective population size by the linkage disequilibrium method (N_{eLD}; 95% confidence intervals indicated within brackets), observed heterozygosity (H_o), average expected heterozygosity corrected for population sample size ($H_{n.b.}$), inbreeding coefficient values (F_{is}), average individual multi-locus heterozygosity (MLH), standardised heterozygosity (SH) and internal relatedness (IR).

Population	Origin	n	Proportion of rare alleles (MAF <5%)	N_{eLD} [95% C.I.]	H_o (± SD)	$H_{n.b.}$ (± SD)	F_{is} (p<0.01)	MLH (± SD)	SH (± SD)	IR (± SD)
Ra (Namarai)	Farm (major island; Viti Levu)	50	11.3%	658.4 [534.8 - 854.9]	0.1338 (±0.1261)	0.2903 (±0.1443)	0.4639	0.1407 (± 0.0189)	1.1226 (± 0.1623)	0.5105 (± 0.0667)
Taveuni (Wailoa)	Farm (offshore island)	43	10.9%	∞ [∞ - ∞]	0.1054 (±0.1155)	0.2943 (±0.1507)	0.5513	0.1052 (± 0.0699)	0.7383 (± 0.3749)	0.6733 (± 0.1780)
Raviravi	Farm (major island; Vanua Levu)	32	10.4%	∞ [2422.5 - ∞]	0.1353 (±0.1325)	0.2950 (±0.1488)	0.4552	0.1465 (± 0.0221)	1.1414 (± 0.1290)	0.4943 (± 0.0813)
Savusavu (Vatubukulaca)	Farm (major island; Vanua Levu)	50	6.5%	∞ [∞ - ∞]	0.0922 (±0.1387)	0.3151 (±0.1414)	0.5239	0.1007 (± 0.0469)	0.8249 (± 0.4129)	0.6760 (± 0.1511)
Savusavu (Wailevu)	Farm (major island; Vanua Levu)	49	8.6%	152.4 [142.0 - 164.3]	0.1258 (±0.1552)	0.3062 (±0.1430)	0.4903	0.1366 (± 0.0149)	1.1138 (± 0.1183)	0.5567 (± 0.0537)
Savusavu (Wailevu, hatchery)	Farm (major island; Vanua Levu)	50	11.4%	5.2 [5.1 - 5.3]	0.1380 (±0.1860)	0.3063 (±0.1540)	0.4370	0.1456 (± 0.0228)	1.1690 (± 0.1727)	0.5713 (± 0.0702)
Lau (Nayau Island)	Wild (archipelago)	50	9.8%	∞ [∞ - ∞]	0.1093 (±0.1176)	0.2975 (±0.1476)	0.5058	0.1111 (± 0.0356)	0.8899 (± 0.2815)	0.6189 (± 0.1246)
Yasawa (Naviti Island)	Wild (archipelago)	35	7.0%	∞ [∞ - ∞]	0.0653 (±0.0956)	0.3113 (±0.1453)	0.6423	0.0703 (± 0.0343)	0.5514 (± 0.2783)	0.7613 (± 0.1229)
Udu Point (Vunikodi)	Wild (major island; Vanua Levu)	18	7.4%	∞ [∞ - ∞]	0.1461 (±0.1535)	0.3169 (±0.1468)	0.4740	0.1522 (± 0.0096)	1.1609 (± 0.0708)	0.4972 (± 0.0337)
Kadavu (Galoa Island)	Wild (archipelago)	25	3.8%	∞ [∞ - ∞]	0.0673 (±0.1322)	0.3449 (±0.1380)	0.6407	0.0695 (± 0.0311)	0.5361 (± 0.2510)	0.7897 (± 0.0950)
Kadavu (Ravitaki)	Wild (archipelago)	25	3.8%	∞ [∞ - ∞]	0.0621 (±0.1131)	0.3444 (±0.1398)	0.6876	0.0687 (± 0.0191)	0.5498 (± 0.1564)	0.7907 (± 0.0584)

Relatedness calculations between individuals revealed no parent-offspring pairs present in the dataset (**Appendix 3.2**). However, full-sib and half-sib relationships were detected for the Savusavu (Vatubukulaca) farm population (with 8 full-sib and 86 half-sib pairs), and 83 full-sib and 116 half-sib pairs identified for the Savusavu hatchery-produced oysters. When between-region relationships were assessed by examining all populations together (**Appendix 3.3**), the degree of relatedness declined with increasing geographic distance. The largest number of full-sib relationships was detected between Savusavu and Lau (25), with lower numbers between Savusavu and Kadavu, Taveuni and the Yasawa archipelago respectively, (4 relationships each). Higher numbers of half-sib relationships between these regions were discovered, particularly between Savusavu and Lau, Taveuni, Kadavu, the Yasawa archipelago and Raviravi (73, 37, 24, 17 and 14 respectively). Between the most distant populations sampled, only 1-2 full-sib and 1-9 half-sib relationships were detected between the Yasawa and Lau, Taveuni and Kadavu populations, respectively. However, 19 half-sib relationships were evident between both Kadavu-Lau and Kadavu-Taveuni.

3.3.3 Resolution of population structure

Pairwise F_{st} estimates (**Table 3.2**) did not significantly depart from zero across almost all populations (average overall F_{st}=0.0028; $p>0.05$), except for the hatchery produced oysters (Savusavu, Wailevu), which showed weak, but significant separation ($p<0.000001$) from four other populations: Ra (farm), Raviravi (farm), Udu Point (wild) and Savusavu, Wailevu (farm). Evaluation of population structure with a DAPC following α-score optimisation to retain 16 informative principal components (**Appendix 3.4**), revealed differentiation across two separate clusters (**Figure 3.2**). The Savusavu hatchery oysters separated from all other populations, with all remaining populations forming a single, diffuse cluster with overlapping 95% inertia ellipses. This separation was confirmed by testing for the actual number of discrete clusters, which was determined to be k=2 (Bayesian Information Criterion (BIC) method**; Appendix 3.5**).

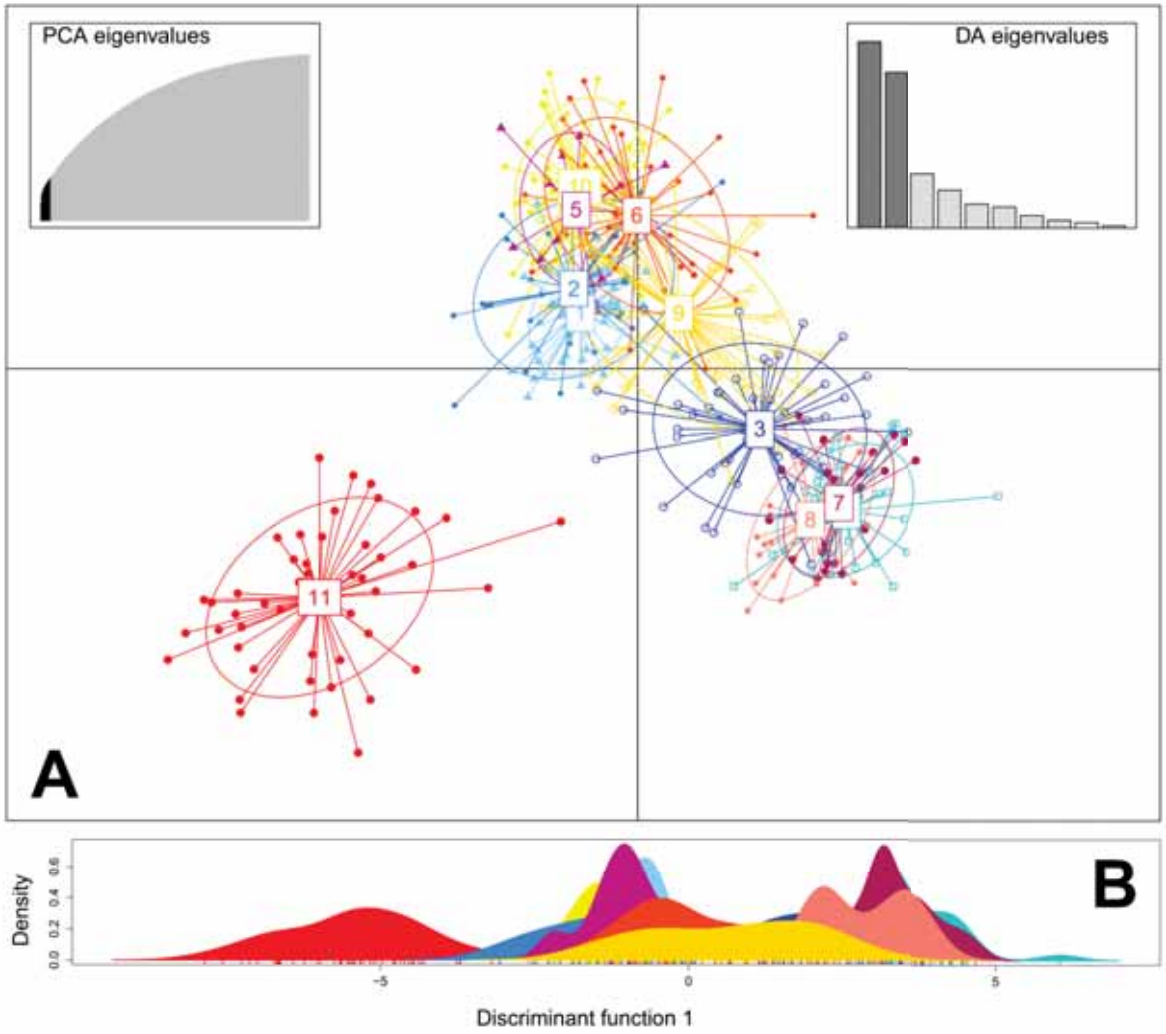


Figure 3.2 Discriminant Analysis of Principal Components (DAPC) scatter plot (A) and individual density plot on the first discriminant function (B), drawn across 427 *P. margaritifera* individuals in the *R* package *adegenet*. Dots represent individuals, with colours denoting sampling origin and inclusion of 95% inertia ellipses. Site colours correspond with **Figure 3.1**, and site numbers are as follows: (**1**) farm site at Namarai, Ra; (**2**) farm site at Raviravi; (**3**) Lau group; (**4**) Yasawa group; (**5**) Udu Point; (**6**) Taveuni; (**7**) Kadavu (Galoa Island); (**8**) Kadavu (Ravitaki); (**9**) farm site at Savusavu (Vatubukulaca); (**10**) farm site at Savusavu (Wailevu) and (**11**) farm site at Savusavu (Wailevu, hatchery produced oysters).

Table 3.2 Population pairwise F_{st} estimates. Estimates were computed using Arlequin (Excoffier *et al.* 2005; Weir and Cockerham 1984 unbiased method), for 4,123 SNP loci in *P. margaritifera* from 11 Fijian populations. Significantly different values at p<0.000001 following 10,000 permutations are indicated with an asterisk.

	Ra (Farm)	Raviravi (Farm)	Lau (Wild)	Yasawa (Wild)	Udu Point (Wild)	Taveuni (Farm)	Kadavu, Galoa (Wild)	Kadavu, Ravitaki (Wild)	Savusavu, Vatubukulaca (Farm)	Savusavu, Wailevu (Farm)
Raviravi (Farm)	0.0009									
Lau (Wild)	0.0050	0.0022								
Yasawa (Wild)	0.0126	0.0057	0.0026							
Udu Point (Wild)	-0.0034	-0.0045	-0.0021	-0.0015						
Taveuni (Farm)	0.0011	-0.0010	0.0044	0.0109	-0.0059					
Kadavu, Galoa (Wild)	0.0095	-0.0001	0.0011	-0.0123	-0.0032	0.0137				
Kadavu, Ravitaki (Wild)	0.0072	0.0003	-0.0039	-0.0112	-0.0063	0.0068	-0.0142			
Savusavu, Vatubukulaca (Farm)	0.0050	-0.0003	0.0014	0.0034	-0.0027	0.0023	0.0019	-0.0033		
Savusavu, Wailevu (Farm)	0.0025	0.0018	0.0104	0.0193	-0.0019	0.0013	0.0173	0.0112	0.0093	
Savusavu, Wailevu (Farm; hatchery)	0.0967*	0.0931*	0.0908	0.1016	0.0850*	0.0909	0.0932	0.0909	0.0873	0.0980*

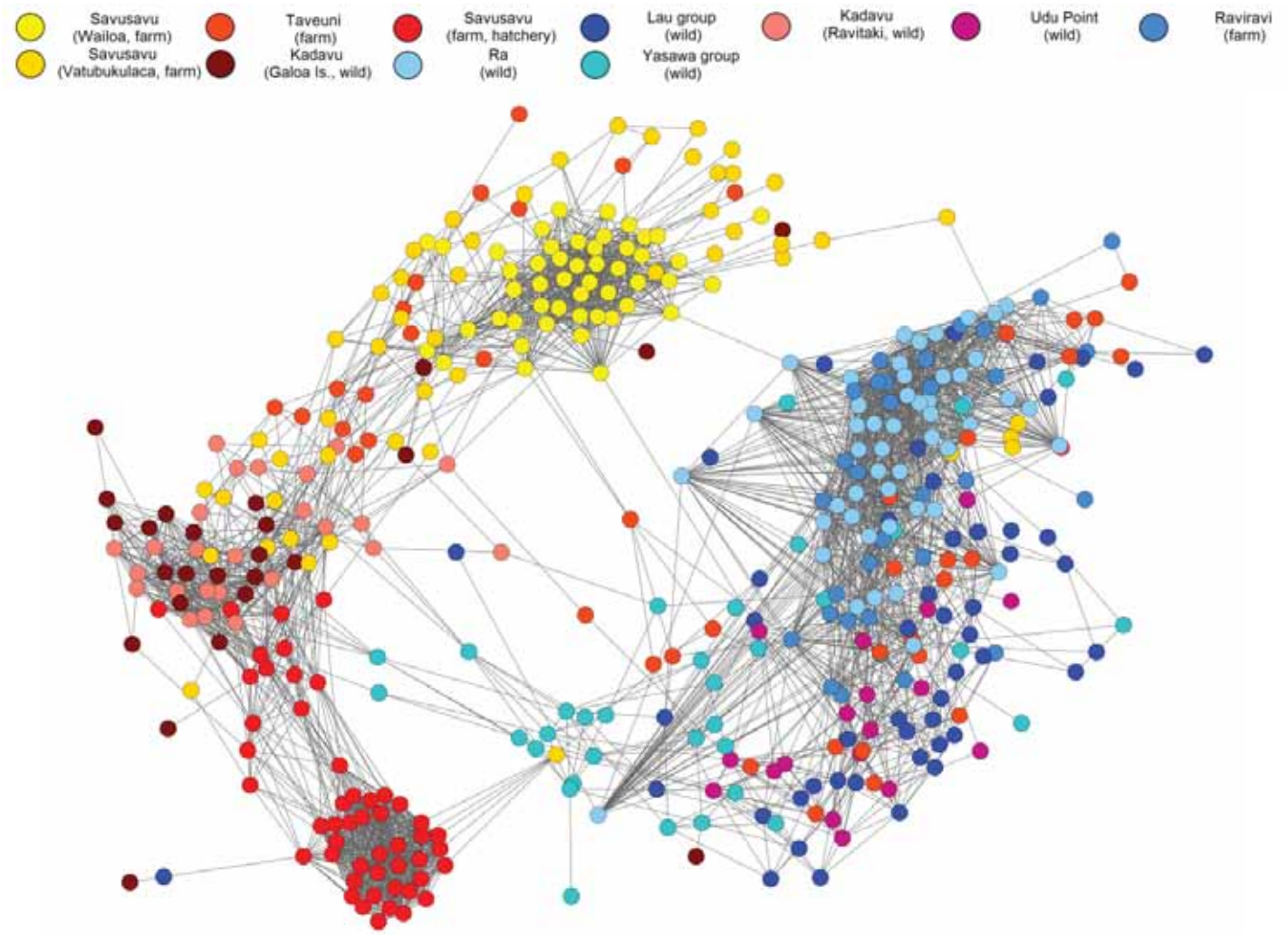


Figure 3.3 Population network of *P. margaritifera* individuals created using the Netview P v.0.4.2.5 pipeline (Steinig *et al.* 2016). The network has been visualised at a maximum number of nearest neighbour (k-NN) threshold of 40, using 4,123 SNPs and 427 individuals. Each dot represents a single individual, and population colours correspond with **Figure 3.1**.

Examination of fine-scale population sub-structure using the NetView P network (**Figure 3.3**) revealed a similar pattern of separation to the DAPC analysis, although with a greater level of individual resolution. Two large genetic groups were resolved, one of which incorporated six populations, while the other comprised a diffuse assemblage of the remaining five populations. The first group included the Savusavu (Wailevu) and Savusavu hatchery oysters, which formed two distinct clusters and remained separate from all other groups. Located between these two clusters, the two Kadavu, as well as the Taveuni and Savusavu (Vatubukulaca) populations also grouped together. The second larger group contained the Ra and Raviravi populations which formed a tight assemblage, along with a less compact cluster containing the Yasawa, Lau and Udu Point oysters. Connectivity between the two larger groups was limited to individuals belonging to the Yasawa, Taveuni, Savusavu (Vatubukulaca) and Lau populations. Identical trends were observed in networks constructed at lower k-NN values ranging from 5 to 35 (results not shown here), with the overall patterns of separation remaining consistent. Results of the hierarchical AMOVA were significant ($p<0.001$), and found that only 2% of the proportion of variation was attributable between wild and farm populations, whereas greater proportions were divided among individuals (68%), among populations (18%) and within individuals (12%).

3.3.4 Examination of adaptive variation

Detection of F_{st} outlier loci at three FDR thresholds of 0.01, 0.05 and 0.10 for each of the three pairwise population comparisons discovered between two and nine directional outlier SNPs jointly identified by Bayescan 2.1 and LOSITAN (**Table 3.3**). These pairwise population comparisons were carried out between Savusavu (Wailevu) and Lau, Udu Point and Kadavu (both populations considered together), as well as the Yasawa archipelago and Lau. These sites were located at maximum geographic distances across the Fiji Islands, positioned across environmental gradients (offshore island vs. mainland island and fringing vs. barrier reef habitats), as well as at opposing points along the major larval transport pathway identified from the particle dispersal simulation analysis. All directional outliers detected by Bayescan were also detected by LOSITAN, and no outlier loci were detected by either platform when all populations were considered together. Bayescan 2.1 analyses failed to detect any balancing outlier loci (zero or negative alpha values) for all pairwise population comparisons, and hence all balancing outliers reported were from LOSITAN computations. LOSITAN runs detected between 43 and 278 balancing loci across all three FDR thresholds

for each pairwise population comparison. In order to select an FDR threshold for accepting a final number of outlier loci for each comparison, QQ plots were constructed for each dataset at all three thresholds. A final stringent FDR threshold of 0.01 was selected on the basis of the QQ plots (**Appendix 3.6**), under which 5, 3 and 2 directional outlier loci were detected between the Savusavu (Wailevu)-Lau, Udu Point-Kadavu and Yasawa-Lau pairwise population comparisons, respectively.

Table 3.3 Numbers of putative directional and balancing F_{st} outlier loci discovered. Tests were carried out at three False Discovery Rate (FDR) thresholds using BayeScan 2.1 (Foll 2012) and LOSITAN (Antao *et al.* 2008). Jointly-identified loci were identified using both outlier detection platforms.

		Directional			Balancing		
Populations compared	**FDR**	**BayeScan 2.1**	**LOSITAN**	**Jointly-identified**	**BayeScan 2.1**	**LOSITAN**	**Jointly-identified**
Savusavu,	0.01	5	28	5	0	197	0
(Wailevu) and	0.05	8	46	8	0	206	0
Lau	0.10	9	96	9	0	248	0
Udu Point and	0.01	3	21	3	0	43	0
both Kadavu	0.05	3	37	3	0	108	0
populations	0.10	4	56	4	0	84	0
Yasawa and	0.01	2	18	2	0	201	0
Lau	0.05	3	46	3	0	278	0
	0.10	4	61	4	0	241	0

To gauge the strength of the selection signal, the average F_{st} values for all directional and balancing outlier loci detected were examined at the selected FDR of 0.01. For the Savusavu (Wailevu)-Lau comparison, the average Bayescan 2.1 F_{st} value was 0.1168. Similarly, average F_{st} values of 0.1025 and 0.1496 were observed for the Yasawa-Lau, and Udu-Kadavu comparisons, respectively. The average LOSITAN F_{st} values for the balancing outliers detected remained consistent for the Savusavu (Wailevu)-Lau, Udu-Kadavu and Yasawa-Lau comparisons, (-0.0343, -0.0464 and -0.0426, respectively). Given this set of results, it appears that any signatures of selection if present, are too weak to be detected and/or indecipherable from the background signal. This was supported by the construction of NJ trees to visualise population structure using directional outlier loci identified for each pairwise population comparison, based on 1-proportion of shared allele distances (results not shown here). All trees failed to show any separation between populations.

3.3.5 Particle dispersal modelling

Simulation of larval transport pathways with the hydrodynamic dispersal model demonstrated broad-scale mixture of particles by surface ocean current systems operating within the Fiji Islands; (see **Figure 3.4** for 2009 particle position outputs at 1, 15, 30 and 60 day time points and **additional file 1** for an animation of the full dispersal simulation over 100 days. 2010 data were very similar to 2009 patterns and are not presented here). A singular dispersal corridor appears to initially drive particles from all seed locations eastwards towards the Lau group of islands for a period of approximately 30 days; after which current movements oscillate across the centre of the Fiji group, while progressing in a southerly direction. Gene flow thus is likely to be homogenous between the Yasawa archipelago, Raviravi and Udu Point through the Bligh Water channel, towards sink locations in the Koro and Lau basins. Reef systems in the Lau group appear to receive particles from all locations in Fiji, although varying degrees of self-recruitment are likely for the Udu Point, Raviravi and Yasawa populations, due to the prevailing current dynamics and architecture of the Great Sea Reef system north of Vanua Levu retaining larvae in those regions. Despite this, a portion of larvae originating in the Yasawa archipelago may recruit along the western coastline of Viti Levu and Ra. Similarly, larvae which are exported from Udu Point and Raviravi may mix with individuals from Savusavu and Taveuni. The lowest degree of mixing is likely to occur between populations located along a north-south axis (e.g. Udu Point and Kadavu), as the dominant dispersal pathway operates in a west to east direction. Interestingly, the simulation indicates that if larvae advected from Kadavu and Lau survive beyond 40 days post-hatching, it may be possible for a few individuals to recruit eastwards onto the reefs of Tongatapu in the Kingdom of Tonga, (approximate position -175° longitude; see Day 60 output in **Figure 3.4**).

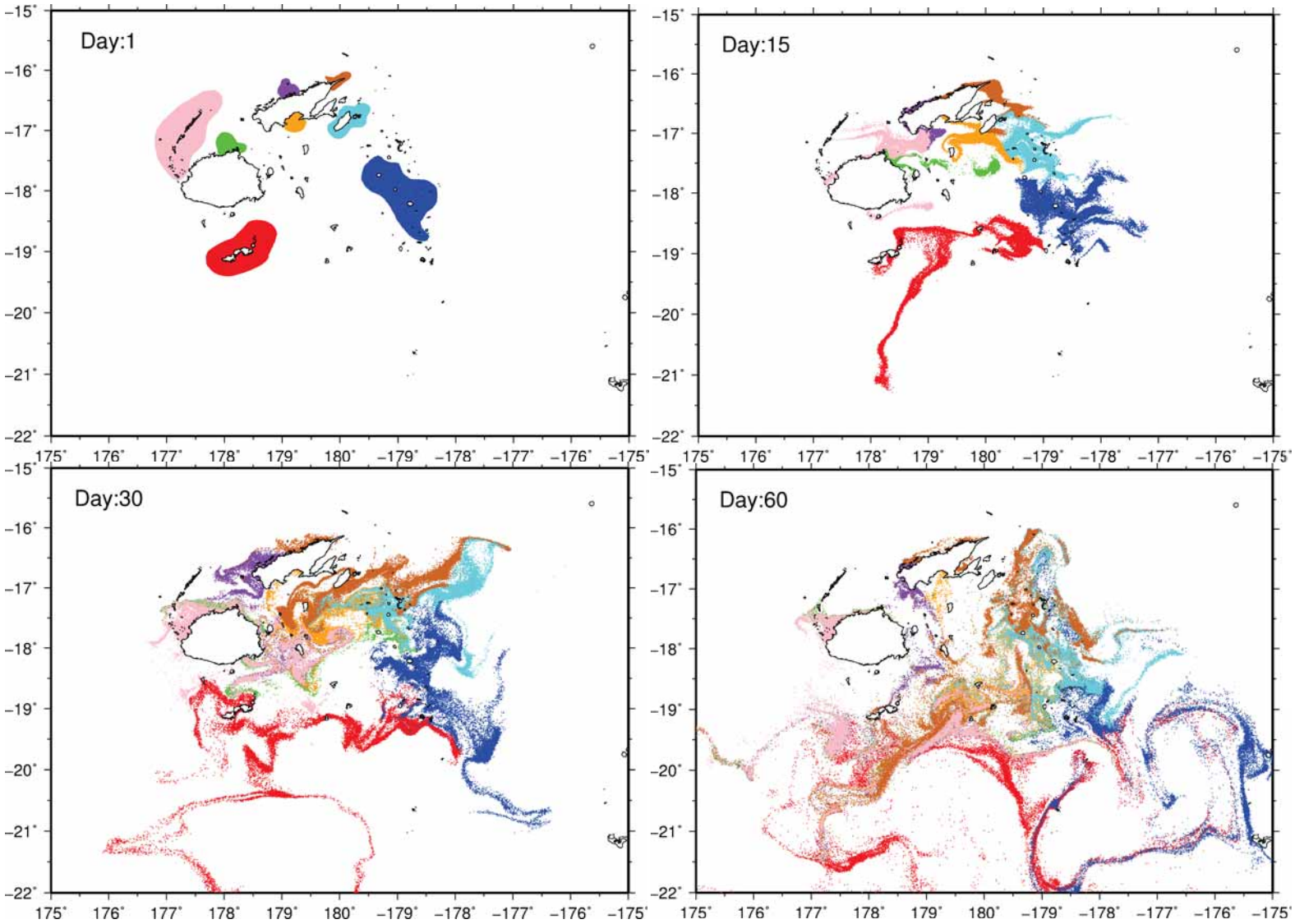


Figure 3.4 Results of 2009 particle dispersal simulation. Particle seed locations are shown in the day 1 position output, with the sampling regions colour coded as follows: Kadavu group (red), Yasawa group (pink), Ra (green), Raviravi (purple), Savusavu (orange), Udu Point (brown), Taveuni (light blue) and the central Lau group (dark blue). Simulated particle positions are shown at 15, 30 and 60 day outputs. An animation of dispersal simulation is provided as **additional file 1**.

3.4 Discussion

By independently evaluating population genomic analyses with hydrodynamic dispersal simulation, it was identified that Fijian *P. margaritifera* display a very shallow pattern of population structure, and are highly likely to constitute a single, biologically significant stock for fishery management. While diffuse patterns of population differentiation are apparent given the resolution of 4,123 SNPs used, the overall pairwise F_{st} estimates are small and not statistically significant (average overall F_{st}=0.0028; p>0.05). Given the largely homogenising larval mass transport pattern resolved using hydrodynamic dispersal simulation and the levels of relatedness between populations, the pattern of structure detected plausibly reflects fine-scale differentiation at the generational and family levels, together with small, isolated patches of localised recruitment (Arnaud-Haond *et al.* 2008). Furthermore, examination of loci under selection failed to detect any signatures of local adaptation, suggesting that environmental differences among populations are insufficiently heterogeneous to drive selection at the spatial scale examined (<400 km). Additionally, if weak local adaptation is present, the very high levels of gene flow between populations would likely override discernible signatures of selection. These results demonstrate the utility of independent population genomic and biophysical datasets for providing insights into the biology and ecology of a broadcast spawning bivalve, and have great potential for application to other marine species with similar life histories, where patterns of genetic structure and connectivity may not be well understood.

3.4.1 Resolution of population structure, diversity and relatedness

A weak pattern of population structure with high levels of connectivity was evident among all populations sampled using both broad-scale (DAPC) and fine-scale (NetView P) methods, mirroring the results of a previous study in Fiji (see chapter 2 and Lal *et al.* 2016b). Investigations of *P. margaritifera* populations elsewhere have yielded similar results, including French Polynesia (Arnaud-Haond *et al.* 2008; Lemer & Planes 2014) and Japan (Durand *et al.* 1993). Considering that *P. margaritifera* is a broadcast spawner with a relatively long PLD of between 26-30 days (Alagarswami *et al.* 1989; Doroudi & Southgate 2003), the degree of larval mixing driven by surface ocean currents (as suggested by the hydrodynamic dispersal simulation), supports the finding that Fijian oysters from all 11 locations sampled may be classified as a singular genetic entity.

Population pairwise F_{st} estimates indicated shallow and non-significant levels of structure, with the hatchery-produced oysters being the only population demonstrating detectable differentiation. This is not surprising considering that this population had undergone a genetic bottleneck through limited broodstock use, and differential larval mortality typical of hatchery rearing conditions. DAPC with BIC analysis, and NetView P network analysis both resolved similar cluster patterns, and overall patterns correlated well with F_{st} results and larval transport pathways inferred from particle dispersal simulation.

The levels of observed heterozygosity (H_o) detected were lower than expected across all populations (Table 1), keeping with the trend of heterozygote deficiency previously observed for *P. margaritifera* in Fiji (Lal *et al.* 2016b), French Polynesia (Arnaud-Haond *et al.* 2003b; Arnaud-Haond *et al.* 2008; Lemer & Planes 2012, 2014; Lemer *et al.* 2011) and Japan (Durand *et al.* 1993). Heterozygote deficits appear to be characteristic of a number of marine molluscs (Lind *et al.* 2009; Miller *et al.* 2013; Peñaloza *et al.* 2014), and in the current study are also likely due to a technical artefact associated with RADseq-based genotyping approaches, where restriction enzyme cut site polymorphisms may cause allelic dropouts (Andrews & Luikart 2014; Puritz *et al.* 2014). While stringent filtering measures were used to reduce the proportion of null alleles present in the final dataset, thorough testing of their effect on H_o, F_{is}, N_{eLD} and population differentiation estimates following the methods of Lal *et al.* (2016b) for *P. margaritifera*, revealed no impact on these metrics.

When assessing populations separately, estimates of individual average multi-locus heterozygosity (MLH), standardised heterozygosity (SH), inbreeding coefficient (F_{is}) and internal relatedness (IR) agreed with trends observed in H_o, which generally showed a lower diversity among pearl oysters sampled from island archipelago populations, compared to those from the larger land masses of Viti Levu and Vanua Levu (e.g. Av.MLH for the Kadavu (Galoa Island) and Raviravi (Vanua Levu) populations were 0.0695 cf. 0.1465 respectively). This observation may indicate higher rates of self-recruitment among island archipelago populations, and fits a growing body of evidence supporting significant self-recruitment for a number of broadcast spawning coral and reef fish species, with geographic setting strongly influencing the degree of larval retention within populations (Jones *et al.* 2009).

Patterns detected in the NetView P network, relatedness analyses and dispersal simulation all indicate support for this observation, as geographically distant populations clustered separately (e.g. Kadavu and Taveuni island sites), and shared fewer pairwise family relationships than others with higher degrees of connectivity either through proximity (e.g. Ra and Raviravi), or position within the major ocean current pathway (e.g. Yasawa and Lau). This was particularly evident between populations <150 Km apart containing 17-73 half-sibs, whereas populations situated farther apart held only 1-9. Examination of pairwise relationships between individuals within populations identified a larger number of full-sib and half-sib relationships for the bottlenecked hatchery produced population, as well as one farmed population sourced from spat collectors. For the latter, it is feasible that several individuals from one or more families remained poorly mixed in the plankton, and subsequently settled together on the spat collectors. This was suggested by Knutsen *et al.* (2011) for their study on Atlantic cod, and similar variability has been observed in hatchery-produced *P. maxima* (Lind *et al.* 2009; Lind *et al.* 2010).

Assessments of N_{eLD} and individual pairwise relationships within populations indicated a generally high degree of connectivity between populations. However, reduced N_{eLD} was detected for three farmed populations, one of which was a hatchery-produced cohort that had experienced a genetic bottleneck as a result of standard hatchery spawning practices (Durand *et al.* 1993; Lal *et al.* 2016b; Lind *et al.* 2009; Lind *et al.* 2010). A possible explanation for the lower N_{eLD} observed for the two other populations may be differential settlement and survival on the spat collectors these oysters were collected from, as previous studies have shown highly variable settlement, survival and predation rates of newly settled *P. margaritifera* spat on collector gear (Doroudi & Southgate 2002; Friedman & Bell 2000; Friedman & Southgate 1999; Pit & Southgate 2003).

The use of hydrodynamic modelling in parallel with genome-wide data for farmed and wild populations, adds fresh perspective for understanding the interaction of geographic and oceanographic influences contributing to population genetic structure in *P. margaritifera*. Studies on the genetic stock structure of this species predominantly originate in French Polynesia, where oysters are found in three distinct types of reef environments (Lemer & Planes 2012, 2014). These comprise high island lagoons with fringing and barrier reef systems with open oceanic circulation (similar to those found in Fiji), atoll lagoons also with open circulation, and closed atoll lagoons with highly reduced circulation (Arnaud-Haond *et*

al. 2008; Lemer & Planes 2012, 2014). Lemer and Planes (2014) detected connectivity at both small (less than 500 km) and large (greater than 1500 km) spatial scales between French Polynesian archipelagos which had open oceanic circulation patterns, mirroring the results of observations for Fijian populations, but also found significant genetic structure for oysters contained within closed atoll lagoons.

3.4.2 Examination of adaptive variation

Understanding levels of adaptive variation is critical for management of translocation, population supplementation and/or assisted migration, in order to avoid negative consequences such as outbreeding depression that may result from moving individuals into an environment they may be maladapted to (Funk *et al.* 2012; Nosil *et al.* 2009). This latter consideration is especially important for aquaculture, as productivity is heavily reliant on stock fitness (Jerry *et al.* 2012; Kvingedal *et al.* 2010; Kvingedal *et al.* 2008). Knutsen *et al.* (2011) in their study on Atlantic cod also failed to detect signatures of selection, despite the species having an extensive north Atlantic natural distribution over known salinity and temperature clines. An explanation they offer for this finding is that their work examined a restricted geographical range, where environmental differences may be small, relative to conspecifics occupying more heterogeneous habitats over the broader species distribution. The situation may be similar for *P. margaritifera* in the present study, and examination of populations across larger spatial scales beyond the Fiji Islands should provide further insights.

The inability of F_{st} outlier testing to discern signatures of selection possibly indicates that the environments oysters were sampled from may be insufficiently heterogeneous to drive local adaptation at an easily detectable threshold. Further considerations include the type of trait under selection (polygenic or monogenic), as well as the opposing dynamics of gene flow against the strength of selection. That is, where local adaptation is present, it may be too weak to be detected by the SNP marker set used and lost to background noise. Nayfa and Zenger (2016) examined three populations of the closely related silver-lip pearl oyster *P. maxima,* from Bali, West Papua and Aru in Indonesia, which were subject to a complex system of prevailing and seasonally reversing surface ocean currents. Evidence of directional selection was detected despite high levels of gene flow, causing divergence between oysters

from Bali and West Papua against those from Aru, and the recommendation for aquaculture was to manage the Aru population separately from Bali and West Papua.

3.4.3 Particle dispersal modelling

Examination of larval dispersal patterns using hydrodynamic modelling alone has been used for a number of marine taxa (Neo *et al.* 2013; Wood *et al.* 2014), including *P. margaritifera* (Thomas *et al.* 2014), but comparatively few studies have sought to combine larval dispersal data with genome-wide population information. Among studies which have coupled oceanographic and genetic methods are White *et al.* (2010a), Galindo *et al.* (2006) and Dao *et al.* (2015) using microsatellite loci, however, the limited number of these markers have provided finite information about fine-scale population structure and adaptive variation (Stapley *et al.* 2010; Zarraonaindia *et al.* 2012).

The discovery of homogenised surface ocean current movement towards the Lau archipelago is well supported by the results of population genomic analyses presented here, particularly regarding broad and fine-scale population differentiation, genetic diversity levels and lack of adaptive variation within and among populations. It is interesting that the major larval sink location is situated in the Lau archipelago, which retained consistency across the 2009-2010 ENSO period. Further examination of fine-scale larval transport pathways is warranted to determine the degree of mixing within the Lau group, and to see if any settlement heterogeneity occurs there. Unfortunately, this was beyond the capability of the HYCOM hydrodynamic model used here, as the data is not captured at a resolution finer than a grid size of 10 km^2 (Chassignet *et al.* 2007; Halliwell 2004). The HYCOM model is the only hydrodynamic model available for the Fiji Islands, however, given the future availability of a finer resolution model, gaining these insights is possible.

For broadcast spawning marine taxa with extended PLD, the inclusion of hydrodynamic dispersal data to better understand population connectivity in the marine environment is indispensable, as assessment of the magnitude of larval movements, along with patterns of current-driven differential recruitment may become possible. Work by Thomas *et al.* (2014) in French Polynesia on connectivity between populations discovered that larval sink and source locations for *P. margaritifera* accounted for 26% and 59% of the variation observed

respectively, underscoring its importance for larval supply and management of farmed and wild pearl oysters.

3.4.4 Implications for fishery management

The persistent problem in stock assessment investigations of determining "biologically meaningful" genetic divergence between populations requires careful evaluation on a case by case basis, with respect to the biological questions being answered (Reiss *et al.* 2009), fishery management goals and the characteristics of the organism(s) involved (Knutsen *et al.* 2011; Waples *et al.* 2008). For high gene flow species where fine-resolution population genomic analyses detect weak divergence by examining neutral and adaptive variation, the use of independent environmental data provides important additional knowledge for informed fishery management decision making.

Given the findings of non-significant population differentiation and the absence of signatures of selection or apparent phenotypic differences among populations, these data support the existence of a singular, biological stock in the Fiji Islands. This suggests that fishery management of *P. margaritifera* in Fiji may be based upon treatment of all sampling locations as one cohesive group (Management Unit). Further evidence of this is found in the independent assessment of population connectivity by hydrodynamic dispersal simulation, which confirms broad scale panmixia across all populations. This finding is promising for developing aquaculture of this species in the country, as it may mean that spat collected in locations which freely exchange recruits can also be grown-out among them (e.g. Kadavu, Ra, Savusavu, Taveuni and Lau). For those populations which experience less connectivity (e.g. Yasawa, Raviravi and Udu Point), further investigation is required to determine if any negative consequences may result from either keeping these groups isolated, or opening them up to translocation.

The small spatial scale of the Fiji Islands and high levels of gene flow apparent for Fijian *P. margaritifera*, may actually facilitate uncomplicated fishery management and aquaculture development of this species in the country, compared to other locations such as French Polynesia, where oysters are distributed over larger scales and across heterogeneous habitats (Lemer & Planes 2014). For French Polynesian populations, Lemer and Planes (2012) and Arnaud-Haond *et al.* (2003b) reported that farmed populations originally sourced from

genetically distinct wild oysters over a period of 20 years, had accumulated higher levels of genetic diversity than their progenitors, potentially providing a risk of outbreeding depression for wild oysters interbreeding with farmed individuals. While it is unlikely that a similar situation could occur for Fijian *P. margaritifera*, there are important lessons to be learnt from the French Polynesian experience. If hatchery production of spat outpaces the collection of wild spat as the primary source of oysters for grow out in the future, any potentially negative consequences as a result of genetic pollution effects could be minimised by careful selection of broodstock to maintain levels of genetic fitness.

3.5 Conclusions

The use of genome-wide SNP data and hydrodynamic particle dispersal modelling have provided valuable insights into the population structure and connectivity of the black-lip pearl oyster in the Fiji Islands, filling a substantial knowledge gap on the stock structure of this species in the country. Simulation of larval transport with hydrodynamic dispersal modelling confirmed the existence of broad-scale connectivity by surface ocean current systems, correlating very well with patterns of differentiation, heterozygosity and adaptive variation discovered in the genetic data. There is strong support for the existence of a singular stock structure in the Fiji Islands, which is promising for developing aquaculture of this species in the country, as it indicates that germplasm transfer is possible between locations that freely exchange recruits. The combined use of both selectively neutral and loci under selection to elucidate fine-scale population variability (or the lack thereof), has high utility for stock assessment in high gene flow species, where biologically meaningful levels of divergence are not immediately apparent. Furthermore, independent assessment of connectivity using environmental data such as particle dispersal simulation, can provide valuable additional information for making fishery management decisions, when patterns in genetic data don't easily lend themselves to the identification of stock boundaries. This study highlights the value of using both genomic and hydrodynamic data, for a comprehensive understanding of population structure and connectivity in broadcast-spawning marine taxa, and utilising the information collectively for aquaculture and sustainable fishery management.

CHAPTER 4: Assessment of range-wide genetic structure and connectivity of *P. margaritifera*

4.1 Introduction

Understanding the patterns and processes shaping population genetic structure across the extent of a species' distribution is an important prerequisite for biological conservation and management efforts, as well as studies of speciation (Guo 2012). For marine taxa, regional fishery management and aquaculture practices also rely on biologically meaningful population structure to delineate discrete stocks (André *et al.* 2011; Lal *et al.* 2016a; Waples *et al.* 2008). The ability to quantify genetic variation across geographical limits may shed light on why a species might demonstrate stable range boundaries, and also permit assessment of the conservation value of central (*C*) versus marginal (*M*) populations (Eckert *et al.* 2008; Guo 2012; Sexton *et al.* 2009). Several studies (reviewed by Eckert et al (2008) and Sexton et al (2009)), have investigated the central-marginal (*C*-*M*) hypothesis, also known as the core-periphery hypothesis (CPH; Brussard 1984; Eckert *et al.* 2008; Liggins *et al.* 2015). While many comparisons between taxa have revealed a general decline in genetic diversity and increased differentiation towards range margins, others show no clear patterns (Guo 2012).

It is expected that the interplay of microevolutionary processes, (namely natural selection, genetic drift and gene flow), will largely determine the magnitude and extent of population structure and connectivity, although the spatial distribution and demographic characteristics of the species may also exert strong influences (Eckert *et al.* 2008; Sexton *et al.* 2009). The CPH provides a model for interpreting how microevolutionary forces shape genetic divergence patterns throughout a species' range. Under this model, a species which colonises a geographical gradient of environmental conditions, is over time expected to exhibit maximised abundance (highest survival, reproduction and growth rates), around a central point where conditions are optimal, while populations become smaller, more fragmented, increasingly divergent and influenced by selective forces towards the periphery (Brussard 1984; Eckert *et al.* 2008; Vucetich & Waite 2003). However, exactly how the abundant centre distribution relates to the partitioning of genetic diversity, patterns of differentiation

and adaptive differences across the *C-M* cline, remains a contentious topic (Eckert *et al.* 2008; Vucetich & Waite 2003). One explanation offered suggests that both effective population size (N_e) and gene flow (m) should be highest at the centre, and lowest at range margins. Consequently, central populations are expected to be less genetically differentiated and possess higher levels of genetic diversity, than those existing at range margins (Brussard 1984; Eckert *et al.* 2008). Furthermore, due to environmental heterogeneity across a *C-M* cline, local adaptation may be observed when comparing populations existing at the core and range peripheries.

While several studies have examined *C-M* genetic patterns in terrestrial taxa (Eckert *et al.* 2008; Hardie & Hutchings 2010), comparatively few investigations have involved marine species (Liggins *et al.* 2015), and marine invertebrates in particular (Liggins *et al.* 2014). Marine systems introduce several challenges for range-wide studies, as >70% of invertebrates and many vertebrates are characterised by large population sizes, high fecundity, external fertilisation and larvae that typically remain in the plankton for several weeks, although this may vary anywhere from a few minutes to years (Broquet *et al.* 2013; Hellberg *et al.* 2002; Limborg *et al.* 2012; Shanks 2009; Strathmann & Strathmann 2007). Consequently, *C-M* patterns compared to terrestrial taxa may differ from expectations under the CPH, as the homogenising influence of gene flow may maintain high connectivity across the *C-M* cline (Liggins *et al.* 2015). Furthermore, divergence and local adaptation may not be as apparent if populations remain highly connected, and environmental gradients are shallow.

Among marine invertebrates, species which are either completely sessile as adults (e.g. barnacles, sponges and ascidians), or possess very limited mobility (e.g. sea urchins, bivalves, gastropods), present additional challenges for assessment of *C-M* trends (Reitzel *et al.* 2013; Thorpe *et al.* 2000). As larvae undergo pelagic dispersal and recruitment, differential selective pressures and survival rates pre- and post-settlement between the plankton and benthos may strongly influence the genetic composition of populations (Addison & Hart 2005; Arnaud-Haond *et al.* 2008). Furthermore, this could mean that the spatial distribution of a population may be limited to isolated biodiversity hotspots (e.g. single bivalve beds), or an entire reef shelf (Gosling 2015; Gosling & Wilkins 1985).

Given the complex nature of the biological and environmental influences at play, it is important to consider multiple sources of information for distribution-wide investigations in

the marine environment, particularly when the species being examined is extensively distributed across heterogenous habitats. Considerations that have been highlighted in previous analyses of *C-M* patterns involving terrestrial taxa, include examination of the geographical direction of the periphery studied, latitudinal effects, the effects of species-range geometry (e.g. shape and size), as well as sampling strategy (Eckert *et al.* 2008; Guo 2012; Hardie & Hutchings 2010). While not all of these may apply to marine scenarios, for taxa that employ a broadcast spawning reproductive strategy, consideration of the extent of ocean current-mediated larval dispersal addresses many of these points (Dao *et al.* 2015; Gaggiotti *et al.* 2009; Hare *et al.* 2000; Lal *et al.* 2016a; Pujolar *et al.* 2011).

Incorporation of environmental data such as dispersal modelling into range-wide studies is capable of offering unprecedented insights into larval dispersal limits (Dao *et al.* 2015; Lal *et al.* 2016a; Neo *et al.* 2013; Thomas *et al.* 2014; Wood *et al.* 2014), and when considered together with both neutral and adaptive patterns of population structure, permit a holistic assessment of concordance with the CPH, or other models of range-wide structuring. The advantage of using independent datasets also includes the potential to reveal and/or corroborate previously undiscovered or poorly understood biogeographic barriers to dispersal, cryptic speciation and regional local adaptation (Grosberg & Cunningham 2001; Krück *et al.* 2013; Nayfa & Zenger 2016; Nielsen *et al.* 2009).

The black-lip pearl oyster *P. margaritifera* (Pteriidae), is a marine bivalve mollusc that has a broad Indo-Pacific distribution (**Figure 4.1**), and is highly valued for cultured pearl and pearl shell production (Southgate *et al.* 2008; Wada & Tëmkin 2008). Aquaculture of this species comprises a valuable industry and important source of coastal community livelihood across almost the entire extent of its distribution (Southgate *et al.* 2008; SPC 2003). While analyses to examine its population structure and connectivity have previously been carried out, these have produced mixed findings, incorporated a range of different marker types (allozymes, mtDNA and microsatellites), and never examined the entirety of the species distribution (Arnaud-Haond *et al.* 2003a; Arnaud-Haond *et al.* 2004; Arnaud-Haond *et al.* 2008; Benzie & Ballment 1994; Durand & Blanc 1988; Durand *et al.* 1993; Lemer & Planes 2012, 2014). The current species description includes a total of six sub-species (Cunha *et al.* 2011; Gervis & Sims 1992; Wada & Tëmkin 2008), which are described exclusively on the basis of variable morphological characters (Jameson 1901). In the Pacific basin, Hawaiian populations are known as *P. margaritifera* var. *galstoffi* (Bartsch, 1931), Cook Islands and

French Polynesian individuals as *P. m.* var. *cummingi* (Reeve, 1857), and all Central and Western Pacific specimens as *P. m.* var. *typica* (Linnaeus, 1758). Indian Ocean populations are represented by *P. m.* var. *persica* (Jameson, 1901; Persian Gulf), *P. m.* var. *erythraensis* (Jameson, 1901; Red Sea) and *P. m.* var. *zanzibarensis* (Jameson, 1901; East Africa, Madagascar and Seychelle Islands (Gervis & Sims 1992)).

Significant genetic heterogeneity has been reported for *P. margaritifera* at nuclear markers (allozymes, anDNA markers and microsatellite loci), at various sites in the Western and Central Pacific (Arnaud-Haond *et al.* 2003a; Durand & Blanc 1988; Wada & Jerry 2008), while contrastingly mitochondrial markers did not (Arnaud-Haond *et al.* 2003a). More recent work using microsatellite loci discovered significant genetic structure both within and between French Polynesian island archipelagos, attributed to "open" and "closed" atoll lagoon hydromorphologies restricting patterns of gene flow (Lemer & Planes 2014). Since then, genome-wide SNPs have been developed and characterised (Lal *et al.* 2016b), and used to investigate stock structure for fishery management and aquaculture in the Fiji Islands (Lal *et al.* 2016a), where a single genetic stock was identified.

Previous studies of range-wide genetic structuring in Pteriid pearl oysters have produced mixed results. Lind *et al.* (2007) reported a reduction in genetic diversity towards the range periphery of the silver-lip pearl oyster, *P. maxima*, which is consistent with CPH assumptions. However, the natural distribution of this species is considerably less extensive than that of *P. margaritifera* (Southgate *et al.* 2008; Wada & Tëmkin 2008). A taxon which has a range similar to that of *P. margaritifera* is the Akoya pearl oyster, currently recognised as the *P. fucata*/*martensii*/*radiata*/*imbricata* species complex (Tëmkin 2010; Wada & Tëmkin 2008). While the population genetic structure of the Akoya complex is pending resolution, it is thought that it may comprise one cosmopolitan, circum-globally distributed species, possessing a very high degree of intraspecific variation across its range (Southgate *et al.* 2008; Wada & Tëmkin 2008; Yu & Chu 2006).

Larval development of *P. margaritifera* occurs over 26-30 days in captivity (Alagarswami *et al.* 1989; Doroudi & Southgate 2003), however, time to settlement may be prolonged if conditions are unfavourable (Pechenik 2006). The high dispersal potential (and thus gene flow) in this species suggests that CPH trends may not be easily identifiable across the broader species range, except perhaps in situations where larval dispersal is restricted by

seascape features (e.g. closed atoll lagoons or current gyres), or at the very limits of the species distribution where favourable habitat is limited, impacting fitness and population growth. Here, I assess populations of *P. margaritifera* across the extent of its Indo-Pacific distribution spanning over 18,000 km, and compare observations with expectations under the CPH and regional morphological sub-divisions. Independent population genomic and hydrodynamic approaches were utilised to assess population genetic structure, adaptive variation and larval connectivity. Through the use of independent biological and environmental datasets, this work sheds light on the links between genetic structure, ecology and oceanography, to reveal how populations of a broadcast spawner can be organised and maintained in the marine environment.

4.2 Methods and materials

4.2.1 Specimen collection, tissue sampling and DNA extraction

Adult and juvenile *P. margaritifera* (n=580) between 5-18 cm in DVM were collected from 14 sites across the species distribution (**Figure 4.1**). All oysters were handled in accordance with James Cook University's animal ethics requirements and guidelines, with permission to collect tissues obtained from local authorities. In the Indian Ocean, oysters were collected from two sites in Tanzania (Mafia Island and Mtwara, n=35 and n=20 respectively), the Persian Gulf (Hendorabi Island, Iran; n=49) and Post Office Island in the Abrolhos Islands group, Western Australia (n=50). All Indian Ocean samples consisted of wild individuals with the exception of the Abrolhos Islands collection, where oysters were hatchery-produced from wild-caught broodstock. In the Western Pacific, oysters were sampled from Checheng, Taiwan (n=24), Nha Trang, Vietnam (n=47) and Manado, Indonesia (n=48). Central Pacific locations were represented by Kavieng, Papua New Guinea (n=38), Gizo Island in the Solomon Islands (n=50), the Great Barrier Reef, Australia (n=35), Savusavu, Yasawa and Lau in the Fiji Islands (n=61) and Tongatapu, Tonga (n=28). In the eastern Pacific, oysters were collected from Manihiki Atoll in the Cook Islands (n=45), and Arutua, French Polynesia (n=50). All Pacific Ocean samples consisted of wild oysters, with the exception of the Cook Islands and French Polynesian samples that were sourced from pearl farm stocks.

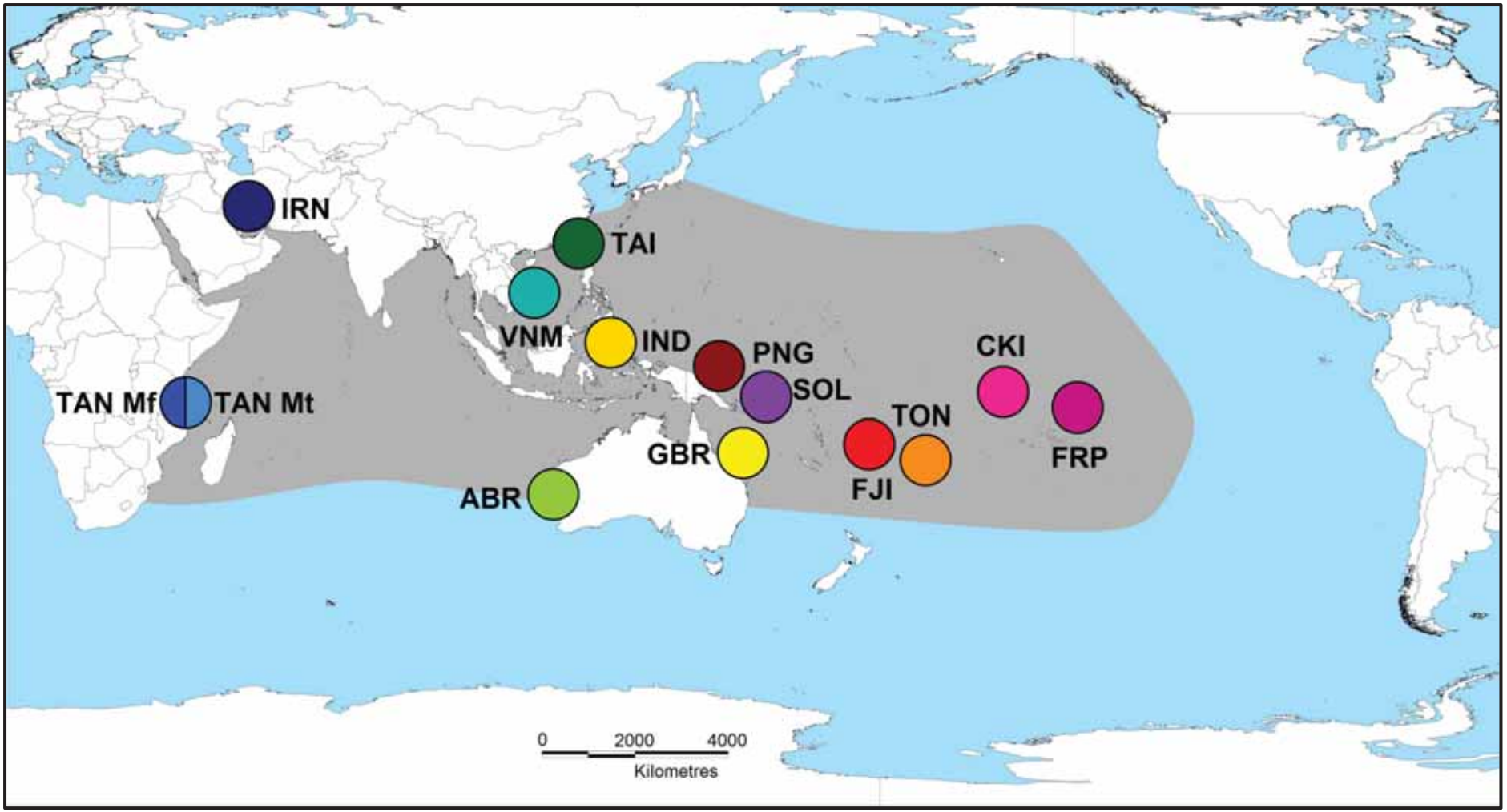


Figure 4.1. Map of global sampling locations from where 580 individuals of *P. margaritifera* were collected. The approximate known distribution and range of the species is presented in grey, and adapted from Wada and Tëmkin (2008). Site codes represent the following locations: **TAN Mf**: Mafia Island, Tanzania (dark blue); **TAN Mt**: Mtwara, Tanzania (light blue); **IRN**: Hendorabi Island, Iran; **TAI**: Checheng, Taiwan; **VNM**: Nha Trang, Vietnam; **IND**: Manado, Indonesia; **AU Abr**: Abrolhos Islands, Australia; **AU GBR**: Great Barrier Reef, Australia; **PNG**: Kavieng, Papua New Guinea; **SOL**: Gizo Island, Solomon Islands; **FJI**: Kadavu, Savusavu, Lau and the Yasawa group, Fiji Islands; **TON**: Tongatapu, Tonga; **CKI**: Manihiki Atoll, Cook Islands and **FRP**: Arutua, French Polynesia.

Proximal mantle and adductor muscle tissues (3 and 6 cm respectively), were collected and handled for gDNA extraction as described in chapter 2. All gDNA samples were subsequently normalised at 100 ng/μL in a 50 μL final volume, and submitted for DArTseq™ 1.0 genotyping at Diversity Arrays Technology PL, Canberra, ACT, Australia.

4.2.2 DArTseq™ 1.0 library preparation and sequencing

Diversity Arrays Technology (DArT PL) proprietary genotyping by sequencing (DArTseq™) reduced-representation libraries were prepared as described by Kilian *et al.* (2012) and Sansaloni *et al.* (2011), with a number of modifications for *P. margaritifera*. Briefly, genome complexity reduction was achieved with a double restriction digest, using a *Pst*I and *Sph*I methylation-sensitive restriction enzyme (RE) combination, in a joint digestion-ligation reaction at 37 °C for 2 hr with 150-200ng gDNA. Because *P. margaritifera* like other bivalve species is highly polymorphic (Harrang *et al.* 2013; Lal *et al.* 2016b), highly repetitive genomic regions were avoided and low copy regions more efficiently targeted for sequence capture with the use of methylation-sensitive REs (Elshire *et al.* 2011).

Custom proprietary barcoded adapters (6-9 bp) were ligated to RE cut-site overhangs as per Kilian *et al.* (2012), with the adapters designed to modify RE cut sites following ligation, to prevent insert fragment re-digestion. The *Pst*I-compatible (forward) adapter incorporated an Illumina flowcell attachment region, sequencing primer and a varying length barcode region (Kilian *et al.* 2012; Ren *et al.* 2015). The reverse adapter also contained a flowcell attachment region, and was compatible with the *Sph*I cut-site overhang. Samples were processed in batches of 94, with 15% of all samples in a batch randomly selected for replication, to provide a basis for assessing region recovery and genotyping reproducibility. Target "mixed" fragments (Ren *et al.* 2015), containing both *Sph*I and *Nla*III cut-sites were selectively amplified using custom designed primers for each sample, under the following PCR conditions: initial denaturation at 94°C for 1 min, then 30 cycles of 94°C for 20s, 58°C for 30s and 72°C for 45s, followed by a final extension step at 72°C for 7 min. Amplified samples were subsequently cleaned using a GenElute PCR Clean-up Kit (Sigma-Aldrich, cat.# NA1020-1KT), on a TECAN Freedom EVO150 automated liquid handler.

To examine fragment size concordance and digestion efficiency, all samples were visualised on a 0.8% agarose gel stained with ethidium bromide (EtBr), and quantified using the ImageJ

software package (Mateos & Pérez 2013). Samples which did not appear to have undergone complete digestion and/or amplification were removed from downstream library preparation. A total of 580 samples were each normalised and pooled using an automated liquid handler (TECAN, Freedom EVO150), at equimolar ratios for sequencing on an Illumina HiSeq 2500 platform. After cluster generation and amplification (HiSeq SR Cluster Kit V4 cBot^{TM}, cat.# GD-401-4001), 77 bp single-end sequencing was performed at the DArT PL facility in Canberra, Australia.

4.2.3 Sequence quality control, marker filtering and genotype calling at DArT PL

Raw reads obtained following sequencing were processed using Illumina CASAVA v.1.8.2 software for initial assessment of read quality, sequence representation and generation of FASTQ files. Filtered FASTQ files were then supplied to the DArT PL proprietary software pipeline DArTtoolbox, which performed further filtering, variant calling and generated final genotypes in sequential primary and secondary workflows (Cruz *et al.* 2013). Within DArTtoolbox, the primary workflow first involved the package DArTsoft14 to remove reads with a quality score <25 from further processing, and apply stringent filtering to the barcode region of all sequences to increase confidence in genomic region recovery. Individual samples were then de-multiplexed by barcode, and subsequently aligned and matched to catalogued sequences in both NCBI GenBank and DArTdb custom databases to check for viral and bacterial contamination, with any matches removed from further processing.

The secondary workflow employed the DArTsoft14 and KD Compute packages along with the DArTdb database, to identify polymorphisms by aligning identical reads to create clusters across all individuals sequenced. These clusters were then catalogued in DArTdb, and matched against each other to create reduced-representation loci (RRL), based on their degree of similarity and size. SNP and reference allele loci were identified within clusters and assigned the following DArT scores: "0"=reference allele homozygote, "1"=SNP allele homozygote and "2"=heterozygote, based on their frequency of occurrence. To ensure robust variant calling, all monomorphic clusters were removed, SNP loci had to be present in both allelic states (homozygous and heterozygous), and a genetic similarity matrix was produced using the first 10,000 SNPs called to assess technical replication error (Robasky *et al.* 2014), and exclude clusters containing tri-allelic or aberrant SNPs and overrepresented sequences.

Once SNP markers had been confidently identified, each locus was assessed in the KD Compute package for homozygote and heterozygote call rate, frequency, polymorphic information content (PIC), average SNP count, read depth and repeatability, before final genotype scores were supplied by DArT PL. Following the receipt of genotype data from DArT PL, the dataset was further filtered to retain only a single, highly informative SNP at each genomic locus. This was achieved by filtering out duplicate SNPs (possessing identical Clone IDs), according to call rate and MAF. Subsequently, loci were screened for call rate, average Polymorphic Information Content (PIC), MAF and average repeatability, to retain SNPs suitable for population genomic analyses. All loci were then tested for departure from HWE using Arlequin v.3.5.1.3 (Excoffier *et al.* 2005), using an exact test with 10,000 steps in the Markov Chain and 100,000 dememorisations. Additionally, all loci were tested for genotypic linkage disequilibrium (LD) in Genepop v.4.3 (Rousset 2008), as per Lal *et al.* (2016b). Two separate datasets were then created, one which contained selectively neutral loci, and the other which included loci putatively under selection. Bayescan v.2.1 and LOSITAN software were used to detect loci under selection, and further details are provided under that section of the methods.

4.2.4 Evaluation of genomic diversity, inbreeding and population differentiation

For assessment of genomic diversity within and between populations, allelic diversity indices were calculated as previously outlined in chapters 2 and 3. Additionally, the mean number of alleles per locus (*A*) was computed using the *diveRsity* (Keenan *et al.* 2013) *R* package, along with the number of private alleles (A_p) with HP-RARE v.1.0 (Kalinowski 2004), according to population groups identified from Netview P and DAPC analyses (see results), due to the levels of genetic divergence observed. Furthermore, rare allelic richness (*Ar*, <5% MAF) was computed manually for each population.

4.2.5 Resolution of broad and fine-scale population structure and connectivity

Pairwise F_{st} estimates for each population were calculated using Arlequin v.3.5.1.3 with 10,000 permutations (Excoffier *et al.* 2005), along with a hierarchical Analysis of Molecular Variance (AMOVA) in the *R* package *Poppr* (Kamvar *et al.* 2014). The AMOVA examined variation between individuals, populations and regions (Pacific vs. Indian Ocean basins). To

assess an isolation by distance (IBD) model of gene flow among populations, Mantel tests were carried out using GenAlEx v.6.5 (Peakall & Smouse 2006), based on pairwise F_{st} and straight-line geographic distance matrices over 10,000 permutations. Mantel tests were performed considering populations within each ocean basin together, separately, and also within Pacific Ocean population clusters identified by DAPC and NetView P analyses. Nei's (1978) standard genetic distances (D_S) between populations were also computed in Genetix v.4.05 with 10,000 permutations (Belkhir *et al.* 1996), and broad-scale population structure visualised by performing a Discriminant Analysis of Principal Components (DAPC) in the *R* package *adegenet* 1.4.2 (Jombart 2008; Jombart & Ahmed 2011; Jombart *et al.* 2010). The DAPC was carried out for all loci, and α-score optimisation used to determine the number of principal components to retain. To reveal any fine-scale stratification within and among all populations, network analysis was carried out using the NetView P pipeline v.0.4.2.5 (Neuditschko *et al.* 2012; Steinig *et al.* 2016). To further investigate the direction and magnitude of migration between populations, migration networks were generated using the *divMigrate* function of the *R* package *diveRsity*, utilising the Nei's G_{st} method (Keenan *et al.* 2013; Sundqvist *et al.* 2013).

4.2.6 Examination of adaptive variation

To first create a selectively neutral dataset for population genomic analyses, a filtered dataset containing 10,683 SNP loci was used as the starting point for this step. Both BayeScan v.2.1 (Foll 2012; Foll & Gaggiotti 2008) and LOSITAN selection detection workbench (Antao *et al.* 2008) software packages were employed to identify candidate loci under selection, at FDRs=0.001, 0.005, 0.01, 0.05 and 0.1 and 0.2. Comprehensive descriptions of the settings used for both software packages were as described in chapter 3. The numbers of loci detected are summarised in **Appendix 4.3**, and verification of these loci was carried out using QQ plots (data not shown)**.** The intended approach was to select loci jointly identified by both Bayescan 2.1 and LOSITAN, however, given the tendency of LOSITAN to overestimate the numbers of loci under selection (Lal *et al.* 2016b; Narum & Hess 2011; Nayfa & Zenger 2016), and disagreement on an appropriate FDR threshold to apply using both methods, a conservative approach was taken where LOSITAN results were disregarded, and the Bayescan 2.1 results at an FDR=0.01 considered. This indicated that a total of 1,059 putatively balancing and directional loci were present in the dataset, and following their removal, a selectively neutral dataset containing 9,624 SNPs remained.

Further population-specific F_{st} outlier tests were used to detect local adaptation, with population pairs tested at FDRs of 0.001, 0.005, 0.01, 0.05 and 0.1 and 0.2. However, testing for F_{st} outliers was restricted to populations sampled from the Pacific Ocean basin, as they were the least differentiated amongst themselves (*i.e.* lowest neutral F_{st} levels <0.11; see results), while all Indian Ocean populations were significantly more divergent. Results of the Bayescan 2.1 and LOSITAN analyses, together with the construction of pairs of Quantile-Quantile plots (QQ-plots), were used to assess the suitability of an FDR threshold for outlier detection between the two methods.

4.2.7 Particle dispersal simulation

To independently evaluate larval connectivity using oceanographic data for comparison with population genomic analyses, larval transport pathways between sampling locations were simulated using the particle dispersal modelling software DisperGPU (https://github.com/CyprienBosserelle/DisperGPU).

4.2.7.1 Hydrodynamic and dispersal numerical models

The particle dispersal model was configured and driven by current velocity output from the global HYbrid Coordinate Ocean Model (HYCOM) data (Chassignet *et al.* 2007; Cummings 2005), as described in chapter 3.

4.2.7.2 Model configuration

Particles were seeded in 11 locations corresponding to locations from where oysters were sampled for genetic analyses (see **Figure 4.5**), which were represented at scales larger than the precise sampling locations, to factor in the extent of surrounding coral reef habitat, as described in chapter 3. All seed areas were also extended farther offshore to account for the fact that the HYCOM model is not adapted for shallow water environments, and does not resolve fine-scale hydrodynamic patterns <10 km (Halliwell 2004). Dispersal simulations for the Tanzanian and Iranian sites were not explored, due to the considerable distances between

locations, and preliminary examination of circulation patterns that predicted a lack of particle admixture.

Within the Pacific basin, *P. margaritifera* is known to have two reproductive events per year, with peaks and duration of spawning events varying by location. In the Indian Ocean, spawning appears to be restricted to a single season (Saucedo & Southgate 2008). A summary of the number and duration of spawning seasons for each sampling location was compiled from literature, to replicate larval supply over the year (see **Appendix 4.1**). At each seed location, 25,600 particles (see Lal *et al.* 2016a) were released per day for 14 days corresponding to the two major documented spawning peaks, and the model run forward in time for 90 and 60 days for each period respectively, within a single calendar year. Simulations were run separately for each of the two spawning periods using HYCOM data for 2015 and 2014, which were selected as these corresponded to an El Niño Southern Oscillation event (ENSO; Song *et al.* 2015; Varotsos *et al.* 2016). This permitted evaluation of any changes in dispersal patterns due to ENSO events over the 2014-2015 time scale.

Particle positions were extracted at time intervals of 60 and 90 days post-seeding for each simulation, and particle displacement visualised using the Generic Mapping Tools package (Wessel *et al.* 2013). Explicit, quantitative correlation of the genetic and hydrodynamic analyses was not possible, as this would have required genetic analysis of oysters at all potential source and sink locations with dense sampling coverage, and modelling of substantially more complex particle competency behaviour than computational resources permitted. Instead, an independent approach was adopted here, to examine congruency of results produced by the two analyses. No mortality or competency behaviour of the particles was simulated.

4.3 Results

4.3.1 SNP filtering

The raw dataset contained a total of 19,666 SNPs genotyped across all 580 individuals, at call rates ranging from 20-100%. The first filtering step undertaken to remove duplicate (clone) SNPs at genomic loci resulted in the removal of 8,079 SNPs (41% loss), after which the dataset was filtered for call rate (65%), average PIC (1%), MAF (2%) and average repeatability (95%). A total of 7 loci were found to deviate from HWE ($p<0.009$), and 99 loci were monomorphic across all 14 populations, which were subsequently removed together with 107 loci under significant LD ($p<0.0001$). These steps collectively resulted in the retention of 10,683 SNPs. Testing of this filtered dataset for F_{st} outlier loci detected 1,059 SNPs determined to be putatively under balancing and directional selection (Bayescan 2.1 results at FDR=0.01; **Appendix 4.3**), and their removal generated a final neutral dataset of 9,624 SNPs. This dataset was used for performing all population genomic analyses, while the original filtered dataset (10,683 SNPs) was retained for investigating adaptive variation.

4.3.2 Population genomic diversity and differentiation

Patterns observed in the mean numbers of alleles per locus (*A*) and rare allelic richness (*Ar*, <5% MAF) were similar, and appeared to vary by ocean basin (**Table 4.1**). Values of *A* for Pacific Ocean populations ranged from 1.6256 (Cook Islands) to 1.8067 (Indonesia), whereas Indian Ocean populations produced values of 1.3934-1.5649 (Tanzania, Mtwara to Abrolhos Islands, Australia). Trends in the total numbers of private alleles (A_p) reflected the divergence between ocean basins and support very limited inter-basin gene flow, with more than 25% of the total SNPs genotyped containing private alleles within each basin; (2,672 and 2,508 for Indian and Pacific Oceans respectively). Within ocean basins, little difference (~2% of total SNPs) was seen among Pacific populations (A_p range of 188-205), while greater differences (~3-5% total SNPs) were observed among the Abrolhos Islands, both Tanzanian, and Iranian sites (290, 354 and 458 respectively).

Table 4.1 Genetic diversity indices for the *P. margaritifera* populations sampled. The parameters calculated include the effective population size by the linkage disequilibrium method (N_{eLD}; 95 % confidence intervals indicated within brackets), mean number of alleles per locus (*A*), rare allelic richness at (*Ar*, MAF <5%), observed heterozygosity (H_o), average expected heterozygosity corrected for population sample size ($H_{n.b.}$), inbreeding coefficient (F_{is}), average individual multi-locus heterozygosity (Av. MLH), homozygosity by locus (HL), standardised heterozygosity (SH) and internal relatedness (IR).

Population	n	Source	N_{eLD} [95 % C.I.]	*A*	*Ar* (<5 %)	H_o (± SD)	$H_{n.b.}$ (± SD)	F_{is} (p<0.01)	Av. MLH (± SD)	HL (± SD)	SH (± SD)	IR (± SD)
Australia Abrolhos Island	50	Farm (hatchery)	9.3 [9.3 - 9.4]	1.5649	0.5446	0.0748 (±0.1244)	0.1655 (±0.1924)	0.5542	0.0914 (±0.0115)	0.8592 (±0.0174)	1.0682 (±0.1457)	0.5765 (±0.0516)
Australia GBR	35	Wild	∞ [∞ - ∞]	1.7603	0.3822	0.0762 (±0.0995)	0.2005 (±0.1771)	0.6265	0.0877 (±0.0044)	0.8618 (±0.0073)	1.0189 (±0.0567)	0.5737 (±0.0222)
Cook Islands	45	Farm (wild origin)	1684.7 [1475.1 - 1963.3]	1.6256	0.4984	0.0728 (±0.1092)	0.1722 (±0.1854)	0.5830	0.0868 (±0.0114)	0.8655 (±0.0179)	1.0066 (±0.1398)	0.5888 (±0.0523)
Fiji Islands	61	Farm (wild origin)	232.4 [229.9 - 234.9]	1.7934	0.3895	0.0929 (±0.1151)	0.1991 (±0.1758)	0.5372	0.1030 (±0.0306)	0.8370 (±0.0475)	1.2189 (±0.3905)	0.5050 (±0.1366)
French Polynesia	50	Farm	299.5 [293.4 - 305.9]	1.7208	0.4334	0.0718 (±0.1002)	0.1883 (±0.1814)	0.6236	0.0844 (±0.0091)	0.8687 (±0.0145)	0.9777 (±0.1132)	0.5965 (±0.0416)
Indonesia	48	Wild	1036.3 [972.6 - 1108.9]	1.8067	0.3635	0.0806 (±0.1027)	0.2054 (±0.1739)	0.6121	0.0925 (±0.0137)	0.8543 (±0.0215)	1.0816 (±0.1730)	0.5568 (±0.0633)
Iran	49	Wild	767.8 [693.1 - 860.3]	1.4402	0.7757	0.0371 (±0.0858)	0.1187 (±0.1795)	0.7008	0.0520 (±0.0039)	0.9378 (±0.0056)	0.5830 (±0.0445)	0.8127 (±0.0145)
Papua New Guinea	38	Wild	199.9 [196.3 - 203.8]	1.7632	0.3774	0.0732 (±0.0967)	0.2007 (±0.1769)	0.6410	0.0847 (±0.0034)	0.8661 (±0.0061)	0.9800 (±0.0399)	0.5884 (±0.0152)
Solomon Islands	50	Wild	119.8 [118.9 - 120.8]	1.8001	0.3748	0.0859 (±0.1077)	0.2019 (±0.1739)	0.5790	0.0965 (±0.0211)	0.8471 (±0.0336)	1.1374 (±0.2709)	0.5323 (±0.0964)
Taiwan	24	Wild	∞ [∞ - ∞]	1.7098	0.3830	0.0741 (±0.1035)	0.2021 (±0.1830)	0.6433	0.0859 (±0.0050)	0.8643 (±0.0080)	0.9947 (±0.0648)	0.5864 (±0.0230)
Tanzania (Mafia Island)	35	Wild	∞ [∞ - ∞]	1.4462	0.6369	0.0410 (±0.0878)	0.1296 (±0.1840)	0.6964	0.0553 (±0.0060)	0.9290 (±0.0083)	0.6149 (±0.0715)	0.7910 (±0.0215)
Tanzania (Mtwara)	20	Wild	∞ [∞ - ∞]	1.3934	0.6485	0.0427 (±0.0951)	0.1256 (±0.1871)	0.6795	0.0557 (±0.0083)	0.9285 (±0.0108)	0.6206 (±0.0994)	0.7898 (±0.0299)
Tonga	28	Wild	120.8 [118.7 - 122.8]	1.6995	0.4062	0.0775 (±0.1076)	0.1954 (±0.1828)	0.6119	0.0889 (±0.0104)	0.8594 (±0.0171)	1.0404 (±0.1326)	0.5714 (±0.0493)
Vietnam	47	Wild	681.5 [651.7 - 714.2]	1.8016	0.3587	0.0775 (±0.0994)	0.2060 (±0.1737)	0.6281	0.0892 (±0.0135)	0.8592 (±0.0215)	1.0378 (±0.1706)	0.5723 (±0.0616)

Average observed heterozygosities were significantly lower ($p<0.05$) than average expected heterozygosities for all populations, and displayed similar variability with the trends observed for *A* and *Ar* values. Pacific Ocean populations displayed generally higher values (H_o: 0.0718-0.0929; $H_{n.b.}$: 0.1722-0.2060), than did Indian Ocean populations (H_o: 0.0371-0.0748; $H_{n.b.}$: 0.1187-0.1655). These patterns also extended to individual average multi-locus heterozygosity (MLH) computations, and measurements of standardised heterozygosity (SH). Average MLH was relatively uniform within Pacific Ocean populations, ranging from 0.0844 (French Polynesia) to 0.1030 (Fiji Islands), which was mirrored in the SH results of 0.9777-1.2189 for the same populations respectively. Within Indian Ocean samples, oysters collected from Tanzanian and Iranian sites showed lower values (MLH: 0.0520-0.0557; SH: 0.5830-0.6206), than animals sampled from the Abrolhos Islands (MLH=0.0914; SH=1.0682).

Inbreeding coefficient (F_{is}) values displayed a similar partitioning by region, with values for Pacific Ocean populations ranging from 0.5372 (Fiji Islands) to 0.6433 (Taiwan), while Indian Ocean animals (with the exception of Abrolhos Islands oysters; F_{is}=0.5542), returned higher values from 0.6795 (Tanzania, Mtwara) to 0.7008 (Iran). Very similar patterns were reported for Fijian oysters in Chapters 2 and 3 (with possible causes discussed), and also evident in related homozygosity by locus (HL) and internal relatedness (IR) multi-locus metrics (see **Table 4.1**). Estimates of effective population size were robust, however, they varied considerably across all sampling locations. Several populations returned infinite N_{eLD} values, including oysters sampled from the GBR, Taiwan and the two Tanzanian locations. Estimates from Solomon Islands samples were at the low end of the range (119.8; [95% CI=118.9-120.8]), while Cook Islands individuals produced higher values (1,684.7; [95% CI=1,475.1-1,963.3]). The lowest estimates were obtained from Abrolhos Islands oysters (9.3; [95% CI=9.3-9.4]), indicating a possible bottleneck, as these animals were F_1 hatchery-produced offspring of wild-caught parents.

4.3.3 Resolution of population structure and migration

Pairwise F_{st} estimates (**Table 4.2**) were highly significant ($p<0.001$) for all population comparisons, with the exception of the two Tanzanian sites (0.0007), and PNG with the Solomon Islands (0.0059). A clear separation in population structure between ocean basins is evident, with pairwise estimates between sites all >0.25, ranging from Tanzania, Mtwara and Indonesia (0.2894), to Iran and the Cook Islands (0.4684).

Table 4.2 Population differentiation estimates for 14 *P. margaritifera* populations sampled. Population pairwise F_{st} estimates computed in Arlequin v.3.5.1.3. are shown below the diagonal, while Nei's (1978) standard genetic distances (D_S) computed in Genetix v.4.05 with 10,000 permutations are reported above. All F_{st} values were significant at p<0.001 following 1,000 permutations. Non-significant F_{st} and D_S values (p>0.05) are presented in bold type.

	Australia Abrolhos Islands	Australia GBR	Cook Islands	Fiji Islands	French Polynesia	Indonesia	Iran	Papua New Guinea	Solomon Islands	Taiwan	Tanzania (Mafia Island)	Tanzania (Mtwara)	Tonga	Vietnam
Australia Abrolhos Islands		0.056	0.095	0.068	0.082	0.053	0.264	0.056	0.057	0.053	0.236	0.238	0.069	0.051
Australia GBR	0.1311		0.033	0.009	0.021	0.009	0.256	**0.005**	0.005	0.009	0.234	0.236	0.011	0.008
Cook Islands	0.2173	0.0816		0.023	0.020	0.044	0.306	0.035	0.035	0.044	0.289	0.291	0.027	0.043
Fiji Islands	0.1526	0.0194	0.0537		0.011	0.018	0.273	0.011	0.010	0.017	0.253	0.255	0.006	0.016
French Polynesia	0.1892	0.0490	0.0502	0.0221		0.031	0.292	0.022	0.021	0.030	0.275	0.277	0.014	0.028
Indonesia	0.1209	0.0211	0.1084	0.0459	0.0759		0.243	0.008	0.009	**0.006**	0.217	0.219	0.020	0.006
Iran	0.4177	0.4145	0.4684	0.3903	0.4438	0.3711		0.255	0.257	0.241	0.071	0.074	0.276	0.239
Papua New Guinea	0.1297	0.0079	0.0862	0.0227	0.0532	0.0166	0.40860		0.005	**0.008**	0.232	0.233	0.013	0.007
Solomon Islands	0.1297	0.0071	0.0835	0.0248	0.0499	0.0208	0.38462	**0.0056**		0.008	0.234	0.237	0.012	0.008
Taiwan	0.1196	0.0172	0.1090	0.0375	0.0739	0.0100	0.41411	0.0148	0.0128		0.217	0.218	0.020	**0.006**
Tanzania (Mafia Island)	0.3508	0.3494	0.4185	0.3374	0.3951	0.3038	0.24438	0.3444	0.3236	0.3394		0.013	0.257	0.214
Tanzania (Mtwara)	0.3402	0.3323	0.4043	0.3235	0.3804	0.2894	0.25340	0.3280	0.3084	0.3200	**0.0069**		0.259	0.216
Tonga	0.1607	0.0235	0.0628	0.0099	0.0294	0.0494	0.44413	0.0279	0.0254	0.0443	0.3846	0.3644		0.019
Vietnam	0.1128	0.0185	0.1062	0.0407	0.0732	0.0124	0.37119	0.0161	0.0174	0.0088	0.3043	0.2906	0.0469	

Within the Pacific, populations appear to be isolated by geographic separation, e.g. pairwise estimates for the GBR and Solomon Islands (0.0078) indicate greater homogeneity than more distant population pairs, such as the Cook Islands and Taiwan (0.1090). Higher degrees of separation are apparent within Indian Ocean populations, with pairwise estimates between Iran and both Tanzanian sites (Mafia Island and Mtwara) being 0.2444 and 0.2534 respectively. The greatest level of differentiation among Indian Ocean sites was detected between the Abrolhos Islands and Iran (0.4177), with oysters from the Abrolhos Islands demonstrating greater similarity with Pacific populations (Abrolhos Islands and GBR pairwise F_{st}=0.1311).

Pairwise Nei's standard genetic distances (D_S) described a similar pattern to the pairwise F_{st} estimates (**Table 4.2**), with the Iranian and two Tanzanian populations displaying marked separation from all other populations (0.214-0.306; $p<0.05$). Partitioning between these populations however, was less evident, with D_S=0.071 and 0.074 respectively (Iran with Mafia Islands and Mtwara). Distances between all Pacific Ocean populations conversely indicated greater homogeneity, ranging from 0.005 (PNG, GBR and Solomon Islands pairwise comparisons), to 0.044 (Cook Islands with Indonesia and Taiwan pairwise comparisons). Oysters collected from the Abrolhos Islands were similarly differentiated, with D_S=0.056 when compared to GBR individuals, and up to D_S=0.082 with French Polynesian animals.

Results of the hierarchical AMOVA carried out between Indian vs. Pacific Ocean basins and populations indicated that 18.11% of the variance originated between ocean basins, with the greatest proportions of variance attributed to within-sample variation (45.79%), and between samples within populations (35.74%). Variation between populations within ocean basins was estimated at just 0.36%, indicating that genotypic variability at the individual oyster level accounted for the majority of the observed variation. Mantel tests indicated isolation by distance dispersal patterns both within each ocean basin (R^2=0.939, p=0.041 and R^2=0.464, p=0.000 for Indian and Pacific oceans respectively), as well as for all populations considered together (R^2=0.613, p=0.000), although additional sampling within each region is needed to confirm the strength of these results. Further Mantel tests within the two largest Pacific Ocean population groupings identified by DAPC and Netview P analyses did not detect significant IBD patterns ($p>0.05$).

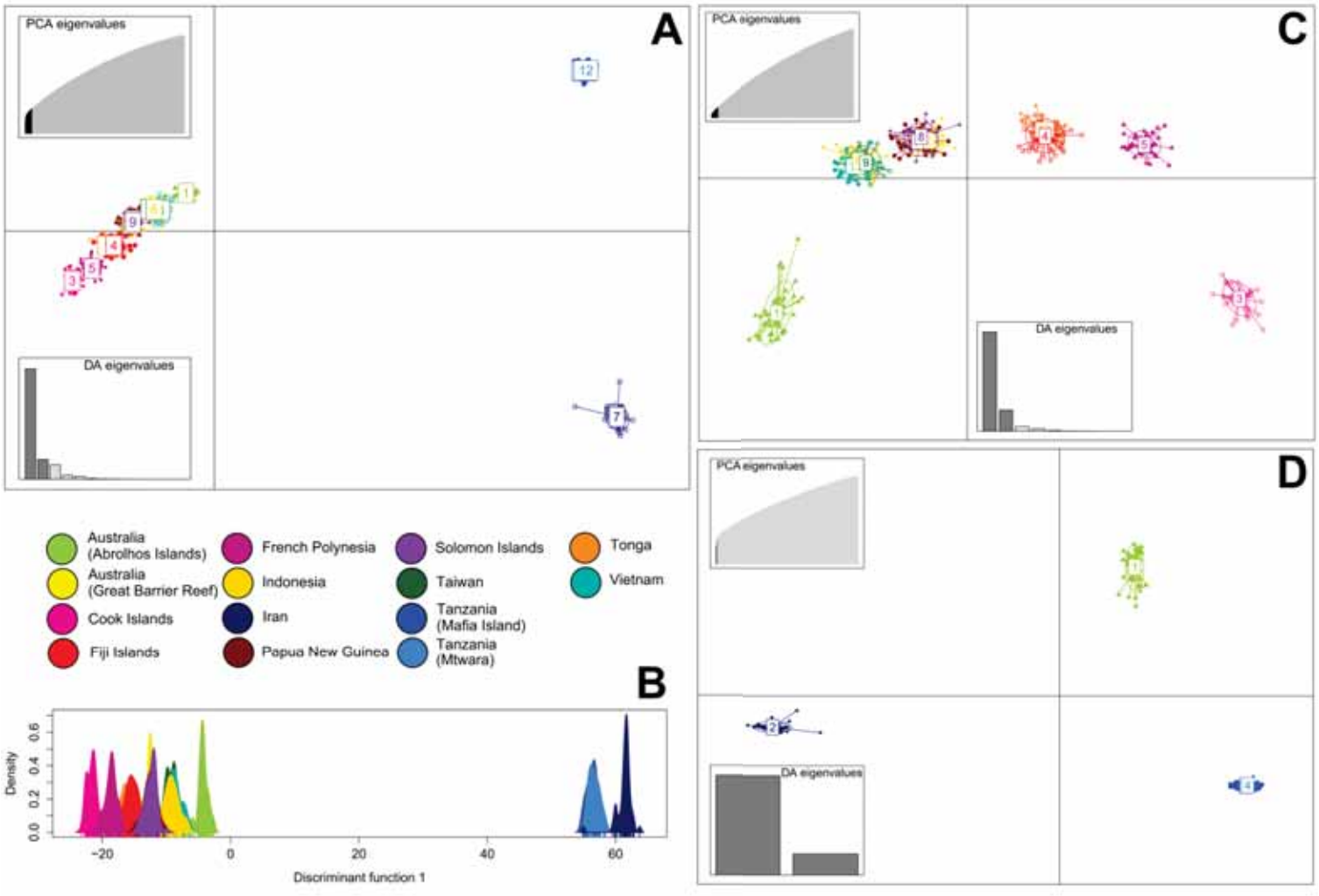


Figure 4.2. Discriminant Analyses of Principal Components (DAPC) carried out using the *R* package *adegenet* to illustrate broad-scale patterns of population structure. Dots on scatterplots represent individuals, with colours denoting sampling origin and inclusion of 95 % inertia ellipses. Scatterplot (**A**) was constructed among all 580 individuals collected from both the Pacific and Indian Ocean sites, while (**B**) is an individual density plot on the first discriminant function for this dataset. Scatterplots (**C**) and (**D**) were constructed on individuals sampled from Pacific Ocean (**C**) and Indian Ocean (**D**) sites only, to clearly identify regional differentiation.

Visualisation of population structure with a DAPC (α-score optimised to retain 22 PCs), revealed clear differentiation between all Pacific Ocean, and both Tanzanian and Iranian populations (**Figures 4.2a** and **4.2b**), when all individuals were analysed together. Additional DAPC analyses involving separation of populations into their respective ocean basins further clarified the patterns observed. Analysis of all populations from the Pacific Ocean (**Figure 4.2c**) revealed clear partitioning of the French Polynesian and Cook Islands oysters from all other populations, while animals sampled from Fiji and Tonga formed a single cluster. Similarly, individuals collected from PNG, Solomon Islands and the GBR formed a single cohesive group, as did oysters sampled from Indonesia, Taiwan and Vietnam. This pattern of separation was confirmed by testing for the actual number of discrete clusters using the BIC method, which was determined to be k=8.

Examination of fine-scale population structure using Netview P (**Figures 4.3a** and **4.3b**) resolved similar patterns of differentiation to the DAPC, but offered greater resolution at the individual oyster level between several population pairs. In particular, when an organic network topology was used (k-NN=40; **Figure 4.3a**), it highlighted the degree of connectivity between the two broad clusters comprising oysters collected from Indonesia, Vietnam and Taiwan, along with individuals sampled from the GBR, Solomon Islands and PNG respectively. Analysis using a circular network topology (k-NN=10; **Figure 4.3b**) made this especially clear, as all individuals from these six locations collapsed into a single cluster. Interestingly, oysters collected from the Abrolhos Islands split into two sub-clusters (**Figure 4.3b**), potentially indicating the presence of family groups, given that all individuals were sampled as a hatchery-produced cohort. Similarly, a closer relationship was apparent between French Polynesian, and Fijian-Tongan samples than with Cook Islands individuals, despite the greater geographic distance separating these populations. This may be due to prevailing ocean current patterns, which could ensure greater connectivity through directional larval dispersal. Networks constructed at lower and higher k-NN thresholds all showed identical differentiation patterns.

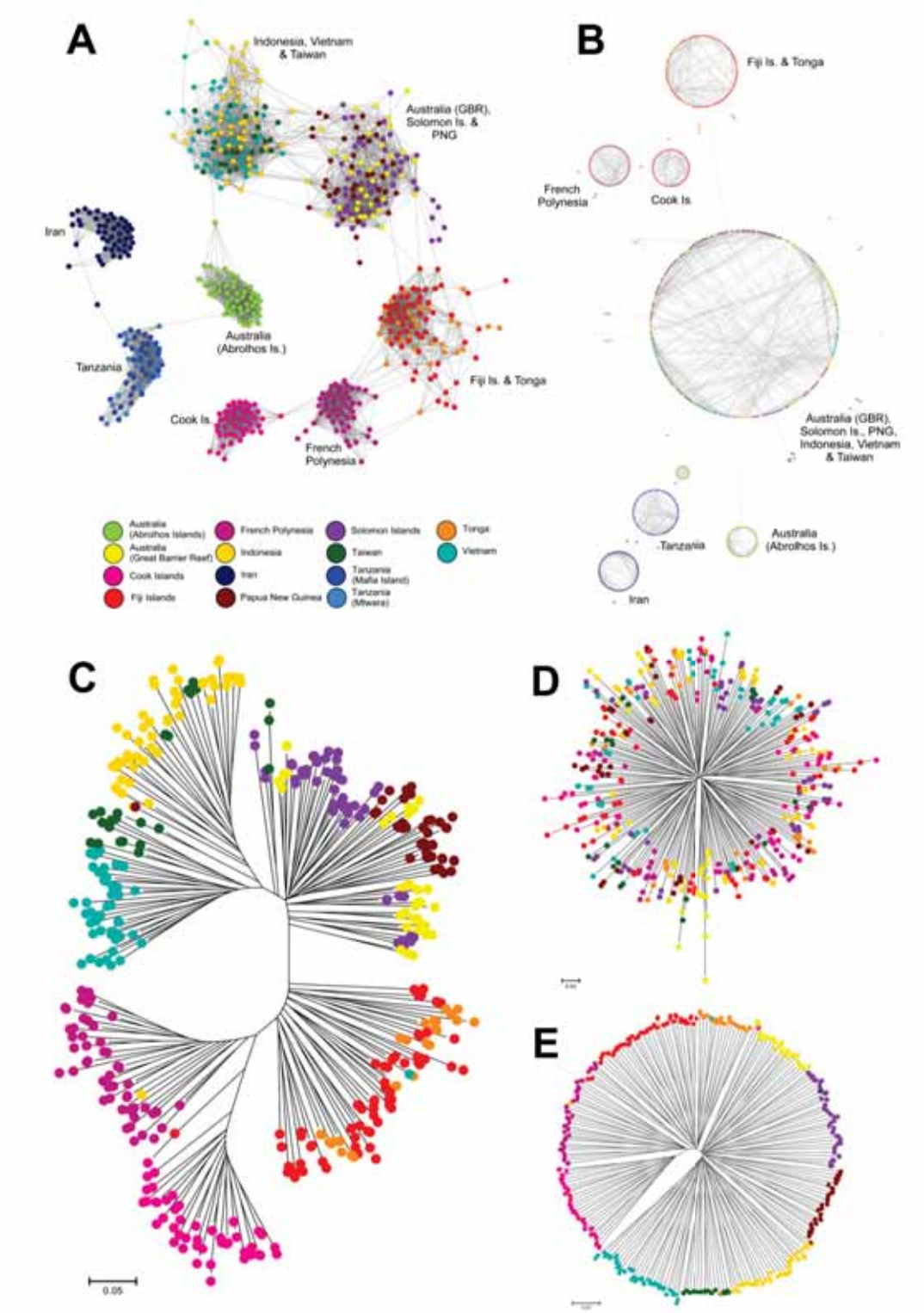


Figure 4.3 Visualisation of population structure among 580 *P. margaritifera* individuals sampled. Fine-scale population networks constructed using the Netview P v.0.4.2.5 pipeline and selectively-neutral loci are shown in (**A**) organic; k-NN=40 and (**B**) circular; k-NN=10 topologies, with each dot representing a single individual. Oysters sampled from the Pacific Ocean had sufficiently low neutral F_{st} levels to permit testing for outlier loci, and Neighbour-Joining trees generated based on 1-psa distance matrices for these individuals are shown in (**C**) and (**D**). The tree displayed in (**C**) was drawn using 89 putatively directional outlier loci detected by both Bayescan 2.1 and LOSITAN at an FDR=0.05, while (**D**) was generated using 37 also jointly-identified putatively balancing loci, at an FDR=0.05. (**E**) shows the arrangement of population structure in these same individuals, but with all loci (9,624 SNPs). The scale bars for (**C**), (**D**) and (**E**) indicate 1-psa genetic distance.

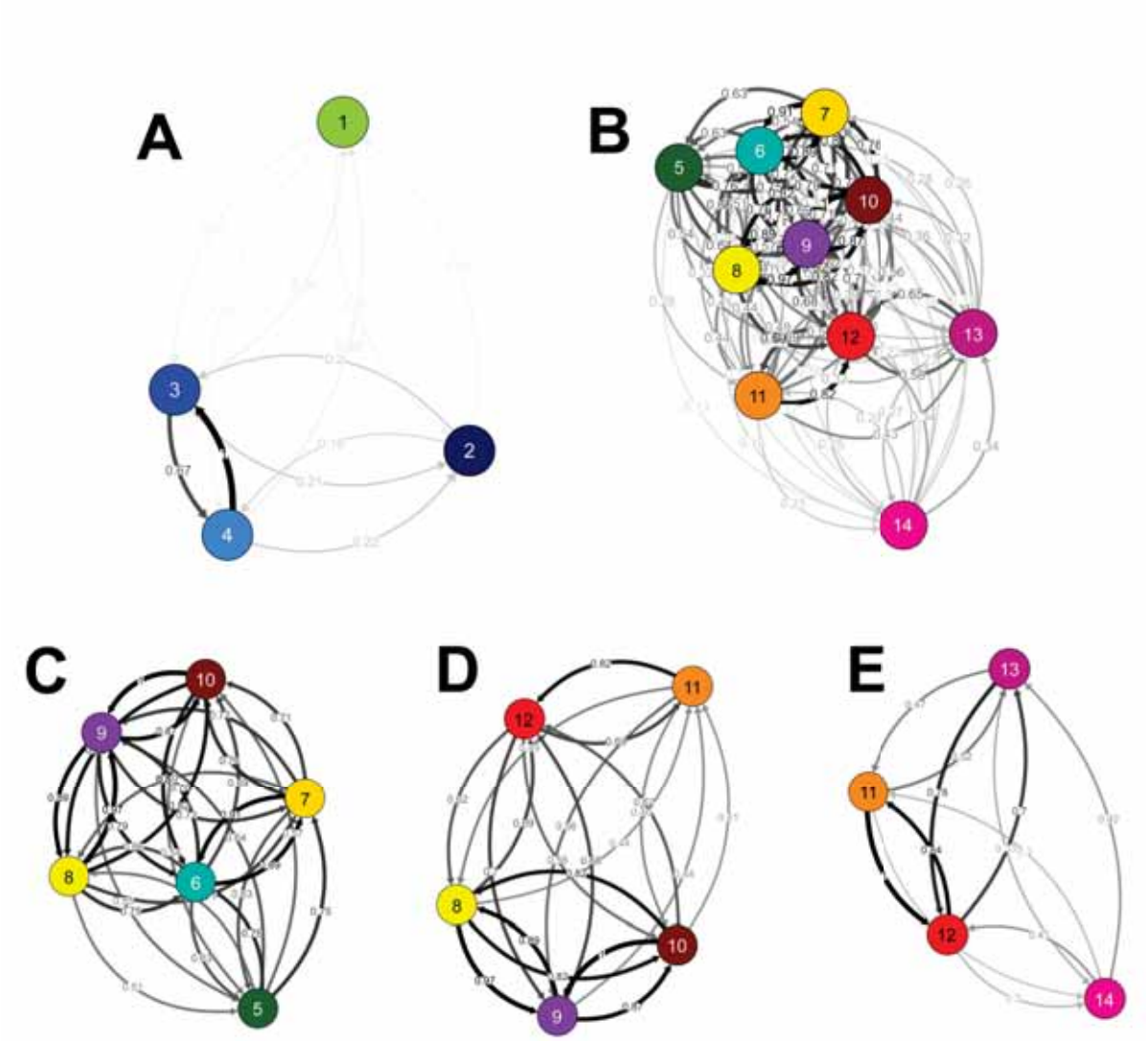


Figure 4.4 Migration networks for *P. margaritifera* populations generated using the *divMigrate* function in *diveRsity* (Keenan *et al.* 2013). Circles represent populations, while arrows indicate the direction and magnitude (arrow edge values) of relative migration levels using Nei's G_{st} method (Rousset 2008; Sundqvist *et al.* 2013). Darker arrows indicate stronger migration relationships compared to lighter arrows. Separate networks are shown for all Indian Ocean populations (**A**) and all Pacific Ocean populations (**B**) sampled. To better visualise separation between all Pacific Ocean populations, further networks have been generated for population groups located in the western Pacific (**C**), western and central Pacific (**D**) and the central and eastern Pacific (**E**). All networks were generated following 1,000 bootstraps and all pairwise relationships are significant ($p<0.01$). Population colour codes correspond to **Figures 4.1-4.3,** and have been numbered as follows. **1:** Australia (Abrolhos Is.), **2**: Iran; **3**: Tanzania (Mafia Is.), **4**: Tanzania (Mtwara), **5**: Taiwan, **6**: Vietnam, **7**: Indonesia, **8**: Australia (GBR), **9**: Solomon Is., **10**: Papua New Guinea, **11**: Tonga, **12**: Fiji Is., **13**: French Polynesia and **14**: Cook Is.

Assessment of migration patterns and gene flow (**Figure 4.4**) using *divMigrate* networks demonstrated nearly identical patterns of population structure between Indian (**Figure 4.4a**) and Pacific (**Figure 4.4b**) Ocean basins, when compared to the DAPC and Netview P networks. These similarities extended to closer examinations of Pacific Ocean populations by sub-region (**Figures 4.4c-e**). Among Indian Ocean populations, directional migration

between both Tanzanian sites was the strongest, but with very little connectivity between these two locations, Iran and the Abrolhos Islands. Connectivity within the Pacific region however, was substantially higher, with only the Cook Islands and French Polynesian populations remaining relatively isolated (**Figures 4.4b**, **4.4e**). Directional migration between Western Pacific sites (Vietnam, Indonesia, Taiwan, PNG, Solomon Islands and GBR) was found to be the strongest (**Figures 4.4c**, **4.4e**), followed by connectivity between the Fiji Islands and Tonga (**Figures 4.4d**, **4.4e**). Despite the geographic proximity of the Cook Islands to the Fiji Islands and Tonga, migration between both these locations and French Polynesia which is more distant was considerably higher.

4.3.4 Examination of adaptive variation

F_{st} outlier tests discovered between 45-137 putatively directional, and 37-216 putatively balancing outlier loci jointly-identified by Bayescan 2.1 and LOSITAN, at six FDR thresholds for Pacific Ocean populations (**Appendix 4.2**). Both platforms failed to detect loci under balancing selection below an FDR=0.01, and based on verification of loci detected at all FDR thresholds using QQ plots, a final stringent FDR threshold of 0.05 was selected. At this FDR, 89 directional and 37 balancing loci were jointly-identified, and used to construct NJ trees to visualise population structure at loci putatively under selection (**Figures 4.3c**, **4.3d** and **4.3e**).

Weak population structure observed at selectively neutral and balancing loci (**Figures 4.3e** and **4.3d** respectively), correlated well with pairwise F_{st} and D_S comparisons. At directional loci however, clear divergence was evident between populations, which corresponded exactly with the five clusters identified by DAPC and Netview P networks in the Pacific Ocean. To gauge the strength of the selection signal, average Bayescan 2.1 F_{st} values among the 89 directional loci were examined, and found to equal 0.1915 (range=0.1012 to 0.4371). Among the 37 balancing loci, average F_{st}=-0.0066 (range=-0.0114 to -0.0031), demonstrating that diffuse population structure (NJ trees **Figures 4.3e** and **4.3d**), becomes apparent when considering these and selectively neutral loci. These results indicate the likely presence of local adaptation acting on the populations examined, which is likely due to the heterogenous habitats occupied by *P. margaritifera* across the Pacific Ocean.

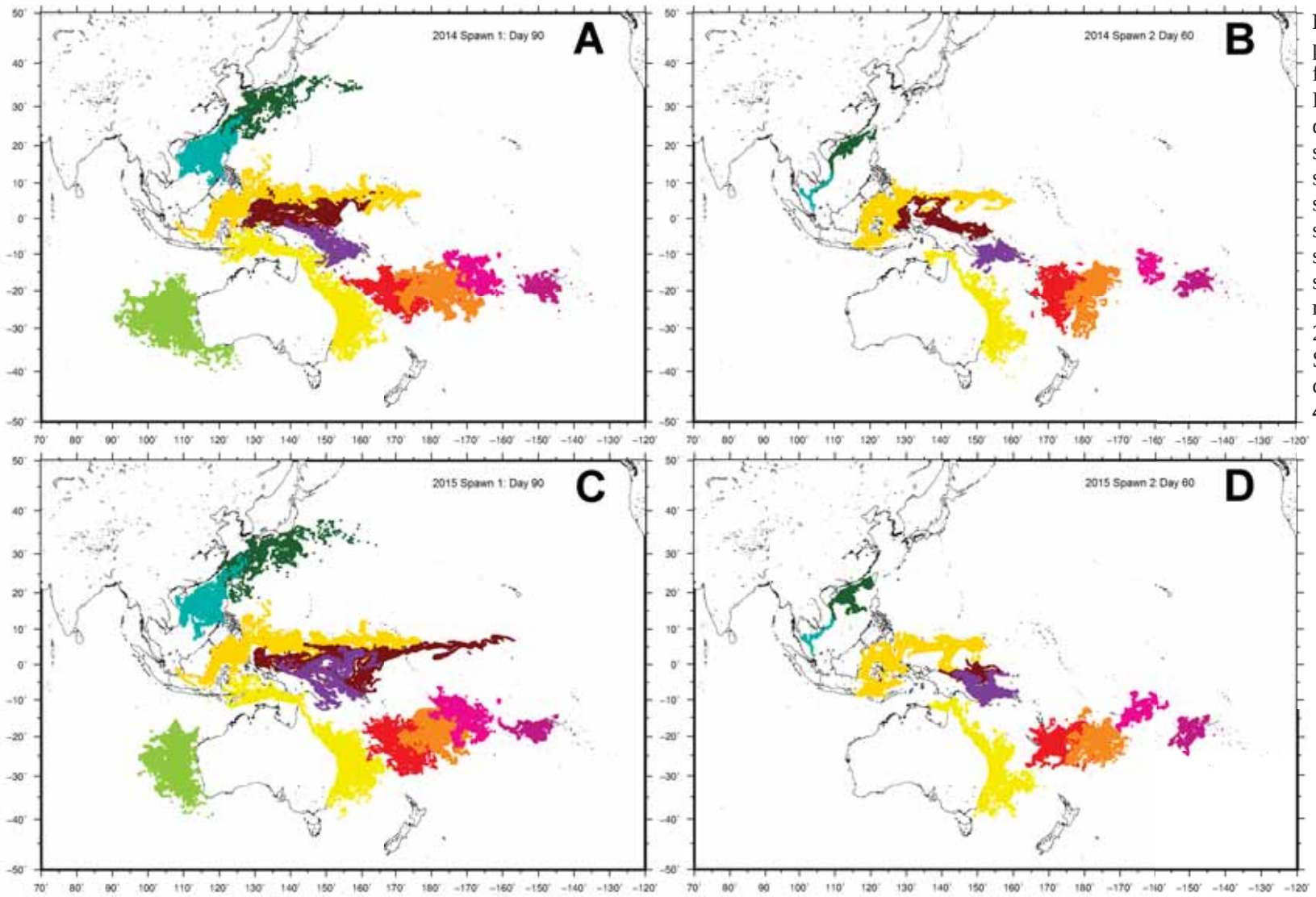


Figure 4.5. Results of particle dispersal simulation for 11 sampling sites. Particle positions are displayed for the following simulations: spawning season 1 for 2014 (**A**), season 2 for 2014 (**B**), season 1 for 2015 (**C**) and season 2 for 2015 (**D**). All season 1 simulations were run for 90 days, and season 2 simulations over 60 days. Sampling site colour codes correspond with **Figures 4.1-4.4**.

4.3.5 Particle dispersal modelling

Simulations of larval transport revealed a high degree of admixture by surface ocean currents within the Pacific basin over both 2014 and 2015 datasets, (**Figure 4.5** and see **additional files 2a-d** for animations of the full dispersal simulations). Interestingly, differences in the direction and extent of dispersal were observed between spawning seasons within either year, than between peak ENSO activity (2014 recorded an El Niño event, which dissipated in 2015). In particular, particles originating in both Taiwan and Vietnam were advected north towards Japan and the Ogasawara Islands archipelago during the first spawning seasons of both 2014 and 2015 (**additional files 2a**, **2c**), while these current patterns reversed during the second spawning seasons, directing particles south across the Vietnamese coastline towards Malaysia (**additional files 2b**, **2d**).

Overall patterns of population structure inferred from DAPC, Netview P and *divMigrate* analyses were highly concordant with simulated dispersal patterns for both ocean basins. At a broad scale, connectivity between the GBR, Solomon Islands, PNG, Indonesia, Vietnam and Taiwan was particularly obvious, together with the Fiji Islands and Tonga. Dispersal patterns for Indian Ocean sampling sites was limited to the Abrolhos Islands, where larval output is likely to spread northwards over much of the Western Australian seaboard (**Figures 4.5a** and **4.5c**). While providing unprecedented insights into the larval connectivity of *P. margaritifera*, these results should not be interpreted as reflecting actual recruitment over the limits of final particle positions. For example, because larval competency behaviour was not modelled, particles originating from the GBR transported into the south Tasman Sea are unlikely to survive due to unfavourable water temperatures in that region.

4.4 Discussion

This study examined population genetic structure and connectivity in the black-lip pearl oyster, over its ~18,000 Km natural distribution. Assessments of differentiation at both neutral and adaptive markers, together with an independent particle dispersal simulation indicate that the evolutionary and physical processes organising population genetic structure are highly complex. At broad and regional scales, surface ocean currents, geographic distance and habitat geomorphology play important roles in regulating connectivity. At sub-regional and local scales, seascape features such as coral atolls, shoals and straits may impede gene flow, and the presence of environmental heterogeneity result in adaptive differences between populations.

In the Pacific Ocean, observations do not lend support for a strong CPH model, where *P. margaritifera* is expected to exhibit reduced diversity and increased differentiation towards its range limits. However, this does not imply that CPH trends are absent, as very high levels of gene flow may conceal *C-M* gradients and sampling may not have detected the true range limits. The presence of local adaptation in habitat sub-regions also supports the presence of hetereogenous environments. Conversely in the Indian Ocean, clear divergence between the marginal populations sampled suggests the presence of *C-M* clines cannot be discounted, and requires further investigation at higher sampling densities, with particular attention to central populations. It is apparent that the mechanisms underlying range-wide genetic structure in *P. margaritifera* are quite complex, and require closer examination to better understand the evolutionary, ecological and physical factors at work.

4.4.1 Basin-wide population structure and connectivity

At a broad scale, *P. margaritifera* populations in the Indian and Pacific Oceans displayed substantial and significant divergence (pairwise F_{st} estimates = 0.2894-0.4684, $p<0.001$). Strong population structure was evident within and between both ocean basins, however, due to the relative isolation of populations between these regions, each is discussed separately.

4.4.1.1 Pacific Ocean

Gene flow among Pacific Ocean populations appears to occur at a basin-wide scale, with pairwise F_{st} estimates reaching a maximum of 0.1090 (Cook Islands and Taiwan), over a distance of approximately 9,900 km. Despite the high degree of admixture among populations, visualisation of population structure (**Figures 4.2**, **4.3**, and **4.4**) resolved five distinct genetic groups. When dispersal simulation data (**Figure 4.5** and **additional files 2a-d**) are compared to genetic differentiation patterns, the physical limits of simulated larval dispersal closely match population groupings. This observation suggests that while surface ocean currents permit sufficient gene flow across the Pacific Ocean to ensure populations retain a high degree of connectivity, circulation patterns and IBD may also facilitate regional larval retention, that stabilises population genetic structure. Because even low levels of gene flow (Slatkin 1987, 1993) are able to prevent population divergence, it is conceivable that standing genetic diversity and structure are maintained by a "founder takes all" density-dependent effect (Waters *et al.* 2013), where individuals arriving after an initial colonisation event may be "blocked" by established conspecifics (Liggins *et al.* 2014; Waters *et al.* 2013).

For the present study, at the geographical limits of the species distribution in the Pacific, decreased differentiation between Taiwan and French Polynesia (F_{st}=0.0739) is evident despite the considerable distance involved (~11,000 km). This observation does not support generalised CPH predictions, and is likely a result of greater connectivity of this population pair through ocean current circulation (Ganachaud *et al.* 2007; Liggins *et al.* 2015). This is corroborated by dispersal simulation data (**additional files 2a** and **2c**), and supported by pairwise migration analyses (**Figure 4**). Larval competency following an extended pelagic dispersal phase is also expected to play a role in recruitment success or failure, as individuals may have greater fitness as a result of shorter and potentially less stressful larval development (Nosil *et al.* 2005; Shima & Swearer 2010). Here, ocean currents may impact recruitment rates by permitting increased larval fitness through reduced transport times, meaning that a population pair separated by greater physical distance may share higher connectivity, compared to a neighbouring population pair where larval plumes are vectored in mutually opposite directions or via circuitous pathways (Shanks 2009; Simpson *et al.* 2014).

Another factor influencing population structure and connectivity is habitat geomorphology, which is particularly evident in the Western Pacific, where long-range larval dispersal is

restricted by the presence of numerous shoals, straits, islands, reefs and semi-enclosed seas (Dao *et al.* 2015). This is reflected in the segregation of Taiwanese, Vietnamese and Indonesian individuals, from oysters collected in PNG, the Solomon Islands and the GBR (**Figures 4.2**, **4.3** and **4.4**). Similar patterns have been documented in several highly-dispersive marine taxa, ranging from a diatom (Godhe *et al.* 2013) and limpet (Hoffman *et al.* 2011), to giant clam (Kochzius & Nuryanto 2008) and mullet (Krück *et al.* 2013).

4.4.1.2 Signatures of selection in the Pacific basin

Similarities in the patterns of population structure obtained at loci under directional selection (**Figures 4.3c-e**), to spatial arrangements generated by DAPC and Netview P networks at selectively neutral loci (**Figures 4.2; 4.3a-b**), reinforce stock boundaries identified for *P. margaritifera* in the Pacific basin. The seascape of the Pacific region has been shaped by complex geological processes, giving rise to considerable habitat heterogeneity (Carpenter 1998; Sanciangco *et al.* 2013). Given the large extent of the species distribution sampled (>11,000 km), it is feasible that the selective differences observed may originate from distinct habitat sub-regions present within the Pacific basin (Barber *et al.* 2002; Bruno & Selig 2007; Wood *et al.* 2014).

For range-wide investigations of genetic structure in broadcast spawning marine species, consideration of adaptive variation can be important for uncovering functional differences between populations that might otherwise go undetected. As an example, adaptive divergence in the Atlantic cod related to temperature and salinity clines across the species distribution was detected by Nielsen *et al.* (2009), but not evident within a restricted portion of its range (Knutsen *et al.* 2011), where environmental differences were predicted to be similar. Similarly, the work described in chapter 3 failed to detect signatures of selection between and among Fijian populations; however, results presented here indicate that detectable selection is evident only at the scale of Fijian and Tongan populations considered together.

In certain situations, adaptive differences in the face of high gene flow are the only discriminating factor through which concise fishery management is possible, by disentangling the effects of selection from demographic history, migration and genetic drift (Bourret *et al.* 2013; Gaggiotti *et al.* 2009; Wang *et al.* 2013). For example, Nayfa and

Zenger (2016) detected divergent selection between three Indonesian populations of the silver-lip pearl oyster *P. maxima* over ~2,000 km, where functional differences had manifested themselves in commercial fitness trait differences (namely growth rate and shell size Kvingedal *et al.* 2010). Because the complex life histories of marine taxa may result in greater vulnerability to pre- and post-settlement selective forces (Nosil *et al.* 2009; Nosil *et al.* 2005), the ability to detect these effects on the genetic composition of populations is critical for informing management for aquaculture, translocation, population supplementation and assisted migration (Funk *et al.* 2012; Hemmer-Hansen *et al.* 2007; Knutsen *et al.* 2011; Miller *et al.* 2012).

4.4.1.3 Indian Ocean

Populations sampled from the Indian Ocean displayed substantial vicariance, with the magnitude of separation between the three distinct genetic groups potentially indicating the presence of distinct ESUs, based on D_S estimates (**Table 4.2**; (Kalinowski 2002; Nei 1972, 1987)). Work by Ranjbar *et al.* (2016) and Cunha *et al.* (2011), suggest that *Pinctada margaritifera* may in fact be a species complex, with populations in the Persian Gulf comprising a distinct ESU. Restriction of gene flow into the Persian Gulf from the greater Indian Ocean by the Strait of Hormuz likely isolates these individuals, and while the current study provides an initial assessment of basin-wide population differentiation for Indian Ocean *P. margaritifera*, further hierarchical sampling is required to determine regional patterns of evolutionary and contemporary genetic structure.

Particular attention to core populations from the central Indian Ocean (Maldives), Madagascar, Arabian Sea, Bay of Bengal, Andaman Sea and Sumatra may resolve these questions, and potentially ascertain the presence of a genetic break between the Indian and Pacific Oceans. Pairwise F_{st} estimates and visualisation of genetic structure between the closest marginal populations from the Western Pacific in the current dataset suggest this is a possibility (see **Table 4.2** and **Figures 4.2**, **4.3**), as similar observations have been recorded for other invertebrate taxa (Barber *et al.* 2002; Richards *et al.* 2016; Williams *et al.* 2002; Wörheide *et al.* 2008).

4.4.2 Patterns across the species' distribution

The CPH predicts that genetic diversity and connectivity should be highest at the centre of a species' range and decrease towards the periphery, however, the data presented here indicate the presence of patterns which are substantially more complex than generalised CPH predictions. For Pacific populations, easily discernible *C-M* trends were absent, and may mean that the homogenising influence of basin-wide current circulation patterns disrupts any obvious patterns. However, ocean currents together with isolation by geographic distance are also likely to maintain sub-regional population structure (e.g. Miller *et al.* 2013 for the surf clam *Donax deltoides*).

Sample collection for the current study was organised according to the published theoretical distribution of *P. margaritifera* (Wada & Tëmkin 2008), and therefore it is possible that the true species distribution limit may not have been sampled, if it in fact extends beyond the current known range. If edge effects of decreased genetic diversity and marked differentiation are present, further sampling and analysis at the periphery of the species distribution in the Pacific Ocean may detect them. The levels of divergence between Indian Ocean oysters could reflect edge effects, considering that individuals were sampled from the ocean basin margins, however, as no central populations were able to be sampled, this observation cannot be substantiated. In addition to the CPH, other theoretical models for describing population organisation such as source-sink interactions, and range edge disequilibrium (Sexton *et al.* 2009) warrant consideration. This is because for many species, range margins are often mobile with expansions and contractions over time, and are the result of numerous biotic and abiotic mechanisms (Eckert *et al.* 2008; Guo 2012; Sexton *et al.* 2009).

4.4.3 Drivers of genetic structure and implications for fishery management

It is evident that the biological and physical processes governing population structure and genetic diversity in *P. margaritifera* are complex. In the Pacific Ocean, the data indicate that ocean currents, seascape features and geographic distances are possible major influences on population connectivity which both disrupts *C-M* clines, and simultaneously stabilises population structure according to basin sub-regions (Wood *et al.* 2014). Broad-scale habitat geomorphology also plays an important role in differentiating populations, by restricting gene flow and influencing sub-regional natural selection. While the sampling scope in the Indian

Ocean was insufficiently dense to determine the existence of *C-M* trends, ocean currents may play a large role in maintaining divergent populations. It is possible that a genetic break between the Indian and Pacific Oceans may exist at the south-east Asian archipelago, and further investigation of these populations could provide answers to this question, as it has for other marine invertebrates (Williams *et al.* 2002; Wörheide *et al.* 2008). Gauging the importance of oceanic circulation for driving population genetic structure and connectivity for *P. margaritifera* would not have been possible without simulations of larval dispersal, and it is evident that oceanographic and/or ecological modelling data is an indispensable component of range-wide investigations of genetic structure in marine organisms, which possess passively dispersing planktonic larvae (Liggins *et al.* 2016; Liggins *et al.* 2013).

Data presented here do not support *P. m.* var. *typica* and *P. m.* var. *cummingi* as valid sub-species classifications in the Pacific Ocean, given the level of broad-scale admixture detected and absence of evidence for distinct ESUs. Unfortunately, as Hawaiian populations could not be sampled, no conclusion as to the status of *P. m.* var. *galstoffi* may be drawn. However, given the ability of larvae to disperse across the Pacific basin over the span of several generations, it is possible that Hawaiian populations may not be as divergent as previously thought (Galstoff 1933). Conversely, *P. m.* var. *zanzibarensis* and *P. m.* var. *persica* in the Indian Ocean may constitute distinct ESUs, given their substantial divergence from all other populations, although denser basin-wide sampling is required for verification. A comprehensive range-wide phylogenetic analysis of *P. margaritifera* is also needed to assess how many ESUs may be present, and to determine if the black-lip pearl oyster represents a true species complex. Because there are discernible regional morphological differences within *P. margaritifera*, there may be parallels with the Akoya species complex, which also displays morphological variability, high levels of gene flow and has a similarly extensive Indo-Pacific distribution (Tëmkin 2010; Wada & Tëmkin 2008).

4.5 Conclusions

The work described in this chapter has permitted examination of the genetic structure and population connectivity of *P. margaritifera* across its ~18,000 km natural distribution for the first time. Overall, it is clear that the mechanisms underlying population organisation are very complex, and driven by several evolutionary, ecological and physical factors, including

surface ocean currents, geographic distances, selective pressures, larval competency and habitat geomorphology. The data generated also holds regional fishery management implications for Pacific populations of *P. margaritifera*, with the discovery of five distinct genetic stocks in the region. Given the economic importance of pearl oyster aquaculture for several Pacific Island nations (Ponia 2010; Southgate *et al.* 2008), these data provide a benchmark for further evaluation of fine-scale population structure at the level of individual countries and territories, to inform localised fishery management policies. For Indian Ocean populations, the clear divergence detected suggests the need for further investigation at higher sampling densities, with particular attention paid to populations centrally located within the ocean basin. Furthermore, the degree of genetic structure detected across both ocean basins underscores the need to examine evolutionary relationships in *P. margaritifera*, to clarify its taxonomy. The results presented here are also extend to fishery assessment, management and aquaculture development exercises in other broadcast spawning marine taxa, as an informed approach for designating stock boundaries relies on robust datasets comprising ecological, evolutionary and physical information.

CHAPTER 5: Assessing the taxonomic identity of *P. margaritifera*

5.1 Introduction

As outlined earlier in this thesis, *P. margaritifera* is a marine bivalve mollusc that has a broad Indo-Pacific distribution spanning over 18,000 km, and is highly valued for cultured pearl and pearl shell production (Southgate *et al.* 2008; Wada & Tëmkin 2008). While several studies have examined contemporary population structure and connectivity in *P. margaritifera* (Arnaud-Haond *et al.* 2003a; Arnaud-Haond *et al.* 2004; Arnaud-Haond *et al.* 2008; Benzie & Ballment 1994; Durand & Blanc 1988; Durand *et al.* 1993; Lal *et al.* 2016a, unpubl.; Lal *et al.* 2016b; Lemer & Planes 2012, 2014), none have investigated its range-wide evolutionary history and species identity.

Descriptions of *P. margaritifera* include a total of six subspecies (Gervis & Sims 1992), along with a former seventh (*P. margaritifera mazatlanica*), that has since been elevated to species rank: the Panamanian pearl oyster *P. mazatlanica* (Hanley, 1856); (Cunha *et al.* 2011; Hanley & Wood 1842; Wada & Tëmkin 2008). The six subspecies of *P. margaritifera* are described exclusively on the basis of morphological characters (Jameson 1901), and are closely associated with their geographic type locations (see **Table 5.1** for a summary). In the Pacific basin, Hawaiian populations are known as *P. margaritifera* var. *galstoffi*, Cook Islands and French Polynesian individuals *P. m.* var. *cummingi*, and all Central and western Pacific specimens *P. m.* var. *typica*. Indian Ocean populations are represented by *P. m.* var. *persica* (Persian Gulf), *P. m.* var. *erythraensis* (Red Sea) and *P. m.* var. *zanzibarensis* (East Africa, Madagascar and Seychelle Islands). Given the phenotypic and adaptive plasticity and morphological diversity present in Pteriid pearl oysters (Masaoka & Kobayashi 2005a, b), and that shell shape and size differences are apparent between populations of *P. margaritifera* (Allan 1959; Wada & Tëmkin 2008), molecular data are required to elucidate evolutionary significant units (ESUs) in this species.

Table 5.1 Currently established subspecies of the black-lip pearl oyster as summarised by Gervis and Sims (1992).

Subspecies	Authority	Regional distribution	References	Remarks
P. margaritifera var. *cummingi*	(Reeve, 1857)	Cook Islands and French Polynesia	Coeroli *et al.* (1984) Galstoff (1933) Gug (1957) Hedley (1924) Jameson (1901) Ranson (1961) Saville-Kent (1905)	Classification appears to be based on morphology, most recently described by Jameson (1901) and Hynd (1955).
P. margaritifera var. *typica*	(Linnaeus, 1758)	Ryukus Is., Japan; Taiwan, Australia (GBR), Fiji Is.	Hedley (1924) Hynd (1955) Saville-Kent (1905)	Classification appears to be based on morphology, most recently described by Jameson (1901) and Hynd (1955).
P. margaritifera var. *galstoffi*	(Bartsch, 1931)	Hawai'i	Cahn (1949) Galstoff (1933) (Wada & Tëmkin 2008)	Classification appears to be based on morphology. Originally described by Bartsch (1931) as *P. galstoffi*.
P. margaritifera var. *erythraensis*	(Jameson, 1901)	Red Sea	Jameson (1901)	Classification appears to be based on morphology, most recently described by Jameson (1901). Length/weight relationships characterised by Elamin and Elamin (2014).
P. margaritifera var. *zanzibarensis*	(Jameson, 1901)	East Africa, Madagascar, Seychelle Is.	Jameson (1901)	Classification appears to be based on morphology, most recently described by Jameson (1901).
P. margaritifera var. *persica*	(Jameson, 1901)	Persian Gulf	Jameson (1901)	Ranjbar *et al.* (2016) suggest that this subspecies is an independent ESU, and should be revised as a separate species named *P. persica.*
P. margaritifera var. *mazatlanica*	(Hanley, 1856)	Baja California, Panama Bay	Jameson (1901) (Hanley & Wood 1842)	Currently recognised as a species in its own right as *P. mazatlanica*, however, the most recent phylogenetic reconstruction by Tëmkin (2010) suggests conspecifity with *P. margaritifera*.

Because of the degree of intraspecific variation documented in *P. margaritifera* across its extensive Indo-Pacific distribution, previous localised molecular studies have suggested that it might constitute a species complex (Cunha *et al.* 2011; Ranjbar *et al.* 2016; Tëmkin 2010; Wada & Tëmkin 2008). Particular examples include a thorough morphological and molecular characterisation of the superfamily Pterioidea by Tëmkin (2010), who reported that *P. mazatlanica* formed an unresolved clade with *P. margaritifera*, suggesting their conspecificity. In the Persian Gulf, Ranjbar *et al.* (2016) using mitochondrial COI data discovered that *P. m.* var. *persica* formed a divergent ESU, and suggested its reclassification as *P. persica*, while *P. m.* var. *zanzibarensis* from Mauritius formed a basal clade to French Polynesian and Japanese specimens (Cunha *et al.* 2011).

Species-level taxonomic relationships in the genus *Pinctada* are somewhat confused, particularly because many earlier descriptions either heavily, or exclusively relied on morphological descriptions of shells, which are now known to display considerable phenotypic, developmental and environmental plasticity (Masaoka & Kobayashi 2005a; Wada & Tëmkin 2008). With the increasing use of molecular tools to resolve both higher and lower level relationships in this taxon, clarity in the taxonomic identity of several species important for cultured pearl production is being established. An example of this which is still pending resolution, is the status of the *P. fucata*/*martensii*/*radiata*/*imbricata* (Akoya pearl oyster) species complex. It is has recently been recognised that this group may comprise one cosmopolitan, circum-globally distributed species, possessing a very high degree of intraspecific variation across its range (Southgate *et al.* 2008; Wada & Tëmkin 2008).

Investigations of the *P. fucata*/*martensii*/*radiata*/*imbricata* complex may also provide insights into unravelling genealogical units in *P. margaritifera*, given that both sets of taxa share extensive geographical distributions. On the basis of morphological data and despite considerable attention, no discrete and diagnostic characters to separate populations according to geographic origin have been established for the Akoya complex (Tëmkin 2010). Additionally, equivalent morphological information for *P. margaritifera*

collected in a systematic fashion remains absent, which needs to be addressed for a comprehensive examination of the taxon, in a parallelised approach with molecular data.

Recent efforts to characterise distribution-wide population genetic structure in *P. margaritifera* (see chapter 4), indicate a high degree of homogeneity within the Pacific basin, bringing into question the subspecies classifications of *P. m.* var. *typica* and *P. m.* var. *cummingi*. Conversely, populations examined from the Indian Ocean displayed substantial vicariance from Pacific Ocean populations, possibly supporting the existence of distinct ESUs in that region. The goals of the present study were to use genome-wide SNP and dominant marker data to resolve species-level phylogenetic relationships within *P. margaritifera*, and thus provide a benchmark for review of the taxon as a species complex. This investigation is the first of its kind to assess individuals spanning the distributional range of the black-lip pearl oyster, and the data generated has high utility for informing regional spatial marine management strategies for conservation and aquaculture efforts, as well as resolving its taxonomic identity.

5.2 Methods and materials

5.2.1 Specimen collection, tissue sampling and DNA extraction

Specimens between 7-18 cm in dorso-ventral measurement (DVM) were collected from several sites spanning the natural distributions of five Pteriid pearl oyster species (**Figure 5.1**). Black-lip pearl oysters (*Pinctada margaritifera*; n=69) were sampled from 14 sites, which for the Indian Ocean included five samples from each of two Tanzanian sites; (Mafia Island and Mtwara), and Post Office Island in the Abrolhos Islands group, Western Australia; with four samples from the Persian Gulf (Hendorabi Island, Iran). Five oysters each were also sampled from western and central Pacific Ocean populations, including Checheng, Taiwan; Nha Trang, Vietnam; Manado, Indonesia; Kavieng, Papua New Guinea; Gizo Island in the Solomon Islands; Arlington, Sudbury and Tongue Reef systems within the Great Barrier Reef (GBR), Australia; Savusavu, Fiji Islands, and

Tongatapu, Tonga, respectively. In the eastern Pacific, five oysters each were collected from Manihiki Atoll in the Cook Islands, and Arutua, French Polynesia respectively.

Silver-lip pearl oyster (*P. maxima*, n=29) specimens were obtained from Hainan Island, China; Phú Quốc, Vietnam; Broome, Western Australia; Thursday Island in the Torres Straits, Australia; Hervey Bay, eastern Australia and Gizo Island in the Solomon Islands. Five samples were obtained from each of these sites, with the exception of Hervey Bay, where four oysters were collected. *P. maxima* was included here as it is the closest known relative of *P. margaritifera* (Yu & Chu 2005). Panamanian pearl oyster (*P. mazatlanica*, n=10) specimens were also collected from a single site at Guaymas, Mexico, while Akoya pearl oysters (*P. fucata martensii*, n=1 and *P. imbricata*, n=3) were collected from Okinawa, Japan and Port Stephens, Australia; respectively. These additional taxa were selected as they were recovered by Yu and Chu (2005) and Cunha *et al.* (2011) as separate clades to *P. margaritifera*, and consequently included to provide a basis for comparison when using conserved dominant loci in the phylogenomic analysis. Penguin's winged pearl oyster *Pteria penguin* was selected as the outgroup taxon, with specimens (n=5) obtained from Savusavu, Fiji Islands.

Tissue sampling, gDNA extraction and quantification were as described in chapter 4. All samples were subsequently normalised at 100 ng/μL in a 50 μL final volume, and submitted for DArTseq™ 1.0 genotyping at Diversity Arrays Technology PL, Canberra, ACT, Australia.

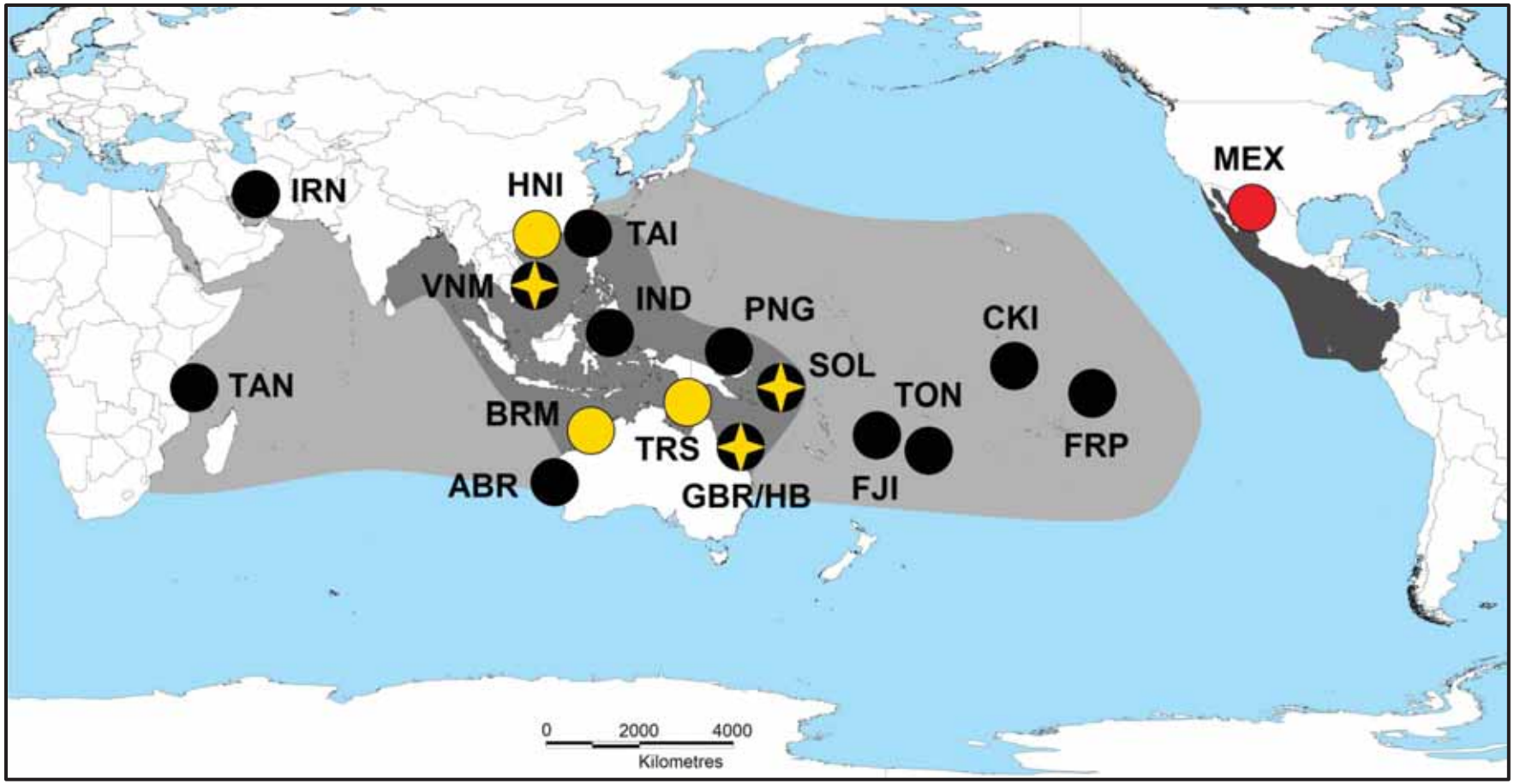


Figure 5.1 Map of sampling locations from where specimens of *P. margaritifera* (n=69; solid black circles and black circles superimposed with yellow stars), *P. maxima* (n=29; yellow circles and black circles superimposed with yellow stars) and *P. mazatlanica* (n=10; solid red circle), were collected. The approximate known distributions of each species are presented in light grey (*P. margaritifera*), medium grey (*P. maxima*) and dark grey (*P. mazatlanica*) respectively; and adapted from Wada and Tëmkin (2008). Site codes represent the following locations: **TAN** Mafia Island and Mtwara, Tanzania; **IRN**: Hendorabi Island, Iran; **TAI**: Checheng, Taiwan (*P. margaritifera*), **HNI**: Hainan Island, China (*P. maxima*); **VNM**: Nha Trang (*P. margaritifera*) and Phú Quốc (*P. maxima*), Vietnam; **IND**: Manado, Indonesia; **ABR**: Abrolhos Islands, Australia; **BRM**: Broome, Australia; **TRS**: Torres Strait, Australia; **GBR/HB**: Great Barrier Reef (*P. margaritifera*) and Hervey Bay (*P. maxima*), Australia; **PNG**: Kavieng, Papua New Guinea; **SOL**: Gizo Island, Solomon Islands; **FJI**: Kadavu, Savusavu, Lau and the Yasawa group, Fiji Islands; **TON**: Tongatapu, Tonga; **CKI**: Manihiki Atoll, Cook Islands; **FRP**: Arutua, French Polynesia and **MEX**: Guaymas, Mexico.

Proximal mantle and adductor muscle tissues (3 and 6 cm respectively) were removed from each specimen and transferred to tubes containing 20% salt saturated dimethyl sulfoxide (DMSO-salt) preservative (Dawson *et al.* 1998). Genomic DNA was extracted using a modified cetyl trimethyl ammonium bromide (CTAB, Amresco, cat. #0833-500G) chloroform/isoamyl alcohol protocol with a warm (30 °C) isopropanol precipitation (Adamkewicz & Harasewych 1996). To clean up all DNA extractions, a Sephadex G50 (GE 2007) spin column protocol was used prior to quantification with a Nanodrop 1000 Spectrophotometer (Thermo Scientific). All samples were subsequently normalised at 100 ng/μL in a 50 μL final volume, and submitted for DArTseq™ 1.0 genotyping at Diversity Arrays Technology PL, Canberra, ACT, Australia.

5.2.2 DArTseqTM 1.0 library preparation and sequencing

Diversity Arrays Technology (DArT PL) proprietary genotyping by sequencing (DArTseq™) reduced-representation libraries were prepared as described in chapter 4. A total of 288 samples were normalised and pooled using an automated liquid handler (TECAN, Freedom EVO150), at equimolar ratios for sequencing in single lanes on the Illumina HiSeq 2500 platform. After cluster generation and amplification (HiSeq SR Cluster Kit V4 cBot™, cat.# GD-401-4001), 77 bp single-end sequencing was performed at the DArT PL facility in Canberra, Australia.

5.2.2.1 Sequence quality control, marker filtering and genotype calling at DArT PL

Raw read processing, initial filtering and genotype calling were as described in chapter 4. In addition to genome-wide SNP loci, presence-absence variant (PAV) markers (termed SilicoDArT loci), were also identified using restriction site-associated (RAD) fragments recovered in the sequence data. SilicoDArTs were scored in a binary fashion, with "1"=RAD fragment presence, "0"=RAD fragment absence and "-"=insufficient counts re-classified as "1"; indicating a hemizygote state. DArTseq PAV markers can be considered to be genome-wide "dominant" markers (Przyborowski *et al.* 2013; Steane *et al.* 2011; Steane *et al.* 2015), and were called based on a minimum reproducibility of

95%. Once SNP and PAV markers had been confidently identified, each locus was assessed in the proprietary KD Compute package for homozygote and heterozygote call rate, frequency, polymorphic information content (PIC), average SNP count, read depth and repeatability, before final genotype scores were supplied by DArT PL.

Following the receipt of genotype data from DArT PL, the SNP dataset was initially filtered using a custom Python script (https://github.com/esteinig/dartQC). This script retained only a single informative SNP (determined by call rate) at each genomic locus, and then filtered all SNPs at a Minor Allele Frequency (MAF) of 2%. Final filtering of the SNP dataset was by call rate (>70%), read depth (>4) and reproducibility (>95%). PAV markers were filtered manually, first to retain the most informative marker at each genomic locus, and then in the order of call rate (100%), MAF (>9.6%), reproducibility (>98%) and read depth (>113).

5.2.3 Assessment of differentiation between population groups and species

Nei's (1973) minimum genetic distances (D_m) between populations were computed in Genetix v.4.05.2 with 1,000 permutations (Belkhir *et al.* 1996) for the SNP dataset, while for the PAV dataset, Nei's unbiased (1972) genetic distances (D) were calculated after Lynch and Milligan (1994), using the AFLP-SURV v.1.0 package and 10,000 permutations (Vekemans 2002; Vekemans *et al.* 2002). Estimates were calculated for both the SNP and PAV datasets, to permit comparison between the two marker types. Computations using the PAV dataset had to be limited to 5,000 loci, as this was the maximum number of sites able to be handled by the software. Random sub-sampling of different sets of 5,000 loci from the larger dataset, and recalculation of genetic distances ensured estimates remained unaffected.

5.2.4 Phylogenomic reconstruction

Population and species-level relationships were reconstructed using neighbour-joining (NJ), maximum-likelihood (ML) and Bayesian inference approaches. The NJ tree was constructed in the MEGA6 software package (Tamura *et al.* 2013). A matrix of Nei's

(1972) genetic distances calculated in Genetix v.4.05 (Belkhir *et al.* 1996) was used as input, and the consensus tree generated over 1000 bootstraps using the SNP dataset. The outgroup taxon selected was *P. maxima*, as the more distant outgroup of *Pt. penguin* was highly divergent from *P. margaritifera*, and as a result received very low genotyping coverage in the SNP dataset (>99% missing data, see results **5.3.1**).

Additionally, the SNPhylo package (Lee *et al.* 2014) was used to perform ML analysis using the SNP dataset. As this dataset had already been filtered, the -r flag was selected which bypassed filtering options for low SNP sample coverage and missing data, and the -l flag set at 1.0 to ignore filtering for linkage disequilibrium (LD) for the final analysis. Preliminary runs were performed both with and without LD pruning to ascertain if this effected changes in tree topology, and no rearrangement of higher-level groups was observed. The transition/transversion ratio was set at 2.0 (-T flag), empirical base frequencies used (-F flag), and constant rate variation set among sites (-R flag). All other options remained at their default settings. SNPhylo first performs a multiple sequence alignment by MUltiple Sequence Comparison by Log-Expectation (MUSCLE; Edgar 2004), and generates a ML tree using the DNAML program in the PHYLIP package (Baum 1989), with bootstrap support provided by the *R* package *phangorn* (Schliep 2011). The final tree was generated following 100,000 bootstraps, with *P. maxima* used as the outgroup taxon.

Bayesian inference of phylogenetic relationships was carried out using only the PAV dataset, with the MrBayes v3.2 package (Drummond & Rambaut 2007; Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003; Ronquist *et al.* 2012). Bayesian reconstruction was also attempted using the SNP dataset, utilising both MrBayes 3.2 and the SNAPP package (Bryant *et al.* 2012). However, these analysis packages are not appropriate for large SNP datasets (*i.e.* MrBayes does not support IUPAC codes and SNAPP is only capable of processing small SNP datasets of ~500 characters), and therefore failed to produce meaningful results. For the PAV dataset, parameter settings were adapted after Koopman *et al.* (2008). The analysis incorporated two runs of 60,000,000 generations each, with each run comprising 4 independent chains. A temperature of 0.10 was set for the heated chains, with a sampling frequency of 1000 and

burn-in fraction of 20%. The burn-in threshold was selected on the basis that both independent runs had achieved convergence (*i.e.* stable log likelihood values reached for all sampled trees, gauged by the average standard deviation of split frequencies). Convergence was also independently assessed using Tracer v.1.6 (Rambaut *et al.* 2003). The Dirichlet prior for state frequencies was set at (90, 10), matching the frequencies of "0" and "1" PAV scores present in the dataset. The outgroup taxon set incorporated all individuals of *Pteria penguin*, together with the additional taxa *P. imbricata* and *P. fucata martensii*. Inclusion of these samples became possible as the PAV dataset was more informative compared to the SNP dataset, for these more evolutionarily distant specimens.

The final trees for the Bayesian analysis were generated by selecting only those post-burn in trees found with the highest individual and cumulative posterior probabilities (p=0.000 and P<0.015 respectively), during Markov Chain Monte Carlo (MCMC) computations. A consensus tree was then constructed from this final credible set of trees using a 50% majority consensus rule in the Dendroscope 3.5.7 package (Huson *et al.* 2007). All phylograms were visualised, inspected and edited in FigTree v.1.4.2 (Rambaut 2014).

5.3 Results

5.3.1 Genotyping and SNP/PAV filtering

The raw SNP dataset contained a total of 23,599 SNPs genotyped across all 118 individuals, at call rates ranging from 20-100%. The first filtering step undertaken to remove duplicate (clone) SNPs at genomic loci resulted in the removal of 5,634 SNPs (24% loss), after which the dataset was filtered for call rate (>70%), average PIC (>1%), MAF (>2%) and average repeatability (>95%). A total of 2 loci were monomorphic within a single taxon, and subsequently removed. These steps collectively resulted in the retention of 8,308 SNPs, across 107 individuals. A total of 11 individuals across 3 taxa; (a single *P. margaritifera* sampled from Iran, along with all *Pt. penguin*, *P. imbricata* and *P. fucata martensii* specimens), were excluded from the final dataset due to poor call

rates (<1%). These taxa were the most divergent compared to all others considered for analysis, and likely experienced poor genotyping coverage due to severe allelic dropout. Consequently, all analyses using the SNP dataset were assigned *P. maxima* as the outgroup.

The raw PAV dataset contained a total of 42,159 variant scores across all individuals, at call rates ranging from 90-100%. No duplicate genomic loci were represented in the dataset, and following filtering for call rate (100%), MAF (>9.6%) and average reproducibility (>98%), 12,212 PAVs remained. No individuals were lost due to poor call rates as with the SNP loci, however the dataset was trimmed to 10,000 PAVs to maximise computational efficiency. All 118 taxa were included in the analyses, and *Pt. penguin* retained as the outgroup taxon.

5.3.2 Assessment of genetic distances between population groups and species

Genetic distance estimates differed between the SNP and PAV datasets as expected due to the different marker systems (**Table 5.2**), although overall trends for the majority of pairwise comparisons were similar. Nei's (1973) minimum genetic distances (D_m) values obtained from the SNP dataset indicated that *P. maxima* from all locations were significantly and substantially divergent from all *P. margaritifera* and *P. mazatlanica* populations, (D_m>0.22; p<0.05). Among sampling sites for *P. maxima*, estimates ranged from 0.023 (Hervey Bay and Vietnam) to 0.069 (Hervey Bay and Solomon Islands), confirming their conspecificity. Pacific Ocean populations of *P. margaritifera* were largely homogenous (D_m=0.006-0.038; p<0.05), and only showed signs of divergence when compared against Indian Ocean populations from Tanzania (D_m=0.076-0.089) and Iran (D_m=0.069-0.113). Interestingly, all *P. mazatlanica* estimates (D_m=0.057-0.133; p<0.05), fell within the range limits of all between-site *P. margaritifera* estimates (D_m=0.006-0.113; p<0.05).

Nei's (1972) unbiased genetic distance (*D*) estimates also indicated clear separation of *P. maxima* from *P. margaritifera* (*D*=0.126-0.296; p<0.05), and *P. mazatlanica* (*D*=0.165-0.307; p<0.05) samples, in keeping with the trend observed for D_m values. Similarly,

broad-scale divergence between *P. mazatlanica* and *P. margaritifera* was not apparent, although the Abrolhos, Tanzanian and Iranian populations did display slightly increased values (0.051, 0.068 and 0.047, respectively), compared to the range estimated for Pacific populations (D=0.034-0.038, $p<0.05$), reflecting divergence between ocean basins.

5.3.3 Phylogenomic reconstruction

5.3.3.1 NJ and ML approaches

Both NJ and ML reconstructions (**Figures 5.2** and **5.3**) produced similar topologies, demonstrating clear separation of *P. margaritifera* and *P. maxima* into species-level clades, as expected. Interestingly, within the larger monophyletic *P. margaritifera* clade, all individuals of *P. mazatlanica* formed a single basal group with short internal branch lengths, and node support for this division in both NJ and ML reconstructions was high (100). Similarly, all Tanzanian with Iranian *P. margaritifera* also formed a distinct monophyletic group, which was basal to the larger monophyletic clade resolved for all other *P. margaritifera* samples. As with the *P. mazatlanica* clade, high bootstrap node support (100) and short internal branch lengths were resolved for this division. Overall, the shallow levels of divergence documented among Pacific Ocean *P. margaritifera* populations suggests their conspecificity, and consequently does not support the current subspecies classifications of *P. margaritifera* var. *typica* and *P. m.* var. *cummingii*. Monophyly of *P. mazatlanica*, as well as of the combined Iranian and Tanzanian populations of *P. margaritifera* within the larger *P. margaritifera* cluster, is suggestive of discrete ESUs present in these locations. Examination of Iranian and Tanzanian specimens in isolation indicates a paraphyletic relationship between the two groups, and underscores the need for further work to resolve evolutionary relationships among Indian Ocean black-lip pearl oyster populations.

5.3.3.2 Bayesian approach for PAV dataset

A total of 480,002 trees were sampled from both runs, and following discard of the burn-in set, 414,101 credible trees remained for calculation of posterior probabilities. The final

average standard deviation of split frequencies achieved was 0.013, with an average potential scale reduction factor (PSRF) for parameter values of 1.000. Within the final set of credible trees, cut-off thresholds of individual ($p=0.000$) and cumulative ($P<0.009$) posterior probabilities were implemented, to select 1,141 highly credible trees with which to construct a consensus phylogram (**Figure 5.4**). The reconstruction resolved three major groups, corresponding to established species-level divisions of *P. fucata martensii* with *P. imbricata* (Akoya species complex), *P. margaritifera* and *P. maxima*. As with the NJ and ML reconstructions, two distinct monophyletic groups were resolved within the larger *P. margaritifera* clade. However, for the sub-group which included *P. mazatlanica*; French Polynesian, Cook Islands, Fijian and Tongan *P. margaritifera* were found to nest together. Similarly, all Abrolhos Islands individuals clustered alongside the Tanzanian and Iranian individuals. It is possible that due to the lower discriminating power of the PAV dataset compared to SNPs at the species level, these more distal relationships were not able to be resolved. This may also explain the similar branch tip lengths observed for all taxa in the final phylogram.

Table 5.2 Pairwise Nei's minimum (D_m, 1973), and Nei's unbiased (D, 1972) genetic distance estimates between sampling locations and species, presented below and above the diagonal for SNP and DArTseq PAV datasets respectively. D_m estimates were computed using Genetix v.4.05.2 (Belkhir *et al.* 1996), while D estimates were computed in AFLP-SURV v.1.0 (Vekemans 2002). All values are significant ($p<0.05$) following permutation. Population groups for *P. margaritifera* were assigned according to chapter 4 genetic clusters, while groups for *P. maxima* were retained according to discrete sampling locations. Sampling site codes are identical to those detailed for **Figure 5.1**.

		P. margaritifera								***P. maxima***						***P. mazatlanica***
	SITE	ABR	CKI	FRP	FIJ+TON	PNG+SOL +GBR	TAI+VNM +IND	TAN	IRN	HB	HNI	SOL	TRS	VNM	BRM	MEX
P. margaritifera	ABR		0.017	0.013	0.013	0.012	0.011	0.035	0.022	0.157	0.173	0.279	0.171	0.141	0.177	0.051
	CKI	0.038		0.002	0.002	0.003	0.005	0.036	0.018	0.142	0.158	0.262	0.156	0.126	0.161	0.037
	FRP	0.036	0.027		0.000	0.001	0.002	0.033	0.014	0.136	0.152	0.255	0.149	0.120	0.155	0.034
	FIJ+TON	0.024	0.018	0.018		0.000	0.001	0.031	0.013	0.138	0.154	0.258	0.152	0.122	0.158	0.035
	PNG+SOL+GBR	0.024	0.018	0.017	0.007		0.000	0.029	0.012	0.137	0.153	0.257	0.151	0.122	0.157	0.037
	TAI+VNM+IND	0.022	0.020	0.018	0.009	0.006		0.026	0.011	0.138	0.154	0.257	0.151	0.122	0.157	0.038
	TAN	0.082	0.089	0.089	0.081	0.078	0.076		0.013	0.172	0.189	0.296	0.186	0.156	0.193	0.068
	IRN	0.103	0.113	0.111	0.104	0.101	0.099	0.069		0.141	0.157	0.261	0.155	0.126	0.161	0.047
P. maxima	HB	0.269	0.286	0.287	0.278	0.277	0.277	0.263	0.255		0.023	0.080	0.006	0.023	0.008	0.182
	HNI	0.259	0.276	0.277	0.267	0.266	0.266	0.256	0.248	0.032		0.066	0.029	0.005	0.037	0.199
	SOL	0.232	0.248	0.247	0.238	0.238	0.237	0.235	0.229	0.069	0.064		0.085	0.078	0.093	0.307
	TRS	0.265	0.282	0.283	0.273	0.273	0.273	0.262	0.255	0.025	0.037	0.068		0.032	0.009	0.197
	VNM	0.271	0.288	0.289	0.279	0.279	0.278	0.265	0.257	0.023	0.029	0.064	0.028		0.039	0.165
	BRM	0.256	0.273	0.274	0.264	0.265	0.264	0.254	0.246	0.032	0.042	0.067	0.037	0.037		0.203
P. mazatlanica	MEX	0.072	0.063	0.058	0.059	0.057	0.058	0.114	0.133	0.289	0.278	0.252	0.284	0.289	0.274	

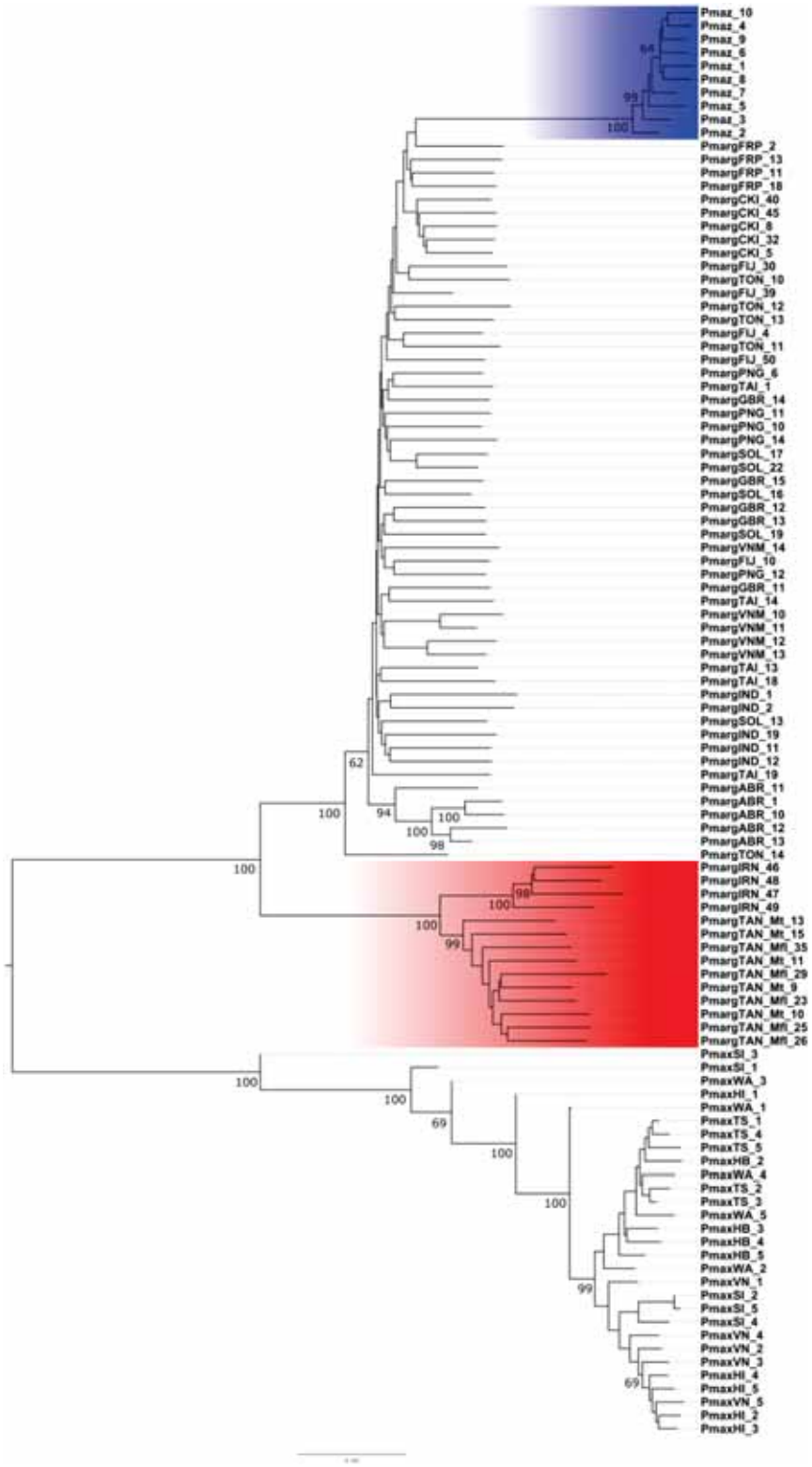


Figure 5.2 Neighbour-joining tree generated in MEGA6 (Tamura *et al.* 2013) using 8,308 SNPs based on Nei's 1972 genetic distances. *P. maxima* was used as the outgroup taxon. Values reported at nodes indicate bootstrap support at a threshold higher than 60 (1000 bootstraps used). Clades for *P. mazatlanica* and Iranian with Tanzanian *P. margaritifera* are highlighted in blue and red respectively.

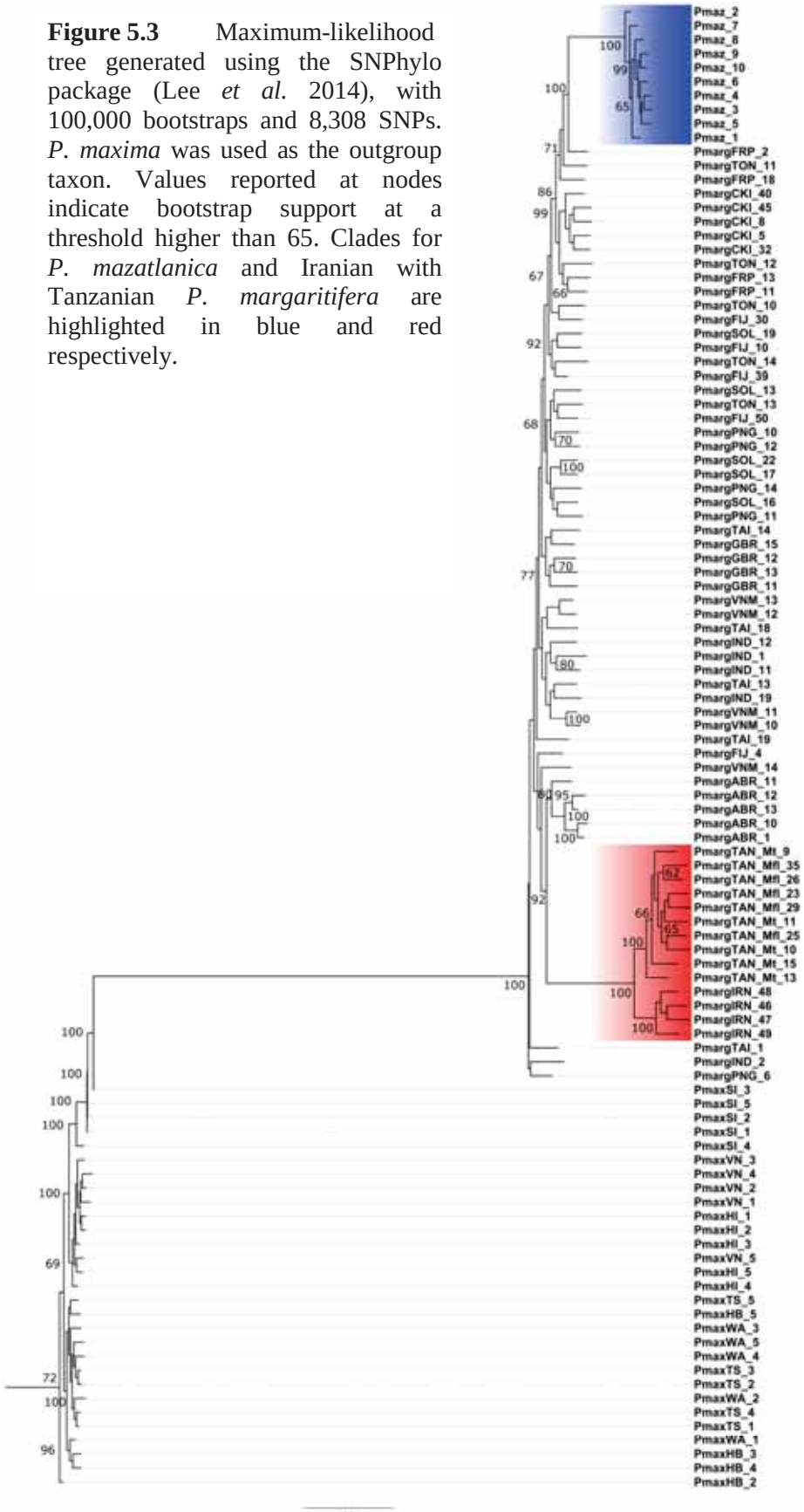


Figure 5.3 Maximum-likelihood tree generated using the SNPhylo package (Lee *et al.* 2014), with 100,000 bootstraps and 8,308 SNPs. *P. maxima* was used as the outgroup taxon. Values reported at nodes indicate bootstrap support at a threshold higher than 65. Clades for *P. mazatlanica* and Iranian with Tanzanian *P. margaritifera* are highlighted in blue and red respectively.

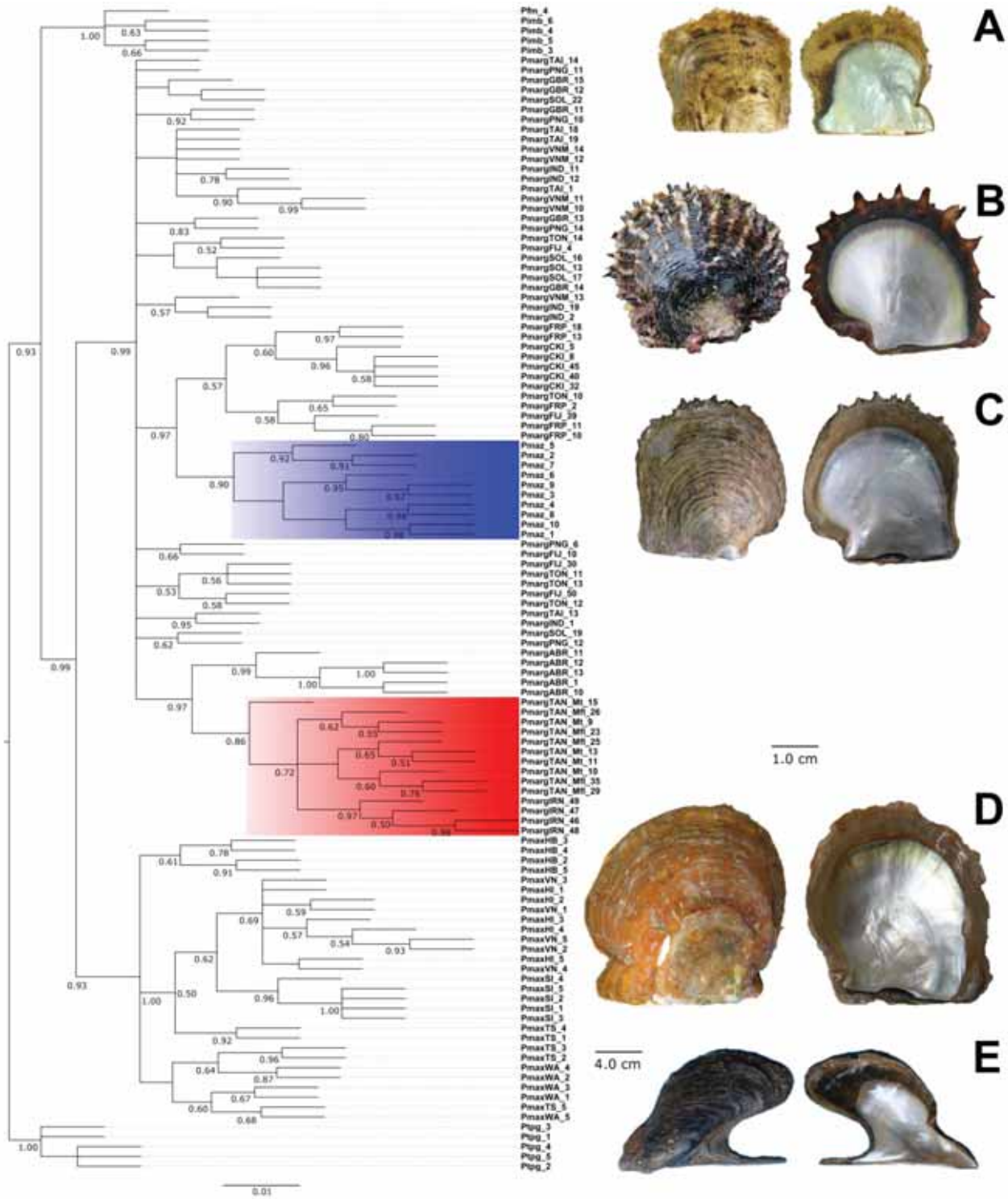


Figure 5.4 Bayesian reconstruction generated using 10,000 DArTseq PAVs in MrBayes v.3.2 (Ronquist *et al.* 2012). The consensus tree reported here was generated from 1,141 of the most credible set of trees (p=0.01), using a 50% majority consensus rule. Posterior probabilities≥0.50 for node support are indicated below the branches. Clades for *P. mazatlanica* and Iranian with Tanzanian *P. margaritifera* are highlighted in blue and red respectively. Representative specimens of *P. imbricata* (**A**; Wakayama, Japan), *P. margaritifera* (**B;** Savusavu, Fiji Is.), *P. mazatlanica* (**C;** Guaymas, Mexico), *P. maxima* (**D;** Bali, Indonesia) and *Pteria penguin* (**E**; location unspecified), are shown to illustrate gross differences in shell morphology. Images **C** and **D** were kindly supplied by Douglas McLaurin-Moreno and David Jones respectively, while **E** and **A** were adapted from Hessel (nd) and the Natural History Museum of Rotterdam (2013), respectively. All specimens are shown at relative scales, with the exception of *Pt. penguin* (4cm scale bar).

5.4 Discussion

The current study represents the most comprehensive evaluation of evolutionary relationships in the black-lip pearl oyster, incorporating two independent high density genome-wide marker sets and sample representation from the entire species distribution. It is clear that the taxonomy of *P. margaritifera* is more complex than previously thought, as a distinct ESU was identified across the Tanzanian and Iranian populations. Additionally, the previously recognised species *P. mazatlanica* nested as a basal clade to Pacific Ocean *P. margaritifera*, suggesting that its status may require revision. In the light of these findings, it is highly likely that *P. margaritifera* may constitute a species complex, which requires further investigation to fully resolve its taxonomic identity.

5.4.1 Evolutionary relationships within *P. margaritifera*

The presence of morphological differences (Allan 1959; Jameson 1901) between locations in *P. margaritifera* is unsurprising, considering that the species distribution spans over 18,000 Km across heterogeneous habitats (see chapter 4; Wada & Tëmkin 2008), and that bivalve molluscs can display very high levels of phenotypic plasticity (Cunha *et al.* 2011; Zhong *et al.* 2014). The high degree of morphological variation and overlapping geographical distributions of many bivalves due to high larval dispersal capability, however, can make the delimitation of population and species boundaries problematic, highlighting the need for molecular information to resolve these differences. The data presented in the current study indicates varying degrees of support for the current morphological subspecies classifications for *P. margaritifera*.

In the Pacific Ocean, the existence of the subspecies *P. m.* var. *typica* and *P. m.* var. *cummingii* is questionable. Work carried out in chapter 4 to examine genetic structure in the Pacific basin indicates a high degree of support for this finding, as results showed *P. margaritifera* experiences very high gene flow between populations separated by several thousands of kilometres. While five discrete Pacific Ocean stocks were identified, populations remained sufficiently undifferentiated to resolve any subspecies groupings. It is possible that the morphological differences recorded between Pacific populations (Allan 1959; Jameson 1901), may be the result of local adaptation and habitat differences. Work presented in chapter 4 detected signatures of selection between all five stocks of *P.*

margaritifera identified in the Pacific, and given the capacity of many bivalve taxa to adapt to diverse environments (Riginos & Cunningham 2005; Takeuchi *et al.* 2016), it is possible that local selective pressures may play a role in manifesting the morphological variation observed.

Despite considerable effort, it was not possible to obtain specimens of Hawaiian *P. margaritifera* for inclusion in the current study, and therefore examination of the status of the third and last Pacific Ocean subspecies described in the literature; *P. m.* var. *galstoffi* (Bartsch 1931; Wada & Tëmkin 2008), remained out of reach. Given the degree of genetic homogeneity between other Pacific populations across similar spatial scales, however, it is possible that Hawaiian populations could also display shallow levels of divergence. This possibility is supported by simulations of larval dispersal by Lal *et al.* (unpubl.), which indicates that larval drift departing from Papua New Guinea has the potential to travel over 5,900 km to reach as far as Kiribati, indicating that Hawaiian populations may also be within reach of neighbouring locations. The closest populations that may be able to supply recruits to the Hawaiian archipelago considering prevailing ocean current circulation are located in Okinawa, Japan, and the Northern Mariana Islands, and the strength of migration between these locations is likely to determine the degree of isolation of Hawaiian *P. margaritifera*. A future study incorporating these populations will be able to address this question.

5.4.2 Taxonomic identity of *P. mazatlanica*

The discovery that *P. mazatlanica* specimens comprised a basal group nested within *P. margaritifera* is interesting, given the lack of an overlap in known species range limits (Wada & Tëmkin 2008). In the light of the substantial trans-Pacific dispersal ability of this species as discussed earlier however, it is entirely possible that recruits originating from eastern Pacific populations could find their way to the Gulf of California, thus maintaining gene flow. Early descriptions of *P. mazatlanica* using morphological characters had in fact classified it as a subspecies of *P. margaritifera* (*P. margaritifera mazatlanica*, and it was noted that its shell morphology appears to be an intermediate form between *P. margaritifera* and *P. maxima* (Hanley & Wood 1842; Jameson 1901; Ranson 1961; Wada & Tëmkin 2008). Assessment of the population genetic structure of *P. mazatlanica* is somewhat confused. Arnaud *et al.* (2000) discovered that mtDNA nucleotide divergence (COI and 12s rRNA) between locations ranged from 0.12-1.3%, while divergence from three individuals of *P. margaritifera*

reached ~4%. Importantly, divergence at mtDNA loci across the broader distribution of *P. margaritifera* remains unknown. Subsequent examinations of species-level taxonomy within the genus *Pinctada* (Tëmkin 2010) using multiple nuclear and mitochondrial loci, found that *P. mazatlanica* formed an unresolved clade with *P. margaritifera*, in concordance with a later reconstruction by Cunha *et al.* (2011), although in their study the latter authors state that results supported the present species-level classification. The genome-wide data presented here provides further strong evidence that these taxa might be conspecific.

These results also support a further observation by Cunha *et al.* (2011), which suggests that French Polynesian *P. margaritifera* are more closely related to *P. mazatlanica* than Indian Ocean specimens. This pattern of differentiation in the data also extended to populations located in the western Pacific in the data, indicating that a clinal effect related to geographic separation may be present, as seen in *Crassostrea* spp. oysters (Zhong *et al.* 2014). The lower than expected levels of divergence separating *P. mazatlanica* from *P. margaritifera*, despite the morphological differences between the two taxa, raises several questions as to the processes driving their divergence, or maintaining genetic structure. One possibility is that *P. mazatlanica* is a distinct ESU within *P. margaritifera*, and that its morphological differences are a consequence of phenotypic and adaptive plasticity. Another scenario is that *P. mazatlanica* may be undergoing incipient speciation (Marques *et al.* 2016), which has been documented in other broadcast spawning marine invertebrates, including bivalves (Dawson 2005; Vierna *et al.* 2014). To resolve these questions, a fine-scale study incorporating range-wide samples of *P. mazatlanica*, together with specimens from marginal eastern Pacific *P. margaritifera* populations (e.g. Hawaii, Kiribati and French Polynesia), is required.

All three phylogenetic reconstruction methods applied in the current study resolved a pattern of paraphyly for Iranian and Tanzanian specimens respectively, which corresponds with the subspecies descriptions of *P. m.* var. *zanzibarensis* (Zanzibar, Madagascar and eastern African coastline) and *P. m.* var. *persica* (Persian Gulf only) (Jameson 1901). This difference in evolutionary trajectories from Pacific Ocean specimens also explains the divergence observed in examination of range-wide population genetic structure in *P. margaritifera* (chapter 4). Separation of Persian Gulf populations as a distinct ESU was also detected by Ranjbar *et al.* (2016), who suggested their reclassification as a species in its own right named *P. persica*. While the findings here do not indicate that specimens from Iran were sufficiently divergent to warrant this elevation in taxonomic rank, further research is required to ascertain

the degree of isolation of Persian Gulf populations from specimens living in the Red Sea (*i.e. P. m. erythraensis*), as well as the broader Indian Ocean.

It is clear that dense sampling of the Indian Ocean is required, as the present study was only able to assess three marginal populations sampled at its geographic limits. At the eastern extent of the Indian Ocean, specimens collected from the Abrolhos Islands (Western Australia) formed a weakly monophyletic clade, which nested closest to the Tanzanian and Iranian specimens, suggesting a restriction in gene flow between the Indian and Pacific Oceans (chapter 4). This observation may also indicate the presence of an Indian-Pacific Ocean genetic break, however a future study incorporating a range of samples from both the Indian Ocean basin centre and periphery is required for confirmation.

5.4.3 Comparison of SNP vs. PAV markers and phylogenetic reconstruction methods

Phylogenetic reconstruction using both genome-wide SNPs and DArTseq PAVs in the current study have demonstrated the utility of both marker types for investigations of contemporary evolutionary relationships in a marine invertebrate. Use of genome-wide SNPs generated by GBS methodology is becoming widespread in phylogenetic investigations (e.g. Bryant *et al.* 2012; Gohli *et al.* 2015; Leaché *et al.* 2015), however, as with population genomic analyses, allelic dropout (ADO) due to restriction cut site mutations in highly polymorphic species can considerably reduce the number of informative markers remaining for final analyses (Andrews & Luikart 2014; Lal *et al.* 2016b; Puritz *et al.* 2014).

A benefit of incorporating dominant markers into phylogenetic analyses of taxa susceptible to ADO, is that a greater overall number of loci may be retained, thus permitting the retention of taxa which otherwise may not meet stringent filtering thresholds. In the current study, the use of PAVs permitted retention of divergent taxa including the *Pt. penguin* outgroup with *P. imbricata* and *P. fucata martensii* specimens, which for the SNP datasets was not possible post-filtering. While care must be exercised when computing and interpreting genetic distances from dominant marker datasets, they can complement the results of SNP datasets by permitting comparative analyses.

Comparative reconstructions using both marker types also permitted mutual assessments of tree topologies, and validated the presence of ESUs among the taxa sampled. Due to the nature of the dominant PAV marker data, it is likely to be useful for resolving higher level taxonomic relationships (e.g. Park *et al.* 2006), while SNP datasets can be used for finer scale resolution of branch tips at the species, or subspecies levels.

5.4.4 The case for the black-lip pearl oyster species complex

The possibility that *P. margaritifera* may constitute a species complex has been suggested by a number of studies, where distinct ESUs have been discovered during localised investigations of genetic structure (Cunha *et al.* 2011; Ranjbar *et al.* 2016; Wada & Tëmkin 2008). The current study incorporating samples spanning the extent of the species distribution contributes further evidence to support this taxonomic classification, and highlights the need for further research to investigate segments of the species range that were not able to be sampled.

Of the five major pearl oyster species which are the focus of commercial aquaculture efforts (Southgate *et al.* 2008; Wada & Tëmkin 2008), the only species distributed over a natural range comparable to that of *P. margaritifera*, is the Akoya complex; *P. fucata/martensii/radiata/imbricata*. The Akoya complex is characterised by substantial intra- and interpopulation morphological variability, local geographic isolation of some populations, human introductions, hybridisation and inconsistent taxonomic practice (Tëmkin 2010; Wada & Tëmkin 2008). Originally, three distinct species were recognised; *P. imbricata* Röding, 1798 (western Atlantic), *P. radiata* ((Leach, 1814); eastern Indian Ocean and Red Sea) and *P. fucata* ((Gould, 1850); Indo-Pacific). Japanese populations were recognised as a distinct species (*P. martensii*, (Dunker, 1872)) or subspecies (*P. fucata martensii*; (Hayami 2000; Matsukuma 2004; Wada & Tëmkin 2008)). These classifications were on the basis of questionable morphological characters, and subsequent molecular analyses (see Wada and Tëmkin (2008) and Tëmkin (2010) for a summary), revealed that Australian, south-east Asian and Japanese populations form a monophyletic group that is highly likely to be conspecific. Furthermore, mating experiments have supported the conspecificity of south-east Asian and Japanese populations (Atsumi *et al.* 2004), and the current consensus is that the Akoya complex may be a cosmopolitan, globally distributed species, possessing substantial intraspecific variation (Wada & Tëmkin 2008). In a comprehensive species-level

investigation of extant Pterioidea using combined morphological and molecular analyses, Tëmkin (2010) also assessed the genus *Pinctada*. This work resolved that *P. imbricata, P. fucata* and *P. radiata* were reciprocally monophyletic, and suggests management as separate ESUs. To settle the taxonomic identity of this group, Tëmkin (2010) proposes that the senior synonym *P. imbricata* be adopted, and that the three distinct genealogical units be recognised as the subspecies *P. imbricata imbricata*, *P. imbricata fucata* and *P. imbricata radiata*.

Given the similarities in morphological variability and the extensive natural distribution of *P. margaritifera* when compared with members of the Akoya species complex, along with findings of the current study, it is certainly feasible that *P. margaritifera* as it is currently known, might comprise a species complex. Ultimate resolution of its taxonomic identity however, will require a large-scale, systematic, molecular and morphological characterisation of samples collected across the entire natural distribution.

5.5 Conclusions

This work represents the first range-wide examination of evolutionary relationships in the black-lip pearl oyster using genome-wide molecular markers, and resolves the taxonomic identity of several populations sampled across its distribution. Additionally, it presents evidence for the conspecificity of *P. mazatlanica* and *P. margaritifera*, and further identifies the presence of discrete ESUs contained within the species distribution. Collectively, these findings provide early indications that *P. margaritifera* is likely to constitute a species complex, and highlight the requirement for further range-wide investigations to fully resolve its taxonomic status. This information is valuable not only for the regional fishery management and aquaculture of *P. margaritifera* given its commercial importance, but also for a broader understanding of the ecology and evolution of similarly widely distributed marine invertebrates.

CHAPTER 6: General discussion

6.1 Future directions for *P. margaritifera*

6.1.1 Fijian pearling industry

Prior to the investigations described in chapters 2 and 3, no information was available on the genetic structure and connectivity of Fijian *P. margaritifera*, and therefore, development of comprehensive fishery management policy and responsible aquaculture practice guidelines was not possible. As a result of these studies, a robust set of genome-wide markers has now been produced and thoroughly tested for the species, which has outlined stock structure for the first time, as well as provided a basis upon which sustainable fishery management can be achieved. Additionally, this data is also highly useful for guiding aquaculture practices; particularly in the management of hatchery-produced oysters to minimise impacts on wild populations (*i.e.* through potential inbreeding or outbreeding effects as seen in French Polynesia: Arnaud-Haond *et al.* 2004; Arnaud-Haond *et al.* 2003b; Arnaud-Haond *et al.* 2008), effective broodstock maintenance to ensure the genetic integrity of farmed oysters, as well as the translocation management of both wild and farmed individuals. Development of a comprehensive translocation protocol for the Fijian pearling industry is a key objective of the ACIAR research project (FIS/2009/057) which funded this PhD research, and the data generated by the investigations described in chapters 2 and 3 will be directly used in its implementation. The major recommendation for farmed oysters is that translocations between sites which freely exchange recruits is permissible, while movement between locations that may share less connectivity should be restricted, until further fine-scale study can establish if translocation is warranted. Such studies should assess any potential negative impacts on existing wild populations, and also establish if any negative stock fitness consequences may arise in the transplanted individuals. For wild oysters, responsible fishery exploitation measures should seek to maintain sustainable spawning-stock biomass within all natural populations, and collection of oysters for aquaculture be limited to spat collection. Furthermore, investigation to investigate the impacts of spat collection on natural populations is warranted, particularly as the Fijian pearling industry develops and intensifies.

6.1.2 Regional fishery management

Despite its commercial importance, there is a distinct lack of information on regional patterns of genetic structure and connectivity in *P. margaritifera*, as outlined in chapters 3 and 4. Research carried out within this PhD candidature fills this substantial knowledge gap, by providing high resolution population genomic data on the *P. margaritifera* resource present in the Indian and Pacific Ocean basins for the first time. While previous studies had examined localised genetic structure and diversity in isolation within the Pacific (e.g. French Polynesia, Japan, Kiribati and the GBR), none had considered basin-wide patterns (Arnaud-Haond *et al.* 2004; Arnaud-Haond *et al.* 2008; Benzie & Ballment 1994; Durand & Blanc 1988; Durand *et al.* 1993; Lemer & Planes 2012, 2014). Contrastingly, Indian Ocean populations had remained almost completely unexamined.

The scope of sampling in the Pacific (14 sites across 12 countries as described in chapter 4), permitted near-complete coverage of the known species distribution, with the notable exceptions of Hawaii and the Federated States of Micronesia. At this scale, the data generated was able to inform regional fishery management of the black-lip oyster resource across country and territorial boundaries. In particular, patterns of genetic structure identified will be valuable for sustainable development of the Fijian and Tongan pearling industries, as their *P. margaritifera* populations are part of a single biological stock. As a result, bilateral cooperation between both countries may be required for the development of joint fishery management and aquaculture strategies. While fine-scale patterns of structure and connectivity are now known for Fijian populations through the study described in chapter 2, a similar study is required for Tonga, with dense sampling coverage across its major archipelagos. Such a study is likely to identify where the outer spatial limits of the joint Fijian-Tongan stock may lie, especially when combined with samples from neighbouring nations (e.g. Vanuatu and New Caledonia to the west, and Niue, Samoa and Tuvalu in the north and east).

Given the discovery of five major stocks in the Pacific region, further localised fine-scale sampling is required within and around their periphery to determine spatial extents for fishery management. As the pearling industries for the Cook Islands and French Polynesia are highly valuable contributors to their respective economies (Ponia 2010), further characterisation of

the respective stocks identified for these countries using high resolution genome-wide markers will be useful. While French Polynesia is a locality where genetic structure and connectivity in *P. margaritifera* has perhaps received the most attention, these studies have all employed lower resolution genetic markers such as microsatellite loci (Arnaud-Haond *et al.* 2003a; Arnaud-Haond *et al.* 2002; Arnaud-Haond *et al.* 2004; Arnaud-Haond *et al.* 2003b; Arnaud-Haond *et al.* 2008; Lemer & Planes 2012, 2014; Lemer *et al.* 2011). With the use of genome-wide SNPs, fine-scale patterns of structure and adaptive variation that may have previously gone undetected could become apparent, which will add valuable information for both fishery management and aquaculture efforts.

The use of genome-wide neutral and adaptive information for identification of discrete stocks also provides a basis for refining the biological definition of a stock for fishery management. While existing criteria mention that demographic parameters are important considerations for managing populations either together or separately (e.g. Carvalho & Hauser 1994; Reiss *et al.* 2009), the identification and consideration of functional differences (e.g. adaptive variation) as highlighted in chapters 3 and 4, is equally important (Nayfa & Zenger 2016). For Fijian *P. margaritifera*, a lack of detectable selective differences among populations indicates that management may proceed by considering all populations as part of a single stock. However, more broadly within the Pacific basin, the adaptive variation identified between the five major stocks suggests that management as separate entities is required.

6.1.3 Taxonomic resolution

Evidence presented in chapter 5 that *P. margaritifera* constitutes a species complex requires further attention, to determine if additional ESUs may be present within the species distribution. Given the relatively sparse sampling achieved within the Indian Ocean, additional sites from this region, along with selected Pacific sites (e.g. Hawaii and Federated States of Micronesia), should be targeted for further study to better understand evolutionary relationships in this taxon. This information will be required if its classification is to be revisited, for detailed characterisation of the species complex.

The presence of a species complex in *P. margaritifera* shares several interesting parallels with the Akoya pearl oyster species complex (*P. fucata/martensii/radiata/imbricata*), as both taxa share very extensive natural ranges, similar life histories, and possess substantial

intraspecific variation (Tëmkin 2010). As the *P. margaritifera* species complex becomes better understood, further information may become available on how large species complexes and/or cryptic species come to exist in aquatic taxa, thus providing additional insights into how evolutionary processes operate within the marine environment (Curini-Galletti & Puccinelli 1998; Lam & Morton 2003).

While the Akoya species complex remains unresolved to an extent (Wada & Tëmkin 2008), the application of genome-wide markers as demonstrated for *P. margaritifera* in chapter 5, may assist in providing clearer answers to questions of genealogical organisation in this taxon and others. Comprehensive analysis of the Pterioidea using nuclear and mtDNA markers, as well as morphological characters by Tëmkin (2010), resolved three distinct ESUs for the Akoya complex, which the author suggests should each be given a subspecies rank. With the addition of genome-wide data, further ESUs may be identified, and a better grasp of broad-scale evolutionary patterns obtained in this commercially important bivalve. However, as genome-wide SNPs remain a relatively new marker type for evolutionary investigations (cf. mtDNA for example), benchmarking divergence levels, species classification thresholds and calibration of molecular clocks to gauge divergence times using these datasets remains under development (Wang & Nielsen 2012). Consequently, until this information becomes available through further investigations on a broad range of taxa, traditional marker sets used for DNA barcoding applications (e.g. mtDNA, ITS), utilised in combination with genome-wide markers (e.g. SNPs, PAVs) may be relied upon to resolve these questions.

Extension of sampling for *P. margaritifera* to include regions immediately outside the currently known species distribution will also be helpful to determine where the natural range terminates, or is likely to be contiguous. A portion of the current distribution that requires dense sampling is Hawaii, with extension towards the Gulf of California, as data presented in chapter 5 strongly indicate that *P. mazatlanica* is a discrete ESU/likely sub-species among Pacific Ocean *P. margaritifera*, instead of a species in its own right. Further comprehensive broad-scale sampling will likely also provide further insights into patterns of genetic structure across the species distribution, as outlined in chapter 4. In particular, core vs. periphery patterns for the Indian Ocean, as well as Pacific Ocean sites which remain unknown may become apparent, which will provide benchmarks for informing regional fishery management and aquaculture.

6.1.4 Pearling industry development

P. margaritifera aquaculture development in terms of selective breeding programmes and characterisation of commercially important traits remains largely under-developed. The greatest progress in these areas of research have been made with the silver-lip pearl oyster *P. maxima*, as well as various members of the Akoya species complex, where large scale selective breeding is underway, the genetic architecture of desirable traits (e.g. shell dimensions, growth rates, nacre colour etc.) is becoming understood, and substantial genomic resources (e.g. high density genome-wide marker sets and linkage maps) are available; (Funabara *et al.* 2013; He *et al.* 2008; Jones *et al.* 2013a; Jones *et al.* 2013b; Jones *et al.* 2014a, b; Miyamoto *et al.* 2013; Mohamed *et al.* 2006; Southgate *et al.* 2008; Yu *et al.* 2009).

There are several reasons why commercial aquaculture of *P. margaritifera* is at a fledgling stage of development, and these include the smaller scale of culture situated primarily in developing countries (c.f. *P. maxima* where operations are mostly large-scale and heavily industrialised), as well as a general lack of data on the natural variation present in wild populations that culture efforts may be able to capitalise on. Furthermore, large-scale commercial production of *P. margaritifera* is largely restricted to the Cook Islands and French Polynesia, with industry development having begun in these countries in the 1970s, whereas *P. maxima* and Akoya complex species have been regularly exploited for longer periods of time (Southgate *et al.* 2008).

Following the contribution of research carried out during this PhD candidature to provide baseline high resolution data on regional and localised genetic variation in *P. margaritifera*, the next steps for aquaculture development will be to understand the extent of natural variation present in wild populations, and to begin to characterise commercially important traits that are of value in selective breeding programmes. While the outcomes of such research may take decades to be realised in commercial production, they provide avenues for achieving productivity gains, as evidenced in similar work carried out for *P. maxima* (Jones *et al.* 2014a, b).

For the Fijian industry, which obtains higher returns on the production of coloured pearls (see **Plate 6.1**) compared to the traditional black pearls for which *P. margaritifera* is known (J. Hunter, *pers. comm.*), investigation of the genetic basis for nacre colour and heritability of

desirable variants, is an important research avenue. The largest Fijian operation J. Hunter Pearls, which contributed an estimated 80% of all marketable pearls produced from Fiji in 2007 (P. Southgate, *pers. comm.*), intentionally utilises orange morphotype saibo donors, instead of black morphotype donors during nucleus implantation. This practice results from reports that orange saibo tissue generates larger proportions of coloured vs. black pearls at harvest (J. Hunter, *pers. comm.*), which systematic investigation using common garden experiments may be able to clarify, and also determine if orange morphs can be selected for in breeding programmes. The study described in chapter 2 successfully detected genomic regions associated with orange vs. black pigmentation, and for future investigation of the genetic architecture of colour traits, a proven genotyping tool and method have been tested and are readily available for use.

The work carried out in this thesis has generated highly valuable genomic resources for *P. margaritifera*, in the form of high quality genome-wide SNP marker sets, which may be used for various future investigations of genetic structure, or trait architecture in this species. Furthermore, the utility of a proven genotyping tool (DArTseq^{TM}) has been demonstrated for delivering these marker sets, which represents a large leap forward and time savings for future marker discovery exercises. The flexibility of the DArTseq^{TM} genotyping tool also lends itself to marker discovery in other species with similar genomic structure and/or life history traits, widening the scope for application to a diverse range of research questions requiring high resolution genetic data.

Plate 6.1 Nacre colour of saibo donor oysters and coloured cultured pearls produced from Fijian *P. margaritifera*. The first shell in the top left photograph is an example of a donor used to produce "Fiji gold" pearls, and the first shell in the top right image is an example of a donor for "chocolate" pearls, all produced by J. Hunter Pearls. The image at lower left illustrates examples of harvested high quality round pearls, with typical colours produced including "chocolate" (second from top), "Fiji gold" (second in bottom row) and "pistachio" (fourth in bottom row). Baroque and circle (grooved) pearls are shown in the lower right photograph. All photographs are reproduced with the permission of J. Hunter Pearls (Fiji).

6.2 Significance of the research and contribution to the field

6.2.1 Utility of genotyping methodology

Population genetic research has benefitted greatly from the next-generation sequencing revolution, and the ability to deliver thousands of truly genome-wide genetic markers to examine a range of questions for both wild and captive organisms (Allendorf *et al.* 2010; Funk *et al.* 2012; Hohenlohe *et al.* 2011). The greatest benefits of NGS technology have arguably been realised for marine taxa, where the detection of fine-scale genetic structure and signatures of selection are now possible for species with highly-dispersive life histories. These levels of discrimination were previously not feasible with traditional markers (e.g. microsatellites and mtDNA), however given this paradigm shift, sufficient sensitivity and resolving power can now be obtained through genome-wide SNPs, to accurately inform conservation and fishery management, aquaculture practices, genome mapping, QTL identification, MAS and various other applications (Angeloni *et al.* 2012; Yue 2013).

Genome-wide SNPs for non-model organisms may be delivered through several genotyping platforms, among which are RADseq and RADseq-based approaches (Puritz *et al.* 2014; Toonen *et al.* 2013; Wang *et al.* 2012), which have promised lower costs and shorter laboratory preparation times compared to older techniques (e.g. microarrays; Davey & Blaxter 2011; Davey *et al.* 2011). Among these, ddRADseq (Peterson *et al.* 2012) was seen to offer a high degree of flexibility for a range of organisms and applications, however since its adoption, was also found to have a number of limitations (Puritz *et al.* 2014). Among the drawbacks of ddRADseq is that for organisms which possess highly polymorphic genomes, restriction site mutations can lead to allelic drop out (ADO), and consequent missing data. This issue was explored in the work carried out for *P. margaritifera* in chapter 2, and a range of laboratory-based and data analysis measures provided for improvement. While ddRADseq remains useful for exploring the genetic architecture of non-model organisms, the issues of ascertainment bias, missing data and standardisation of genome sampling (Arnold *et al.* 2013; Puritz *et al.* 2014), are substantial challenges that remain to be addressed.

An example of a potential solution to improve on the limitations of ddRADseq, is the DArTseq™ GBS method provided by the commercial genotyping service provider Diversity Arrays Technology (DArT PL). This platform was used for the studies described in chapters 4 and 5, and offered several improvements in the quality of genome-wide SNP marker sets obtained, over the ddRADseq approach utilised in chapters 2 and 3. DArTseq™ is a modified ddRADseq protocol, and incorporates a number of advancements over the original method developed by Peterson *et al.* (2012). These improvements include a more standardised sampling of the genome with methylation sensitive restriction enzymes, incorporation of region recovery diagnostics, standardising fragment sizes before sequencing and inclusion of fixed sample replication (Kilian *et al.* 2012; Przyborowski *et al.* 2013; Sansaloni *et al.* 2011). With these innovations, missing data is minimised, larger numbers of SNPs common to all individuals are typed, ascertainment bias is reduced, and taxa that are potentially problematic for RADseq-based genotyping; (such as *P. margaritifera* and other species with highly polymorphic genomes), can now be successfully studied. The examination of distribution-wide genetic structure in divergent populations and evolutionary relationships for *P. margaritifera* as described in chapters 4 and 5 respectively, illustrate the utility of this platform for investigations in other taxa that are faced with similar limitations.

An additional advantage of DArTseq™ genotyping is the provision of PAV or "dominant" genome-wide markers (Przyborowski *et al.* 2013; Steane *et al.* 2011; Steane *et al.* 2015). While the application of PAVs for contemporary population genomic analyses largely remains under evaluation (Lucas-Lledó *et al.* 2014; Tan *et al.* 2012), their utility was demonstrated in a Bayesian reconstruction of evolutionary relationships in *P. margaritifera* (described in chapter 5). Where genome-wide SNPs excluded an outgroup taxon and two congeneric taxa from analysis due to low genotyping coverage as a result of high divergence, use of the PAV marker set enabled their inclusion. A recommendation following phylogenomic reconstruction using both genome-wide SNPs and PAVs, is that due to the differing characteristics of both marker sets, comparative reconstructions may provide valuable information on the taxa sampled. PAV data while

offering lower resolution compared to SNP marker sets, could be used for resolving higher relationships (e.g. genus level or higher), while SNP data may be especially useful for lower level resolution at branch tips (e.g. species or sub-species levels). While phylogenomic analyses remain a relatively new toolset for investigating evolutionary relationships, their utility continues to be proven for a range of taxa and research questions, including cryptic lineages in *Eucalyptus* spp. (Steane *et al.* 2011; Steane *et al.* 2015), phenotypic plasticity in *Crassostrea* spp. oysters (Zhong *et al.* 2014) and resolution of a sub-species complex in Afrocanarian blue tits (Gohli *et al.* 2015).

6.2.2 Combined use of independent genomic and environmental datasets

Marine taxa which employ a broadcast spawning reproductive strategy present a number of challenges for determination of population genetic structure and connectivity, due to the potentially high levels of gene flow that may occur between individuals separated by large geographic distances (André *et al.* 2011; Waples 1998; Waples & Gaggiotti 2006). Ways in which the life history traits of marine invertebrates can confound efforts to determine genetic structure for fishery assessment is discussed in detail in chapter 3, where the importance of multidisciplinary approaches is also highlighted. Investigations that use multiple toolsets such as biophysical and demographic methods, in combination with genome-wide genetic data to investigate genetic structure and population boundaries, may provide for mutually-corroborative results, thus increasing overall confidence that findings are biologically relevant.

Studies carried out in chapter 3 and 4 illustrate the utility and value of using independent genomic and hydrodynamic datasets particularly well for *P. margaritifera* at both local and regional spatial scales, as patterns of genetic structure and population connectivity were strongly supported by particle dispersal simulations in both investigations. While hydrodynamic modelling alone has been used for assessment of dispersal in several marine taxa (Neo *et al.* 2013; Thomas *et al.* 2014; Wood *et al.* 2014), consideration of independent and explicit oceanic circulation information, together with high resolution genomic data as presented here remains a novel approach (Liggins *et al.* 2013). Because population connectivity for the vast majority of highly-dispersive marine taxa is realised

through larval transport, inclusion of hydrodynamic data is strongly recommended for future investigations of genetic structure and gene flow pathways.

The task of modelling larval transport and recruitment in the marine environment is immense, due to the inherent complexities of the biological and physical mechanisms involved, and the effects of their interaction. Because the usefulness of any model is often limited in part to the accuracy of its inputs, there is generally room for improvement. Use of the global HYCOM model remains the best solution for oceanic circulation data derived through meteorological forcing, as many global regions (e.g. Fiji and much of the Pacific), lack comprehensive and publicly available hydrodynamic datasets. For regions where finer-resolution models are available, these may be considered for use together, or in conjunction with HYCOM data, depending on the research questions involved. For example, a number of region-specific models are available for Australia, including OzROMS, SAROMS, SEAROMS, ROAM and the eReefs suite for the GBR (Joseph *et al.* 2014; Oke *et al.* 2014). There are a number of additional analysis tools available for integrating genetic and environmental datasets (see Liggins *et al.* 2013) and (Safner *et al.* 2011)), including the generation of resistance surfaces and predictive modelling to identify correlations with genetic data (e.g. (McRae 2006; Peterman 2014). These methodologies are highly useful for providing a better understanding of the processes driving genetic structure and connectivity in the marine environment, given the inherent difficulties in making empirical measurements.

Lagrangian particle models as described in chapters 3 and 4 are most accurate for simulation of passively drifting particles, which fit the known egg and larval characteristics of many marine invertebrate taxa. However, for application to organisms that possess larvae with effective swimming ability (e.g. post-flexion finfish or decapodid crustaceans), substantially more complex particle behaviour needs to be modelled. Furthermore, the addition of particle competency limitations, such as freezing particle movements after a period of time for "settlement", or reducing the total numbers of particles over the simulation to replicate larval mortality, are refinements that may become possible in the future as the toolset develops, and computational constraints are overcome through technological advances.

6.3 Conclusions

For all species exploited through capture fisheries or commercial scale aquaculture, a thorough understanding of genetic structure, diversity and population connectivity is required to inform sustainable fishery management and responsible culture practices. Wildlife management programmes, such as restocking or conservation efforts for threatened or overexploited taxa, also rely on the availability of these datasets.

The work carried out over this PhD candidature has provided comprehensive, high resolution data on both contemporary and evolutionary genetic structure in *P. margaritifera* at both local and regional scales, filling several substantial knowledge gaps for this commercially important bivalve. Apart from the provision of high quality genome-wide SNP marker sets, it also demonstrates the utility and flexibility of a robust genotyping tool for this species, which also readily lends itself to use in a wide range of other taxa, including other pearl oyster species and marine broadcast spawners.

Within the Fiji Islands, sustainable fishery management of wild populations and responsible aquaculture practices can now be realised for *P. margaritifera*, with the availability of data on genetic stock structure and population connectivity. The next steps for the Fijian pearling industry will be to develop guidelines on culture practices and translocation of both wild and farmed oysters, that will become increasingly important as existing farms grow, culture operations expand to new areas, and hatchery production increases. Given the economic importance of coloured pearl production, future research directed at examining the genetic basis of these traits, and to gauge if commercial gains may be possible through selective breeding efforts, is warranted.

More broadly at both regional and species distribution levels, this work has raised a number of questions for future research, including determination of genetic structuring within the Indian Ocean, as well as Pacific Ocean sites to better understand the drivers of population organisation. The evolutionary relationships among *P. margaritifera* populations have now been clarified using molecular data for the first time, and the

discovery that it constitutes a species complex also highlights areas for further investigation to produce a complete picture of its taxonomic identity.

In summary, the series of studies presented here represent major progress made towards the domestication of *P. margaritifera*, with respect to the genetic characterisation of populations for informing both fishery and aquaculture management. The genomic tools and data generated also lend themselves to extensive application in many other broadcast spawning marine taxa, which often despite their commercial and/or ecological importance, remain poorly understood.

REFERENCES

Adamkewicz SL, Harasewych MG (1996) Systematics and biogeography of the genus *Donax* (Bivalvia: Donacidae) in eastern North America. *American Malacological Bulletin* **13**, 97-103.

Addison JA, Hart MW (2005) Spawning, copulation and inbreeding coefficients in marine invertebrates. *Biology Letters* **1**, 450-453.

Alagarswami K, Dharmaraj S, Chellam A, Velayudhan TS (1989) Larval and juvenile rearing of black-lip pearl oyster, *Pinctada margaritifera* (Linnaeus). *Aquaculture* **76**, 43-56.

Alho JS, Välimäki K, Merilä J (2010) Rhh: an R extension for estimating multilocus heterozygosity and heterozygosity–heterozygosity correlation. *Molecular Ecology Resources* **10**, 720-722.

Allan J (1959) *Australian Shells with Related Animals Living in the Sea, in Freshwater and on the Land* Charles T. Branford Company, Boston, Massachussetts, USA.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews: Genetics* **11**, 697-709.

Anderson SC, Mills Flemming J, Watson R, Lotze HK (2011) Rapid Global Expansion of Invertebrate Fisheries: Trends, Drivers, and Ecosystem Effects. *PloS ONE* **6**, e14735.

André C, Larsson LC, Laikre L, *et al.* (2011) Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. *Heredity* **106**, 270-280.

Andrews KR, Luikart G (2014) Recent novel approaches for population genomics data analysis. *Molecular Ecology* **23**, 1661-1667.

Angeloni F, Wagemaker N, Vergeer P, Ouborg J (2012) Genomic toolboxes for conservation biologists. *Evolutionary Applications* **5**, 130-143.

Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a F_{st}-outlier method. *BMC Bioinformatics* **9**, 323.

Arjarasirikoon U, Kruatrachue M, Sretarugsa P, Chitramvong Y, Jantataeme S (2004) Gametogenic processes in the pearl oyster, *Pteria penguin* (Roding, 1798) (Bivalvia, Mollusca). *Journal of Shellfish Research* **23**, 403-410.

Arnaud-Haond S, Bonhomme F, Blanc F (2003a) Large discrepancies in differentiation of allozymes, nuclear and mitochondrial DNA loci in recently founded Pacific populations of the pearl oyster *Pinctada margaritifera*. *Journal of Evolutionary Biology* **16**, 388-398.

Arnaud-Haond S, Boudry P, Saulnier D, *et al.* (2002) New anonymous nuclear DNA markers for the pearl oyster *Pinctada margaritifera* and other *Pinctada* species. *Molecular Ecology Notes* **2**, 220-222.

Arnaud-Haond S, Vonau V, Bonhomme F, *et al.* (2004) Spatio-temporal variation in the genetic composition of wild populations of pearl oyster (*Pinctada margaritifera cumingii*) in French Polynesia following 10 years of juvenile translocation. *Molecular Ecology* **13**, 2001-2007.

Arnaud-Haond S, Vonau V, Bonhomme F, *et al.* (2003b) Spat collection of the pearl oyster (*Pinctada margaritifera cumingii*) in French Polynesia: an evaluation of the potential impact on genetic variability of wild and farmed populations after 20 years of commercial exploitation. *Aquaculture* **219**, 181-192.

Arnaud-Haond S, Vonau V, Rouxel C, *et al.* (2008) Genetic structure at different spatial scales in the pearl oyster (*Pinctada margaritifera cumingii*) in French Polynesian lagoons: beware of sampling strategy and genetic patchiness. *Marine Biology* **155**, 147-157.

Arnaud S, Monteforte M, Galtier N, Bonhomme F, Blanc F (2000) Population structure and genetic variability of pearl oyster *Pinctada mazatlanica* along Pacific coasts from Mexico to Panama. *Conservation Genetics* **1**, 299-307.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology* **22**, 3179-3190.

Atsumi T, Komaru A, Okamoto H (2004) Genetic relationship among the Japanese pearl oyster *Pinctada fucata martensii* and foreign pearl oysters. *Fish Genetics and Breeding Science* **33**, 135-142.

Babraham Bioinformatics (2007) FastQC. Babraham Institute, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Ball RD (2013) Designing a GWAS: Power, Sample Size, and Data Structure. In: *Genome-Wide Association Studies and Genomic Prediction* (eds. Gondro C, van der Werf J, Hayes B), pp. 37-97. Humana Press.

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. *Molecular Ecology* **11**, 659-674.

Bartsch P (1931) A new pearl oyster from the Hawaiian Islands. *Proceedings of the United States National Museum* **79**, 1-2.

Baum BR (1989) Phylip: Phylogeny Inference Package (Version 3.2). *The Quarterly Review of Biology* **64**, 539-541.

Beaumont AR (1991) Genetic studies of laboratory reared mussels, *Mytilus edulis*: heterozygote deficiencies, heterozygosity and growth. *Biological Journal of the Linnean Society* **44**, 273-285.

Beer AC, Southgate PC (2000) Collection of pearl oyster (family Pteriidae) spat at Orpheus Island, Great Barrier Reef (Australia). *Journal of Shellfish Research* **19**, 821-826.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996) *GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations.* Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier, France. http://www.genetix.univ-montp2.fr/genetix/intro.htm

Benzie JAH, Ballment E (1994) Genetic differences among black-lipped pearl oyster (*Pinctada margaritifera*) populations in the western Pacific. *Aquaculture* **127**, 145-156.

Berry O, England P, Marriott RJ, Burridge CP, Newman SJ (2012) Understanding age-specific dispersal in fishes through hydrodynamic modelling, genetic simulations and microsatellite DNA analysis. *Molecular Ecology* **21**, 2145-2159.

Bourret V, Kent MP, Primmer CR, *et al.* (2013) SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology* **22**, 532-551.
Broquet T, Viard F, Yearsley JM (2013) Genetic drift and collective dispersal can result in chaotic genetic patchiness. *Evolution* **67**, 1660-1675.
Brownlow RJ, Dawson DA, Horsburgh GJ, Bell JJ, Fish JD (2008) A method for genotype validation and primer assessment in heterozygote-deficient species, as demonstrated in the prosobranch mollusc *Hydrobia ulvae*. *BMC Genetics* **9**, 55-55.
Bruno JF, Selig ER (2007) Regional Decline of Coral Cover in the Indo-Pacific: Timing, Extent, and Subregional Comparisons. *PloS ONE* **2**.
Brussard PF (1984) Geographic Patterns and Environmental Gradients: The Central-Marginal Model in *Drosophila* Revisited. *Annual Review of Ecology and Systematics* **15**, 25-64.
Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. *Molecular Biology and Evolution*.
Cahn A (1949) Pearl culture in Japan (ed. Service FaW), pp. 1-91. United States Department of the Interior.
Cannuel R, Beninger PG, McCombie H, Boudry P (2009) Gill Development and Its Functional and Evolutionary Implications in the Blue Mussel *Mytilus edulis* (Bivalvia: Mytilidae). *Biological Bulletin* **217**, 173-188.
Carlson CS, Smith JD, Stanaway IB, Rieder MJ, Nickerson DA (2006) Direct detection of null alleles in SNP genotyping data. *Human Molecular Genetics* **15**, 1931-1937.
Carpenter KE (1998) An introduction to the oceanography, geology, biogeography, and fisheries of the tropical and subtropical western and central Pacific. In: *FAO Species Identification Guide for Fishery Purposes. The Living Marine Resources of the Western Central Pacific* (eds. Carpenter KE, Niem VH), pp. 1-19. Food and Agriculture Organization of the United Nations (FAO), South Pacific Forum Fisheries Agency (FFA) and Norwegian Agency for International Development (NORAD). Rome, Italy.
Carvalho GR, Hauser L (1994) Molecular genetics and the stock concept in fisheries. *Reviews in Fish Biology and Fisheries* **4**, 326-350.
Catchen J, Bassham S, Wilson T, *et al.* (2013) The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular Ecology* **22**, 2864-2883.
Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes, Genomes, Genetics* **1**, 171-182.
Chassignet EP, Hurlburt HE, Smedstad OM, *et al.* (2007) The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system. *Journal of Marine Systems* **65**, 60-83.
Coeroli M, De Gaillande D, Landret JP, Aquacop (1984) Recent innovations in cultivation of molluscs in French Polynesia. *Aquaculture* **39**, 45-67.
Cotter AJR, Burt L, Paxton CGM, *et al.* (2004) Are stock assessment methods too complicated? *Fish and Fisheries* **5**, 235-254.

Cruz VMV, Kilian A, Dierig DA (2013) Development of DArT Marker Platforms and Genetic Diversity Assessment of the U.S. Collection of the New Oilseed Crop *Lesquerella* and Related Species. *PloS ONE* **8**, e64062.
Cummings JA (2005) Operational multivariate ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society* **131**, 3583-3604.
Cunha RL, Blanc F, Bonhomme F, Arnaud-Haond S (2011) Evolutionary Patterns in Pearl Oysters of the Genus *Pinctada* (Bivalvia: Pteriidae). *Marine Biotechnology* **13**, 181-192.
Curini-Galletti M, Puccinelli I (1998) The *Gyratrix hermaphroditus* species complex (Kalyptorhynchia: Polycystididae) in marine habitats of eastern Australia. *Hydrobiologia* **383**, 287-298.
Dąbrowski MJ, Pilot M, Kruczyk M, *et al.* (2014) Reliability assessment of null allele detection: inconsistencies between and within different methods. *Molecular Ecology Resources* **14**, 361-373.
Dao HT, Smith-Keune C, Wolanski E, Jones CM, Jerry DR (2015) Oceanographic Currents and Local Ecological Knowledge Indicate, and Genetics Does Not Refute, a Contemporary Pattern of Larval Dispersal for The Ornate Spiny Lobster, *Panulirus ornatus* in the South-East Asian Archipelago. *PloS ONE* **10**.
Davey JW, Blaxter ML (2011) RADSeq: next-generation population genetics. *Briefings in Functional Genomics* **9**, 416-423.
Davey JW, Hohenlohe PA, Etter PD, *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews: Genetics* **12**, 499-510.
Dawson MN (2005) Incipient speciation of *Catostylus mosaicus* (Scyphozoa, Rhizostomeae, Catostylidae), comparative phylogeography and biogeography in south-east Australia. *Journal of Biogeography* **32**, 515-533.
Dawson MN, Raskoff KA, Jacobs DK (1998) Field preservation of marine invertebrate tissue for DNA analyses. *Molecular Marine Biology and Biotechnology* **7**, 145-152.
Do C, Waples RS, Peel D, *et al.* (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources* **14**, 209-214.
Domingos JA, Smith-Keune C, Robinson N, *et al.* (2013) Heritability of harvest growth traits and genotype–environment interactions in barramundi, *Lates calcarifer* (Bloch). *Aquaculture* **402–403**, 66-75.
Doroudi MS, Southgate PC (2002) The effect of chemical cues on settlement behaviour of blacklip pearl oyster (*Pinctada margaritifera*) larvae. *Aquaculture* **209**, 117-124.
Doroudi MS, Southgate PC (2003) Embryonic and larval development of *Pinctada margaritifera* (Linnaeus, 1758). *Molluscan Research* **23**, 101-107.
Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 1-8.
Durand P, Blanc F (1988) Genetic diversity in a tropical marine bivalve - *Pinctada margaritifera* (Linne, 1758). *Bulletin De La Societe Zoologique De France - Evolution Et Zoologie* **113**, 293-304.
Durand P, Wada KT, Blanc F (1993) Genetic variation in wild and hatchery stocks of the black pearl oyster, *Pinctada margaritifera*, from Japan. *Aquaculture* **110**, 27-40.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology* **17**, 1170-1188.
Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.
Elamin EM, Elamin SEM (2014) Biometric relationships of the mother of pearl oyster (*Pinctada margaritifera* var *erythraensis*) from Dongonab Bay, Red Sea. *International Journal of Science, Environment* **3**, 1193-1204.
Elshire RJ, Glaubitz JC, Sun Q, *et al.* (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PloS ONE* **6**, e19379.
Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.
FAO (2015) The State of World Fisheries and Aquaculture 2014: Opportunities and challenges. In: *The State of World Fisheries and Aquaculture (SOFIA)* (ed. FAO), p. 243. Food and Agriculture Organisation (FAO) of the United Nations, Rome, Italy.
Foll M (2012) BayeScan v2.1 User Manual. *Ecology* **20**, 1450-1462.
Foll M, Gaggiotti O (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **180**, 977-993.
Friedman K, Kronen M, Vunisea A, *et al.* (2010) Fiji Islands Country Report: Profiles and Results from Survey Work at Dromuna, Muaivuso, Mali and Lakeba. (September to November 2002, April to June 2003, June and July 2007, and February 2009). In: *Pacific Regional Oceanic and Coastal Fisheries Development Programme (PROCFish/C/CoFish)*, p. 503. Secretariat of the Pacific Community, Noumea, New Caledonia.
Friedman KJ, Bell JD (2000) Shorter immersion times increase yields of the blacklip pearl oyster, *Pinctada margaritifera* (Linne.), from spat collectors in Solomon Islands. *Aquaculture* **187**, 299-313.
Friedman KJ, Southgate PC (1999) Growout of Blacklip Pearl Oysters, *Pinctada margaritifera* collected as wild spat in the Solomon Islands. *Journal of Shellfish Research* **18**, 159-167.
Fromentin J-M, Bonhommeau S, Arrizabalaga H, Kell LT (2014) The spectre of uncertainty in management of exploited fish stocks: The illustrative case of Atlantic bluefin tuna. *Marine Policy* **47**, 8-14.
Frost LA, Evans BS, Jerry DR (2007) Loss of genetic diversity due to hatchery culture practices in barramundi (*Lates calcarifer*). *Aquaculture* **272, Supplement 1**, S272.
Funabara D, Watanabe D, Satoh N, Kanoh S (2013) Genome-wide survey of genes encoding muscle proteins in the pearl oyster, *Pinctada fucata*. *Zoolog Sci* **30**.
Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution* **27**, 489-496.
Gaggiotti OE, Bekkevold D, Jørgensen HBH, *et al.* (2009) Disentangling the Effects of Evolutionary, Demographic, and Environmental Factors Influencing Genetic Structure of Natural Populations: Atlantic Herring as a Case Study. *Evolution* **63**, 2939-2951.

Galindo HM, Olson DB, Palumbi SR (2006) Seascape Genetics: A Coupled Oceanographic-Genetic Model Predicts Population Structure of Caribbean Corals. *Current Biology* **16**, 1622-1626.
Galstoff PS (1933) *Pearl and Hermes Reef, Hawaii: hydrographical and biological observations* Bernice P. Bishop Museum, Honolulu, Hawaii.
Ganachaud A, Kessler W, Wijffels S, *et al.* (2007) Southwest Pacific Ocean Circulation and Climate Experiment (SPICE): Part I Scientific Background. . In: *CLIVAR Publication Series*, *NOAA OAR Special Report*, pp. 1-46. NOAA/OAR/PMEL, Seattle, Washington, USA.
GE (2007) Illustra AutoSeq G-50 and AutoSeq 96 dye terminator removal. Data file 28-9175-28. In: *Illustra AutoSeq G-50 documents*, pp. 1-4. GE Healthcare UK Limited, Buckinghamshire, United Kingdom.
Gervis MH, Sims NA (1992) *The Biology and Culture of Pearl Oysters (Bivalvia: Pteriidae)* ICLARM, Manila, Philippines.
Godhe A, Egardt J, Kleinhans D, *et al.* (2013) Seascape analysis reveals regional gene flow patterns among populations of a marine planktonic diatom. *Proceedings of the Royal Society of London B: Biological Sciences* **280**.
Gogarten SM, Bhangale T, Conomos MP, *et al.* (2012) GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. *Bioinformatics* **28**, 3329-3331.
Gohli J, Leder EH, Garcia-del-Rey E, *et al.* (2015) The evolutionary history of Afrocanarian blue tits inferred from genomewide SNPs. *Molecular Ecology* **24**, 180-191.
Gosling EM (1982) Genetic Variability in Hatchery-Produced Pacific Oysters (*Crassostrea gigas* Thunberg). *Aquaculture* **26**, 273-287.
Gosling EM (2015) Fisheries and management of natural populations. In: *Marine Bivalve Molluscs* (ed. Gosling EM), pp. 270-324. Wiley Blackwell, West Sussex, United Kingdom.
Gosling EM, Wilkins NP (1985) Genetics of settling cohorts of *Mytilus edulis* (L.): Preliminary observations. *Aquaculture* **44**, 115-123.
Grosberg RK, Cunningham CW (2001) Genetic structure in the sea: from populations to communities. In: *Marine Community Ecology* eds. Bertness MD, Gaines SD, Hay ME), pp. 61-84. Sinauer Associates, Sunderland, Massachussetts, USA.
Gug M (1957) Mother-of-pearl industry in French Oceania. In: *SPC Quarterly Bulletin*, pp. 19-21. South Pacific Commission, Noumea, New Caledonia.
Guo Q (2012) Incorporating latitudinal and central-marginal trends in assessing genetic variation across species ranges. *Molecular Ecology* **21**, 5396-5403.
Halliwell GR (2004) Evaluation of vertical coordinate and vertical mixing algorithms in the HYbrid-Coordinate Ocean Model (HYCOM). *Ocean Modelling* **7**, 285-322.
Hanley SCT, Wood W (1842) *An illustrated and descriptive catalogue of recent bivalve shells* Williams and Norgate, London, United Kingdom.
Hardie DC, Hutchings JA (2010) Evolutionary ecology at the extremes of species' ranges. *Environmental Reviews* **18**, 1-20.
Hare PM, Palumbi RS, Butman AC (2000) Single-step species identification of bivalve larvae using multiplex polymerase chain reaction. *Marine Biology* **137**, 953-961.
Harrang E, Lapègue S, Morga B, Bierne N (2013) A High Load of Non-neutral Amino-Acid Polymorphisms Explains High Protein Diversity Despite Moderate Effective

Population Size in a Marine Bivalve With Sweepstakes Reproduction. *G3: Genes|Genomes|Genetics* **3**, 333-341.
Hauser L, Baird M, Hilborn RAY, Seeb LW, Seeb JE (2011) An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular Ecology Resources* **11**, 150-161.
Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries* **9**, 333-362.
Hayami I (2000) Family Pteriidae Order Pterioida. In: *Marine Mollusks in Japan* (ed. Okutani T), pp. 879-883. Tokai University Press, Tokyo, Japan.
Hayes B (2013) Overview of Statistical Methods for Genome-Wide Association Studies (GWAS). In: *Genome-Wide Association Studies and Genomic Prediction* (eds. Gondro C, van der Werf J, Hayes B), pp. 149-170. Humana Press.
He M, Guan Y, Yuan T, Zhang H (2008) Realized heritability and response to selection for shell height in the pearl oyster *Pinctada fucata* (Gould). *Aquaculture Research* **39**, 801-805.
Hedgecock D, Gaffney PM, Goulletquer P, *et al.* (2005) The case for sequencing the Pacific Oyster genome. *Journal of Shellfish Research* **24**, 429-441.
Hedgecock D, Li G, Hubert S, Bucklin K, Ribes V (2004) Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific Oyster, *Crassostrea gigas*. *Journal of Shellfish Research* **23**, 379-385.
Hedley C (1924) Australian pearl fisheries. In: *The Australian Museum Magazine*, pp. 5-11. The Australian Museum, Australia.
Hellberg ME, Burton RS, Neigel JE, Palumbi SR (2002) Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science* **70**, 273-290.
Helyar SJ, Hemmer-Hansen J, Bekkevold D, *et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* **11**, 123-136.
Hemmer-Hansen J, Nielsen EE, Grønkjær P, Loeschcke V (2007) Evolutionary mechanisms shaping the genetic population structure of marine fishes; lessons from the European flounder (*Platichthys flesus* L.). *Molecular Ecology* **16**, 3104-3118.
Hessel P (nd) Pteria penguin (180mm), colno 34008. www.fossilshells.nl, online.
Hoffman JI, Peck LS, Linse K, Clarke A (2011) Strong Population Genetic Structure in a Broadcast-Spawning Antarctic Marine Invertebrate. *Journal of Heredity* **102**, 55-66.
Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources* **11**, 117-122.
Huang H, Knowles LL (2014) Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation Study of RAD Sequences. *Systematic Biology*.
Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.
Huson DH, Richter DC, Rausch C, *et al.* (2007) Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* **8**, 1-6.

Hwang JJ, Okutani T (2003) Taxonomy and Distribution of the Genera *Pteria* and *Pinctada* (Bivalvia: Pteriidae) in Taiwan. *Journal of the Fisheries Society of Taiwan* **30**, 199-216.
Hynd JS (1955) A Revision of the Australian Pearl-shells, Genus *Pinctada* (Lamellibranchia). *Marine and Freshwater Research* **6**, 98-138.
Jabbour-Zahab R, Chagot D, Blanc F, Grizel H (1992) Mantle histology, histochemistry and ultrastructure of the pearl oyster *Pinctada margaritifera* (L.). *Aquatic Living Resources* **5**, 287-298.
Jackson TR, Martin-Robichaud DJ, Reith ME (2003) Application of DNA markers to the management of Atlantic halibut (*Hippoglossus hippoglossus*) broodstock. *Aquaculture* **220**, 245-259.
Jameson LH (1901) On the Identity and Distribution of the Mother-of-Pearl Oysters; with a Revision of the Subgenus *Margaritifera*. *Proceedings of the Zoological Society of London* **70**, 372-394.
Jerry DR, Kvingedal R, Lind CE, *et al.* (2012) Donor-oyster derived heritability estimates and the effect of genotype × environment interaction on the production of pearl quality traits in the silver-lip pearl oyster, *Pinctada maxima*. *Aquaculture* **338–341**, 66-71.
Johnson M, Zaretskaya I, Raytselis Y, *et al.* (2008) NCBI BLAST: a better web interface. *Nucleic Acids Research* **36**, W5-W9.
Johnston SE, Orell P, Pritchard VL, *et al.* (2014) Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (Salmo salar). *Molecular Ecology* **23**, 3452-3468.
Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405.
Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*.
Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**, 94.
Jones D, Jerry D, Khatkar M, Raadsma H, Zenger K (2013a) A high-density SNP genetic linkage map for the silver-lipped pearl oyster, *Pinctada maxima*: a valuable resource for gene localisation and marker-assisted selection. *BMC Genomics* **14**, 810.
Jones DB, Jerry DR, Forêt S, Konovalov DA, Zenger KR (2013b) Genome-Wide SNP Validation and Mantle Tissue Transcriptome Analysis in the Silver-Lipped Pearl Oyster, *Pinctada maxima*. *Marine Biotechnology* **15**, 647-658.
Jones DB, Jerry DR, Khatkar MS, *et al.* (2014a) Determining genetic contributions to host oyster shell growth: Quantitative trait loci and genetic association analysis for the silver-lipped pearl oyster, *Pinctada maxima*. *Aquaculture* **434**, 367-375.
Jones DB, Jerry DR, Khatkar MS, *et al.* (2014b) Quantitative trait loci and genetic association analysis reveals insights into complex pearl quality traits in donor silver-lipped pearl oysters. *Aquaculture* **434**, 476-485.
Jones GP, Almany GR, Russ GR, *et al.* (2009) Larval retention and connectivity among populations of corals and reef fishes: history, advances and challenges. *Coral Reefs* **28**, 307-325.

Joseph E, Pattiaratchi C, Wijeratne S (2014) Simulating the dynamics of continental shelf waves generated by tropical cyclones in Western Australia. In: *Australian Coastal and Ocean Modelling and Observations Workshop (ACOMO) 2014*, pp. 295-301, 7-8th October 2014, Shine Dome, Australian Academy of Sciences, Canberra

Kai W, Nomura K, Fujiwara A, *et al.* (2014) A ddRAD-based genetic map and its integration with the genome assembly of Japanese eel (*Anguilla japonica*) provides insights into genome evolution after the teleost-specific genome duplication. *BMC Genomics* **15**, 233.

Kalinowski ST (2002) Evolutionary and statistical properties of three genetic distances. *Molecular Ecology* **11**, 1263-1273.

Kalinowski ST (2004) Counting Alleles with Rarefaction: Private Alleles and Hierarchical Sampling Designs. *Conservation Genetics* **5**, 539-543.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16**, 1099-1106.

Kalinowski ST, Wagner AP, Taper ML (2006) ML-RELATE: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes* **6**, 576-579.

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**, e281.

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* **4**, 782-788.

Kilian A, Wenzl P, Huttner E, *et al.* (2012) Diversity arrays technology: A generic genome profiling technology on open platforms. In: *Methods in Molecular Biology*, pp. 67-89.

Knowlton N (1993) Sibling Species in the Sea. *Annual Review of Ecology and Systematics* **24**, 189-216.

Knutsen H, Olsen EM, Jorde PE, Espeland SH (2011) Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. *Molecular Ecology* **20**, 768.

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. *Molecular Ecology* **17**, 3775-3787.

Koopman WJM, Wissemann V, De Cock K, *et al.* (2008) AFLP markers as a tool to reconstruct complex relationships: A case study in *Rosa* (Rosaceae). *American Journal of Botany* **95**, 353-366.

Krück NC, Innes DI, Ovenden JR (2013) New SNPs for population genetic analysis reveal possible cryptic speciation of eastern Australian sea mullet (*Mugil cephalus*). *Molecular Ecology Resources* **13**, 715-725.

Kvingedal R, Evans BS, Lind CE, *et al.* (2010) Population and family growth response to different rearing location, heritability estimates and genotype × environment interaction in the silver-lip pearl oyster (*Pinctada maxima*). *Aquaculture* **304**, 1-6.

Kvingedal R, Evans BS, Taylor JJU, Knauer J, Jerry DR (2008) Family by environment interactions in shell size of 43-day old silver-lip pearl oyster (*Pinctada maxima*), five families reared under different nursery conditions. *Aquaculture* **279**, 23-28.

Lal MM, Southgate PC, Jerry DR, Bosserelle C, Zenger KR (2016a) A parallel population genomic and hydrodynamic approach to fishery management of highly-dispersive marine invertebrates: the case of the Fijian black-lip pearl oyster *Pinctada margaritifera*. *PloS ONE*.

Lal MM, Southgate PC, Jerry DR, Bosserelle C, Zenger KR (unpubl.) Ocean currents and seascape features regulate genetic structure in a broadcast spawning marine invertebrate: an examination of the black-lip pearl oyster *Pinctada margaritifera* across its 18,000 Km Indo-Pacific distribution.

Lal MM, Southgate PC, Jerry DR, Zenger KR (2016b) Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. *Marine Genomics* **25**, 57-68.

Lam K, Morton B (2003) Morphological and ITS1, 5.8S, and Partial ITS2 Ribosomal DNA Sequence Distinctions Between Two Species *Playtygyra* (Cnidaria: Scleractinia) from Hong Kong. *Marine Biotechnology* **5**, 555-567.

Laugen AT, Engelhard GH, Whitlock R, *et al.* (2014) Evolutionary impact assessment: accounting for evolutionary consequences of fishing in an ecosystem approach to fisheries management. *Fish and Fisheries* **15**, 65-96.

Leaché AD, Banbury BL, Felsenstein J, de Oca An-M, Stamatakis A (2015) Short Tree, Long Tree, Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies. *Systematic Biology* **64**, 1032-1047.

Lee T-H, Guo H, Wang X, Kim C, Paterson AH (2014) SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* **15**, 1-6.

Lemer S, Planes S (2012) Translocation of wild populations: conservation implications for the genetic diversity of the black-lipped pearl oyster *Pinctada margaritifera*. *Molecular Ecology* **21**, 2949-2962.

Lemer S, Planes S (2014) Effects of habitat fragmentation on the genetic structure and connectivity of the black-lipped pearl oyster *Pinctada margaritifera* populations in French Polynesia. *Marine Biology* **161**, 2035-2049.

Lemer S, Rochel E, Planes S (2011) Correction Method for Null Alleles in Species with Variable Microsatellite Flanking Regions, A Case Study of the Black-Lipped Pearl Oyster *Pinctada margaritifera*. *Journal of Heredity* **102**, 243-246.

Lepais O, Weir JT (2014) SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. *Molecular Ecology Resources* **14**, 1314-1321.

Lewison RL, Soykan CU, Franklin J (2009) Mapping the bycatch seascape: multispecies and multi-scale spatial patterns of fisheries bycatch. *Ecological Applications* **19**, 920-930.

Liggins L, Booth DJ, Figueira WF, *et al.* (2015) Latitude-wide genetic patterns reveal historical effects and contrasting patterns of turnover and nestedness at the range peripheries of a tropical marine fish. *Ecography* **38**, 1212-1224.

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, southwest Pacific. *Bulletin of Marine Science* **90**, 379-397.

Liggins L, Treml EA, Possingham HP, Riginos C (2016) Seascape features, rather than dispersal traits, predict spatial genetic patterns in co-distributed reef fishes. *Journal of Biogeography* **43**, 256-267.

Liggins L, Treml EA, Riginos C (2013) Taking the Plunge: An Introduction to Undertaking Seascape Genetic Studies and using Biophysical Models. *Geography Compass* **7**, 173-196.
Limborg MT, Helyar SJ, De Bruyn M, *et al.* (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology* **21**, 3686-3703.
Lind CE, Evans BS, Knauer J, Taylor JJU, Jerry DR (2009) Decreased genetic diversity and a reduced effective population size in cultured silver-lipped pearl oysters (*Pinctada maxima*). *Aquaculture* **286**, 12-19.
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. *Molecular Ecology* **16**, 5193-5203.
Lind CE, Evans BS, Taylor JJU, Jerry DR (2010) The consequences of differential family survival rates and equalizing maternal contributions on the effective population size (Ne) of cultured silver-lipped pearl oysters, *Pinctada maxima*. *Aquaculture Research* **41**, 1229-1242.
Lucas-Lledó JI, Vicente-Salvador D, Aguado C, Cáceres M (2014) Population genetic analysis of bi-allelic structural variants from low-coverage sequence data with an expectation-maximization algorithm. *BMC Bioinformatics* **15**, 1-13.
Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Molecular Ecology* **3**, 91-99.
Markey KL, Abdo DA, Evans SN, Bosserelle C (2016) Keeping It Local: Dispersal Limitations of Coral Larvae to the High Latitude Coral Reefs of the Houtman Abrolhos Islands. *PloS ONE* **11**.
Marques DA, Lucek K, Meier JI, *et al.* (2016) Genomics of Rapid Incipient Speciation in Sympatric Threespine Stickleback. *PloS Genetics* **12**, e1005887.
Masaoka T, Kobayashi T (2005a) Natural hybridization between *Pinctada fucata* and *Pinctada maculata* inferred from internal transcribed spacer regions of nuclear ribosomal RNA genes. *Fisheries Science* **71**, 829-836.
Masaoka T, Kobayashi T (2005b) Species identification of *Pinctada imbricata* using intergenic spacer of nuclear ribosomal RNA genes and mitochondrial 16S ribosomal RNA gene regions. *Fisheries Science* **71**, 837-846.
Mateos JM, Pérez JP (2013) *Image Processing with ImageJ* Packt Publishing Ltd, Olton, Birmingham, GBR.
Matsukuma A (2004) Family Pteriidae Order Pterioida. In: *Encyclopedia of Shellfish* (ed. Okutani T), p. 288. Sekaibunkasha, Tokyo, Japan.
McAndrew B, Napier J (2011) Application of genetics and genomics to aquaculture development: current and future directions. *The Journal of Agricultural Science* **149**, 143-151.
McGinty EL, Zenger KR, Jones DB, Jerry DR (2012) Transcriptome analysis of biomineralisation-related genes within the pearl sac: Host and donor oyster contribution. *Marine Genomics* **5**, 27-33.
McRae BH (2006) Isolation by Resistance. *Evolution* **60**, 1551-1561.
Miller AD, Versace VL, Matthews TG, Montgomery S, Bowie KC (2013) Ocean currents influence the genetic structure of an intertidal mollusc in southeastern Australia –

implications for predicting the movement of passive dispersers across a marine biogeographic barrier. *Ecology and Evolution* **3**, 1248-1261.
Miller MR, Brunelli JP, Wheeler PA, *et al.* (2012) A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular Ecology* **21**, 237-249.
Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza LL (1995) MICROSAT Version 1.4d: a computer program for calculating various statistics on microsatellite allele data. University of Stanford, Stanford, California, U.S.A.
Miyamoto H, Endo H, Hashimoto N, *et al.* (2013) The diversity of shell matrix proteins: genome-wide investigation of the pearl oyster, *Pinctada fucata*. *Zoolog Sci* **30**.
Mohamed KS, Kripa V, Velayudhan TS, Appukuttan KK (2006) Growth and biometric relationships of the pearl oyster *Pinctada fucata* (Gould) on transplanting from the Gulf of Mannar to the Arabian Sea. *Aquaculture Research* **37**, 725-741.
Nagai K, Yano M, Morimoto K, Miyamoto H (2007) Tyrosinase localization in mollusc shells. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **146**, 207-214.
Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11**, 184-194.
Nayfa MG, Zenger KR (2016) Unravelling the effects of gene flow and selection in highly connected populations of the silver-lip pearl oyster (*Pinctada maxima*). *Marine Genomics*.
Nei M (1972) Genetic Distance between Populations. *The American Naturalist* **106**, 283-292.
Nei M (1987) Genetic distance and molecular phylogeny. In: *Population Genetics and Fishery Management* (eds. Ryman N, Utter F), pp. 193-223. University of Washington Press, Seattle, United States of America.
Neo ML, Erftemeijer PLA, Beek KL, *et al.* (2013) Recruitment constraints in Singapore's fluted giant clam (*Tridacna squamosa*) population--A dispersal model approach. *PloS ONE* **8**.
Neuditschko M, Khatkar MS, Raadsma HW (2012) NETVIEW: A High-Definition Network-Visualization Approach to Detect Fine-Scale Population Structures from Genome-Wide Patterns of Variation. *PloS ONE* **7**, e48375.
Nielsen EE, Hemmer-Hansen J, Poulsen NA, *et al.* (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology* **9**, 276-276.
Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Reviews: Genetics* **12**, 443-451.
Noble TH, Smith-Keune C, Jerry DR (2014) Genetic investigation of the large-scale escape of a tropical fish, barramundi *Lates calcarifer*, from a sea-cage facility in northern Australia. *Aquaculture Environment Interactions* **5**, 173-183.
Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**, 375–402.
Nosil P, Vines TH, Funk DJ (2005) Perspective: Reproductive Isolation Caused by Natural Selection against Immigrants from Divergent Habitats. *Evolution* **59**, 705-719.
Oke PR, Roughan M, Brinkman R, Baird ME (2014) Coastal and shelf oceanography in Australia, pp. 1-8. National Marine Science Committee, Online.

Ovenden JR, Berry O, Welch DJ, Buckworth RC, Dichmont CM (2015) Ocean's eleven: a critical evaluation of the role of population, evolutionary and molecular genetics in the management of wild fisheries. *Fish and Fisheries* **16**, 125-159.
Palumbi SR (1994) Genetic Divergence, Reproductive Isolation, and Marine Speciation. *Annual Review of Ecology and Systematics* **25**, 547-572.
Palumbi SR (2003) Population Genetics, Demographic Connectivity, and the Design of Marine Reserves. *Ecological Applications* **13**, S146-S158.
Park J-M, Kim G-Y, Lee S-J, *et al.* (2006) Comparison of RAPD, AFLP, and EF-1α Sequences for the Phylogenetic Analysis of *Fusarium oxysporum* and Its formae speciales in Korea. *Mycobiology* **34**, 45-55.
Passfield K (1995) Report of a pearl oyster survey of W. Vanua Levu, Beqa, Totoya and Makogai, in the Fiji Islands (Draft), pp. 1-24. Fiji Fisheries Division, Suva, Fiji Islands.
Patterson K, Cook R, Darby C, *et al.* (2001) Estimating uncertainty in fish stock assessment and forecasting. *Fish and Fisheries* **2**, 125-157.
Pauly D, Christensen V, Guenette S, *et al.* (2002) Towards sustainability in world fisheries. *Nature* **418**, 689-695.
Pauly D, Zeller D (2016) Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining. *Nat Commun* **7**.
Peakall ROD, Smouse PE (2006) GENEALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.
Pechenik JA (2006) Larval Experience and Latent Effects: Metamorphosis Is Not a New Beginning. *Integrative and Comparative Biology* **46**, 323-333.
Peñaloza C, Bishop SC, Toro J, Houston RD (2014) RAD Sequencing reveals genome-wide heterozygote deficiency in pair crosses of the Chilean mussel *Mytilus* spp. In: *10th World Congress on Genetics Applied to Livestock Production*, pp. 1-3, Vancouver, British Columbia, Canada.
Peterman WE (2014) ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *bioRxiv*.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PloS ONE* **7**, e37135.
Pit JH, Southgate PC (2003) Fouling and predation; how do they affect growth and survival of the blacklip pearl oyster, *Pinctada margaritifera*, during nursery culture? *Aquaculture International* **11**, 545-555.
Plutchak LL, Simmons RE, Woodruff DS (2006) Multilocus allozyme heterozygote deficiencies in *Crepidula onyx*: geographic and temporal patterns among adult snails in Mission Bay, California. *Journal of Molluscan Studies* **72**, 337-348.
Ponia B (2010) A review of aquaculture in the Pacific Islands 1998 - 2007: Tracking a decade of progress through official and provisional statistics. In: *Aquaculture Technical Papers*, p. 38. Secretariat of the Pacific Community, Noumea, New Caledonia.
Pons M, Branch TA, Melnychuk MC, *et al.* (2016) Effects of biological, economic and management factors on tuna and billfish stock status. *Fish and Fisheries*.
Pool JE, Hellmann I, Jensen JD, Nielsen R (2010) Population genetic inference from genomic sequence variation. *Genome Research* **20**, 291-300.

Pouvreau S, Prasil V (2001) Growth of the black-lip pearl oyster, *Pinctada margaritifera*, at nine culture sites of French Polynesia: synthesis of several sampling designs conducted between 1994 and 1999. *Aquatic Living Resources* **14**, 155-163.

Pouvreau S, Tiapari J, Gangnery A, *et al.* (2000) Growth of the black-lip pearl oyster, *Pinctada margaritifera*, in suspended culture under hydrobiological conditions of Takapoto lagoon (French Polynesia). *Aquaculture* **184**, 133-154.

Przyborowski JA, Sulima P, Kuszewska A, Zaluski D, Kilian A (2013) Phylogenetic Relationships between Four *Salix* L. Species Based on DArT Markers. *International Journal of Molecular Sciences* **14**, 24113-24125.

Pujolar JM, Bevacqua D, Andrello M, *et al.* (2011) Genetic patchiness in European eel adults evidenced by molecular genetics and population dynamics modelling. *Molecular Phylogenetics and Evolution* **58**, 198-206.

Pujolar JM, Jacobsen MW, Als TD, *et al.* (2014) Genome-wide single-generation signatures of local selection in the panmictic European eel. *Molecular Ecology* **23**, 2514-2528.

Purcell S, Neale B, Todd-Brown K, *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575.

Puritz JB, Matz MV, Toonen RJ, *et al.* (2014) Demystifying the RAD fad. *Molecular Ecology*, n/a-n/a.

Rambaut A (2014) *FigTree v.1.4.2: tree drawing tool.* Andrew Rambaut. http://tree.bio.ed.ac.uk/software/figtree/

Rambaut A, Suchard M, Drummond AJ (2003) *Tracer v.1.6.* http://tree.bio.ed.ac.uk/software/tracer/

Ranjbar MS, Zolgharnien H, Yavari V, *et al.* (2016) Rising the Persian Gulf Black-Lip Pearl Oyster to the Species Level: Fragmented Habitat and Chaotic Genetic Patchiness in *Pinctada persica*. *Evolutionary Biology* **43**, 131-143.

Ranson G (1961) *Missions dans le Pacifique récifs coralliens, huîtres perlières* Lechevalier, Paris, France.

Rasic G, Filipovic I, Weeks A, Hoffmann A (2014) Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, *Aedes aegypti*. *BMC Genomics* **15**, 275.

Reiss H, Hoarau G, Dickey-Collas M, Wolff WJ (2009) Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish and Fisheries* **10**, 361-395.

Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM (2013) Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology* **22**, 2953-2970.

Ren R, Ray R, Li P, *et al.* (2015) Construction of a high-density DArTseq SNP-based genetic map and identification of genomic regions with segregation distortion in a genetic population derived from a cross between feral and cultivated-type watermelon. *Molecular Genetics and Genomics* **290**, 1457-1470.

Richards ZT, Berry O, van Oppen MJH (2016) Cryptic genetic divergence within threatened species of *Acropora* coral from the Indian and Pacific Oceans. *Conservation Genetics* **17**, 577-591.

Riginos C, Cunningham CW (2005) INVITED REVIEW: Local adaptation and species segregation in two mussel (*Mytilus edulis* × *Mytilus trossulus*) hybrid zones. *Molecular Ecology* **14**, 381-400.
Robasky K, Lewis NE, Church GM (2014) The role of replicates for error mitigation in next-generation sequencing. *Nat Rev Genet* **15**, 56-62.
Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.
Ronquist F, Teslenko M, van der Mark P, *et al.* (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology* **61**, 539-542.
Rose RA (1990) *A manual for the artificial propagation of the silverlip or goldlip pearl oyster, Pinctada maxima (Jameson) from Western Australia.* .
Rose RA, Baker SB (1994) Larval and spat culture of the Western Australian silver- or goldlip pearl oyster, *Pinctada maxima* Jameson (Mollusca: Pteriidae). *Aquaculture* **126**, 35-50.
Rotterdam NM (2013) Pinctada imbricata. Natural History Museum Rotterdam, online.
Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.
Safner T, Miller MP, McRae BH, Fortin M-J, Manel S (2011) Comparison of Bayesian Clustering and Edge Detection Methods for Inferring Boundaries in Landscape Genetics. *International Journal of Molecular Sciences* **12**, 865-889.
Sanciangco JC, Carpenter KE, Etnoyer PJ, Moretzsohn F (2013) Habitat Availability and Heterogeneity and the Indo-Pacific Warm Pool as Predictors of Marine Species Richness in the Tropical Indo-Pacific. *PloS ONE* **8**, e56245.
Sansaloni C, Petroli C, Jaccoud D, *et al.* (2011) Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. *BMC Proceedings* **5**, P54-P54.
Saucedo PE, Southgate PC (2008) Reproduction, Development and Growth. In: *The Pearl Oyster* (eds. Southgate PC, Lucas JS), pp. 133-186. Elsevier, Amsterdam, the Netherlands.
Saville-Kent W (1905) Torres Straits pearlshell fisheries. In: *Queensland parliamentary papers: Session 2 of 1905. Vol 2. Report to both Houses of Parliament*, pp. 1075-1078, Brisbane, Australia.
Schliep KP (2011) phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592-593.
Sekino M, Saitoh K, Yamada T, *et al.* (2003) Microsatellite-based pedigree tracing in a Japanese flounder *Paralichthys olivaceus* hatchery strain: implications for hatchery management related to stock enhancement program. *Aquaculture* **221**, 255-263.
Sexton JP, McIntyre PJ, Angert AL, Rice KJ (2009) Evolution and Ecology of Species Range Limits. *Annual Review of Ecology, Evolution, and Systematics* **40**, 415-436.
Shanks AL (2009) Pelagic Larval Duration and Dispersal Distance Revisited. *Biological Bulletin* **216**, 373-385.
Shima JS, Swearer SE (2010) The legacy of dispersal: larval experience shapes persistence later in the life of a reef fish. *Journal of Animal Ecology* **79**, 1308-1314.

Shirai S, Nakamura S (1994) *Pearls and Pearl Oysters of the World* Marine Planning Co., Okinawa, Japan.
Siegel DA, Kinlan BP, Gaylord B, Gaines SD (2003) Lagrangian descriptions of marine larval dispersion. *Marine Ecology Progress Series* **260**, 83-96.
Siegel DA, Mitarai S, Costello CJ, *et al.* (2008) The stochastic nature of larval connectivity among nearshore marine populations. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 8974-8979.
Simpson SD, Harrison HB, Claereboudt MR, Planes S (2014) Long-Distance Dispersal via Ocean Currents Connects Omani Clownfish Populations throughout Entire Species Range. *PloS ONE* **9**, e107610.
Sims M, Cox T, Lewison R (2008) Modeling spatial patterns in fisheries bycatch: improving bycatch maps to aid fisheries management. *Ecological Applications* **18**, 649-661.
Slate J, David P, Dodds KG, *et al.* (2004) Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity* **93**, 255-265.
Slatkin M (1987) Gene Flow and the Geographic Structure of Natural Populations. *Science* **236**, 787-792.
Slatkin M (1993) Isolation by Distance in Equilibrium and Non-Equilibrium Populations. *Evolution* **47**, 264-279.
Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431-432.
Solé-Cava AM, Thorpe JP (1991) High levels of genetic variation in natural populations of marine lower invertebrates. *Biological Journal of the Linnean Society* **44**, 65-80.
Song Z, Shu Q, Bao Y, Yin X, Qiao F (2015) The prediction on the 2015/16 El Niño event from the perspective of FIO-ESM. *Acta Oceanologica Sinica* **34**, 67-71.
Southgate PC (2008) Pearl Oyster Culture. In: *The Pearl Oyster* (eds. Southgate PC, Lucas JS), pp. 231-272. Elsevier, Amsterdam, the Netherlands.
Southgate PC, Strack E, Hart A, *et al.* (2008) Exploitation and Culture of Major Commercial Species. In: *The Pearl Oyster* (eds. Southgate PC, Lucas JS), pp. 303-355. Elsevier, Amsterdam, the Netherlands.
SPC (2003) Profiles of high interest aquaculture commodities for Pacific Islands countries, p. 71. Secretariat of the Pacific Community, Noumea, New Caledonia.
SPC (2007) SPC Aquaculture Action Plan 2007, p. 31. Secretariat of the Pacific Community, Noumea, New Caledonia.
Stapley J, Reger J, Feulner PGD, *et al.* (2010) Adaptation genomics: the next generation. *Trends in Ecology and Evolution* **25**, 705-712.
Steane DA, Nicolle D, Sansaloni CP, *et al.* (2011) Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping. *Molecular Phylogenetics and Evolution* **59**, 206-224.
Steane DA, Potts BM, McLean E, *et al.* (2015) Genome-wide scans reveal cryptic population structure in a dry-adapted eucalypt. *Tree Genetics & Genomes* **11**, 32-45.

Steinig EJ, Neuditschko M, Khatkar MS, Raadsma HW, Zenger KR (2016) NetView P: A network visualization tool to unravel complex population structure using genome-wide SNPs. *Molecular Ecology Resources*, 1-12.
Stickney RR (2005) Aquaculture. In: *Encylopedia of Coastal Science* (ed. Schwartz ML), pp. 33-38. Springer Netherlands, Netherlands.
Strathmann MF, Strathmann RR (2007) An Extraordinarily Long Larval Duration of 4.5 Years from Hatching to Metamorphosis for Teleplanic Veligers of *Fusitriton oregonensis*. *The Biological Bulletin* **213**, 152-159.
Suárez-Ulloa V, Fernández-Tajes J, Manfrin C, *et al.* (2013) Bivalve Omics: State of the Art and Potential Applications for the Biomonitoring of Harmful Marine Compounds. *Marine Drugs* **11**, 4370-4389.
Sundqvist L, Zackrisson M, Kleinhans D (2013) Directional genetic differentiation and asymmetric migration. *arXiv pre-print: arXiv:1304.0118v2*, 14.
Takeuchi T, Kawashima T, Koyanagi R, *et al.* (2012) Draft Genome of the Pearl Oyster *Pinctada fucata*: A Platform for Understanding Bivalve Biology. *DNA Research* **19**, 117-130.
Takeuchi T, Koyanagi R, Gyoja F, *et al.* (2016) Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zoological Letters* **2**, 1-13.
Takgi R, Miyashita T (2014) A cDNA Cloning of a Novel Alpha-Class Tyrosinase of Pinctada fucata: Its Expression Analysis and Characterization of the Expressed Protein. *Enzyme Research* **2014**, 9.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*.
Tan S, Zhong Y, Hou H, Yang S, Tian D (2012) Variation of presence/absence genes among *Arabidopsis* populations. *BMC Evolutionary Biology* **12**, 1-11.
Tëmkin I (2010) Molecular phylogeny of pearl oysters and their relatives (Mollusca, Bivalvia, Pterioidea). *BMC Evolutionary Biology* **10**, 1-28.
Thomas Y, Dumas F, Andréfouët S (2014) Larval Dispersal Modeling of Pearl Oyster *Pinctada margaritifera* following Realistic Environmental and Biological Forcing in Ahe Atoll Lagoon. *PloS ONE* **9**, e95050.
Thorpe JP, Solé-Cava AM, Watts PC (2000) Exploited marine invertebrates: genetics and fisheries. *Hydrobiologia* **420**, 165-184.
Toonen RJ, Puritz JB, Forsman ZH, *et al.* (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ* **1**.
Tranter DJ (1958) Reproduction in Australian pearl oysters (Lamellibranchia). IV. *Pinctada margaritifera* (L.) . *Australian Journal of Marine and Freshwater Research* **9**, 509-523.
Varotsos CA, Tzanis CG, Sarlis NV (2016) On the progress of the 2015–2016 El Niño event. *Atmos. Chem. Phys.* **16**, 2007-2011.
Vekemans X (2002) AFLP-SURV version 1.0. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium., Brussels, Belgium.
Vekemans X, Beauwens T, Lemaire M, Roldán-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology* **11**, 139-151.

Vierna J, Cuperus J, Martínez-Lage A, *et al.* (2014) Species delimitation and DNA barcoding of Atlantic *Ensis* (Bivalvia, Pharidae). *Zoologica Scripta* **43**, 161-171.
Viikmäe B, Torsvik T, Soomere T (2013) Impact of horizontal eddy diffusivity on Lagrangian statistics for coastal pollution from a major marine fairway. *Ocean Dynamics* **63**, 589-597.
Vilisoni MTJ (2012) *Recruitment patterns of molluscs in Savusavu Bay, Fiji with emphasis on the Blacklip Pearl Oyster, Pinctada margaritifera (Linnaeus, 1758)* Unpublished thesis, University of the South Pacific.
Vucetich JA, Waite TA (2003) Spatial patterns of demography and genetic processes across the species' range: Null hypotheses for landscape conservation genetics. *Conservation Genetics* **4**, 639-645.
Wada KT, Jerry DR (2008) Population Genetics and Stock Improvement. In: *The Pearl Oyster* (eds. Southgate PC, Lucas JS), pp. 437-471. Elsevier, Amsterdam, the Netherlands.
Wada KT, Tëmkin I (2008) Taxonomy and Phylogeny: Commercial Species. In: *The Pearl Oyster* (eds. Southgate PC, Lucas JS), pp. 54-57. Elsevier, Amsterdam, the Netherlands.
Wang L, Liu S, Zhuang Z, *et al.* (2013) Population Genetic Studies Revealed Local Adaptation in a High Gene-Flow Marine Fish, the Small Yellow Croaker (*Larimichthys polyactis*). *PloS ONE* **8**, e83493.
Wang S, Meyer E, McKay JK, Matz MV (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat Meth* **9**, 808-810.
Wang Y, Nielsen R (2012) Estimating population divergence time and phylogeny from single-nucleotide polymorphisms data with outgroup ascertainment bias. *Molecular Ecology* **21**, 974-986.
Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity* **89**, 438-450.
Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **15**, 1419-1439.
Waples RS, Punt AE, Cope JM (2008) Integrating genetic data into management of marine resources: how can we do it better? *Fish and Fisheries* **9**, 423-449.
Ward RD (2000) Genetics in fisheries management. *Hydrobiologia* **420**, 191-201.
Waters JM, Fraser CI, Hewitt GM (2013) Founder takes all: density-dependent processes structure biodiversity. *Trends in Ecology & Evolution* **28**, 78-85.
Weersing K, Toonen RJ (2009) Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series* **393**, 1-12.
Wessel P, Smith WHF, Scharroo R, Luis J, Wobbe F (2013) Generic Mapping Tools: Improved Version Released. *Eos, Transactions American Geophysical Union* **94**, 409-410.
White C, Selkoe KA, Watson J, *et al.* (2010a) *Ocean currents help explain population genetic structure*.
White TA, Stamford J, Rus Hoelzel A (2010b) Local selection and population structure in a deep-sea fish, the roundnose grenadier (*Coryphaenoides rupestris*). *Molecular Ecology* **19**, 216-226.

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology* **14**, 611-619.
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 and Comparative Biology* **42**, 941-952.
Willing E-M, Dreyer C, van Oosterhout C (2012) Estimates of Genetic Differentiation Measured by Fst Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. *PloS ONE* **7**, e42649.
Wood S, Paris CB, Ridgwell A, Hendy EJ (2014) Modelling dispersal and connectivity of broadcast spawning corals at the global scale. *Global Ecology and Biogeography* **23**, 1-11.
Wörheide G, Epp LS, Macis L (2008) Deep genetic divergences among Indo-Pacific populations of the coral reef sponge *Leucetta chagosensis* (Leucettidae): Founder effects, vicariance, or both? *BMC Evolutionary Biology* **8**, 1-18.
Yu DH, Chu KH (2005) Phylogenetics of the common pearl oysters in the genus *Pinctada*: evidence from nrDNA ITS sequence. *Biodiversity Science* **13**, 315-323.
Yu DH, Chu KH (2006) Genetic variation in wild and cultured populations of the pearl oyster *Pinctada fucata* from southern China. *Aquaculture* **258**, 220-227.
Yu DH, Huang G-j, Guo Y-h, Wang X-y, Wang A-m (2009) Genetic characteristics of hybrid populations derived by crossing Chinese and Indian pearl oysters, *Pinctada fucata*, based on AFLP markers. *African Journal of Agricultural Research* **4**, 659-664.
Yue GH (2013) Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish and Fisheries* **15**, 376-396.
Zarraonaindia I, Iriondo M, Albaina A, Pardo MA (2012) Multiple SNP markers reveal fine-scale population and deep phylogeographic structure in European anchovy (*Engraulis encrasicolus* L.). *PloS ONE* **7**, e42201.
Zenger KR, Khatkar MS, Cavanagh JA, Hawken RJ, Raadsma HW (2007a) Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into Australian and global population variability, including impact of selection. *Anim Genet* **38**, 7-14.
Zenger KR, Khatkar MS, Tier B, *et al.* (2007b) QC analyses of SNP array data: experience from a large population of dairy sires with 23.8 million data points *Proceedings of the Association for the Advancement of Animal Breeding and Genetics* **17**, 123-126.
Zhang G, Fang X, Guo X, *et al.* (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49-54.
Zhong X, Li Q, Yu H, Kong L (2014) SNP Mining in *Crassostrea gigas* EST Data: Transferability to Four Other *Crassostrea* Species, Phylogenetic Inferences and Outlier SNPs under Selection. *PloS ONE* **9**, e108256.
Zouros E, Foltz DW (1984) Possible explanations of heterozygote deficiency in bivalve molluscs. *Malacologia* **25**, 583-591.

ADDITIONAL FILES

Please note that all additional files are animations presented in a .gif file format, and need to be opened in a web browser to display correctly. These have been supplied as separate files together with the electronic copy of this thesis.

Additional file 1. Animation of full particle dispersal model simulation run for 2009 over 100 days. Particle seed location colour codes are identical to those described in **Fig 3.4**.

Additional file 2a. Animation of particle dispersal model simulation using 2014 HYCOM data for spawning season 1. Particle seed location colour codes for 11 populations are identical to those described in **Figure 4.1**.

Additional file 2b. Animation of particle dispersal model simulation using 2014 HYCOM data for spawning season 2. Particle seed location colour codes for 10 populations are identical to those described in **Figure 4.1**.

Additional file 2c. Animation of particle dispersal model simulation using 2015 HYCOM data for spawning season 1. Particle seed location colour codes for 11 populations are identical to those described in **Figure 4.1**.

Additional file 2d. Animation of particle dispersal model simulation using 2015 HYCOM data for spawning season 2. Particle seed location colour codes for 10 populations are identical to those described in **Figure 4.1**.

APPENDICES

CHAPTER 2

Appendix 2.1 Sequencing recovery rates and SNP identification at each filtering step.

Metric	Values
Number of raw reads	295,678,888
Number of retained reads	276,918,263
Overall number of stacks	163,106
Number of SNP loci retained	21,331
Average number of stacks per individual	31,122
Average read depth per stack	12.2
Range of numbers of stacks	8149 - 11573
Range of stack depths	1- 55,777

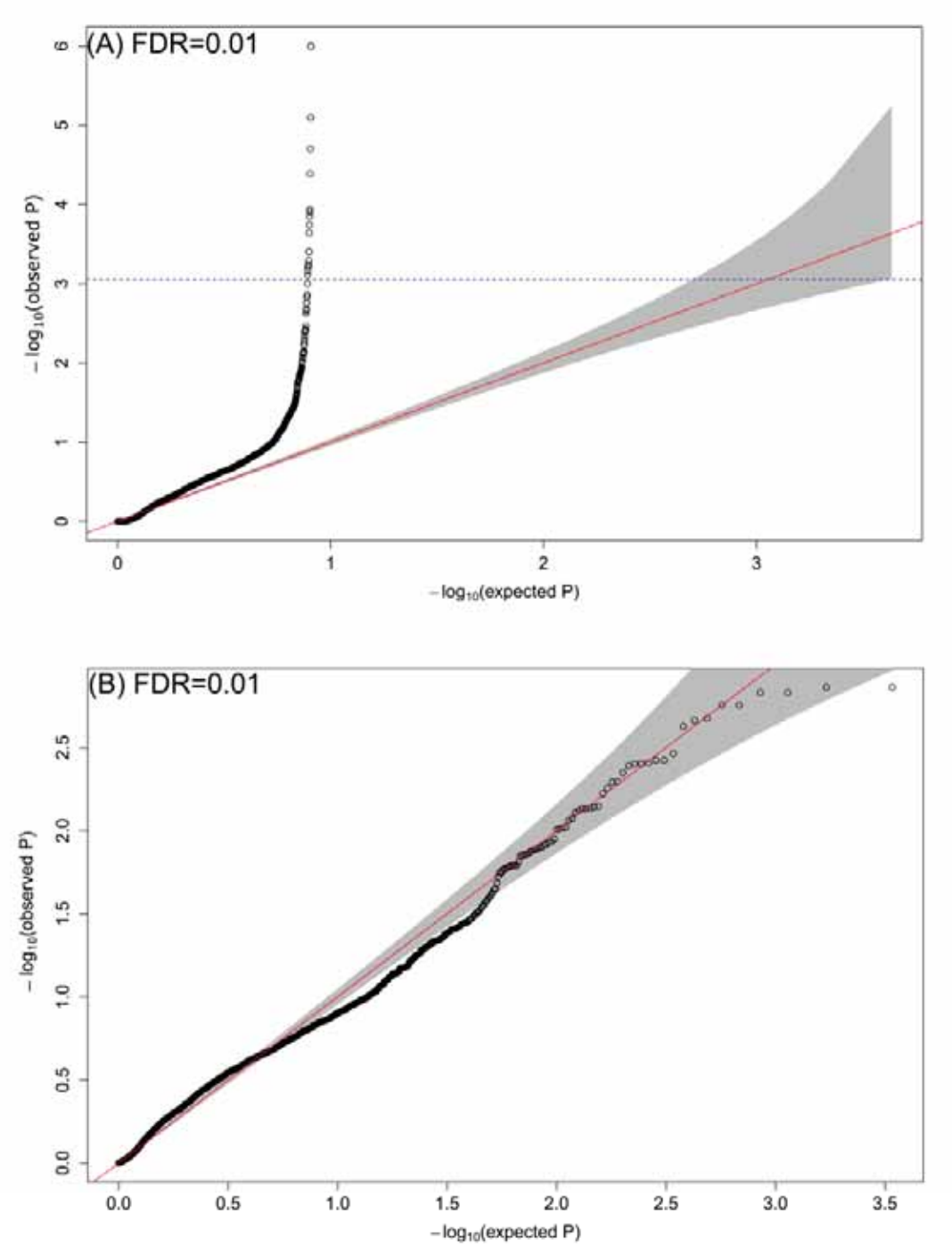
Appendix 2.2 Numbers of putative directional F_{st} outlier loci discovered in *P. margaritifera* at five False Discovery Rate thresholds using Bayescan 2.1 (Foll, 2012) and LOSITAN (Antao et al. 2008).

Population	FDR	Bayescan 2.1	LOSITAN*	Jointly-identified
Namarai (wild)	0.01	1	42	1
	0.05	2	95	2
	0.10	3	116	3
	0.15	4	121	4
	0.20	5	121	5
Raviravi (wild)	0.01	1	62	1
	0.05	2	105	2
	0.10	3	100	3
	0.15	3	112	3
	0.20	6	112	6
Savusavu (wild)	0.01	1	44	1
	0.05	1	92	1
	0.10	4	103	3
	0.15	5	104	5
	0.20	7	104	7

*Cutoffs for Namarai were at F_{st}>0.27, for Raviravi at F_{st}>0.41 and Savusavu (wild) at F_{st}>0.30.

Appendix 2.3 Table of descriptions returned from a BLAST search of 204 putatively associated loci after case control testing in *P. margaritifera*. The first 5 loci listed (indicated with an asterisk) are those which returned hits for pearl oyster shell matrix and melanin biosynthesis pathways.

Locus number	Accession number	Description	Max score	Total score	Query coverage	E value	Identity
1*	DQ116438.1	*Pinctada fucata* mantle protein 11 (MG11) mRNA, complete cds	24.7	24.7	6%	9.7	100%
2*	HE610377.1	*Pinctada margaritifera* mRNA for tyrosinase 1	22.9	22.9	6%	3.4	100%
3*	HE610378.1	*Pinctada margaritifera* mRNA for tyrosinase 2	35.6	35.6	9%	0.0005	100%
4*	KJ533305.1	*Pinctada maxima* tyrosinase A2 (TyrA2) mRNA, complete cds	26.5	26.5	9%	2.5	89%
5*	KJ533314.1	*Pinctada maxima* tyrosinase B5 (TyrB5) mRNA, complete cds	22.9	22.9	6%	3.1	100%
6	AB823700.1	*Pinctada fucata* mRNA for matrix protein, complete cds	24.7	24.7	6%	9.7	100%
7	EU887510.1	*Pinctada fucata* serine protease mRNA, complete cds	24.7	24.7	8%	9.7	94%
8	FJ267519.1	*Pinctada fucata* galectin mRNA, complete cds	26.5	26.5	8%	2.8	94%
9	GQ452847.1	*Pinctada maxima* mitochondrion, complete genome	24.7	49.5	17%	9.7	100%
10	GU971706.1	*Pinctada martensi* clone Pm58 microsatellite sequence	26.5	26.5	7%	2.8	100%
11	HE610379.1	*Pinctada margaritifera* mRNA for egf-like 1	24.7	24.7	9%	0.98	89%
12	HE610387.1	*Pinctada margaritifera* mRNA for prism uncharacterized shell protein 1 (pusp1 gene)	24.7	24.7	6%	9.7	100%
13	HM357125.1	*Pinctada maxima* microsatellite M1331 sequence	22.9	22.9	6%	3.1	100%
14	HM467838.1	*Pinctada margaritifera* mitochondrion, complete genome	24.7	24.7	6%	9.7	100%
15	HQ014601.1	*Pinctada fucata* tandem-repeat galectin mRNA, complete cds	26.5	26.5	8%	2.8	94%
16	JQ898347.1	*Pinctada fucata* tumor necrosis factor receptor-associated factor 3 (TRAF3) mRNA, complete cds	28.3	28.3	9%	0.79	94%
17	KF017176.1	*Pinctada maxima* microsatellite D2-E09 sequence	22.9	22.9	6%	3.1	100%
18	KF017277.1	*Pinctada fucata* manganese superoxide dismutase gene, complete cds	24.7	24.7	6%	9.7	100%
19	KF524261.1	*Pinctada martensi* heat shock protein 60 mRNA, complete cds	24.7	24.7	9%	9.7	89%
20	KJ010543.1	*Pinctada martensi* cytochrome P450 family 4 (CYP4) mRNA, complete cds	22.9	22.9	7%	5.4	93%
21	KJ010544.1	*Pinctada martensi* aryl hydrocarbon receptor mRNA, complete cds	22.9	22.9	6%	5.4	100%
22	KJ907377.1	*Pinctada margaritifera* sex determining protein Fem-1 like protein mRNA, complete cds	24.7	24.7	6%	0.99	100%
23	KM593798.1	*Pinctada fucata* NFAT protein (NFAT) gene, complete cds	24.7	24.7	8%	9.7	94%
24	KM593798.1	*Pinctada fucata* NFAT protein (NFAT) gene, complete cds	24.7	24.7	8%	8.7	94%
25	KM874283.1	*Pinctada margaritifera* vitellogenin-6 mRNA, partial cds	22.9	22.9	6%	3.4	100%
26	KM874284.1	*Pinctada margaritifera* GATA-type zinc finger protein 1 mRNA, complete cds	22.9	22.9	6%	3.4	100%
27	KP276261.1	*Pinctada martensii* mineralization related protein 2 mRNA, complete cds	24.7	24.7	6%	9.7	100%



Appendix 2.4 Verification of outlier loci detected in the Namarai population colour morphotypes using Quantile-Quantile plots (QQ plots) at an FDR of 0.01. **(A)** displays the p value distributions of all SNP loci while **(B)** displays the distribution when all outlier loci are removed. The dashed blue line indicates the threshold of outlier data and corresponds to a p value of $1.355e^{-3}$ for **(A)**. The red line indicates y=x linearity for conformity to a normal distribution, with the surrounding grey area approximating a 95 % confidence interval.

CHAPTER THREE:

Appendix 3.1 Sequencing recovery rates and SNP identification at each filtering step in the STACKs 1.20 pipeline.

Metric	Value
Number of raw reads	765,273,656
Number of retained reads	725,064,036
Overall number of stacks	303650
Number of SNP loci retained	42,341
Average number of stacks per individual	33,738.94
Average read depth per stack	17.81
Range of numbers of stacks	2,062 - 95,560
Range of stack depths	8.47 - 51.94

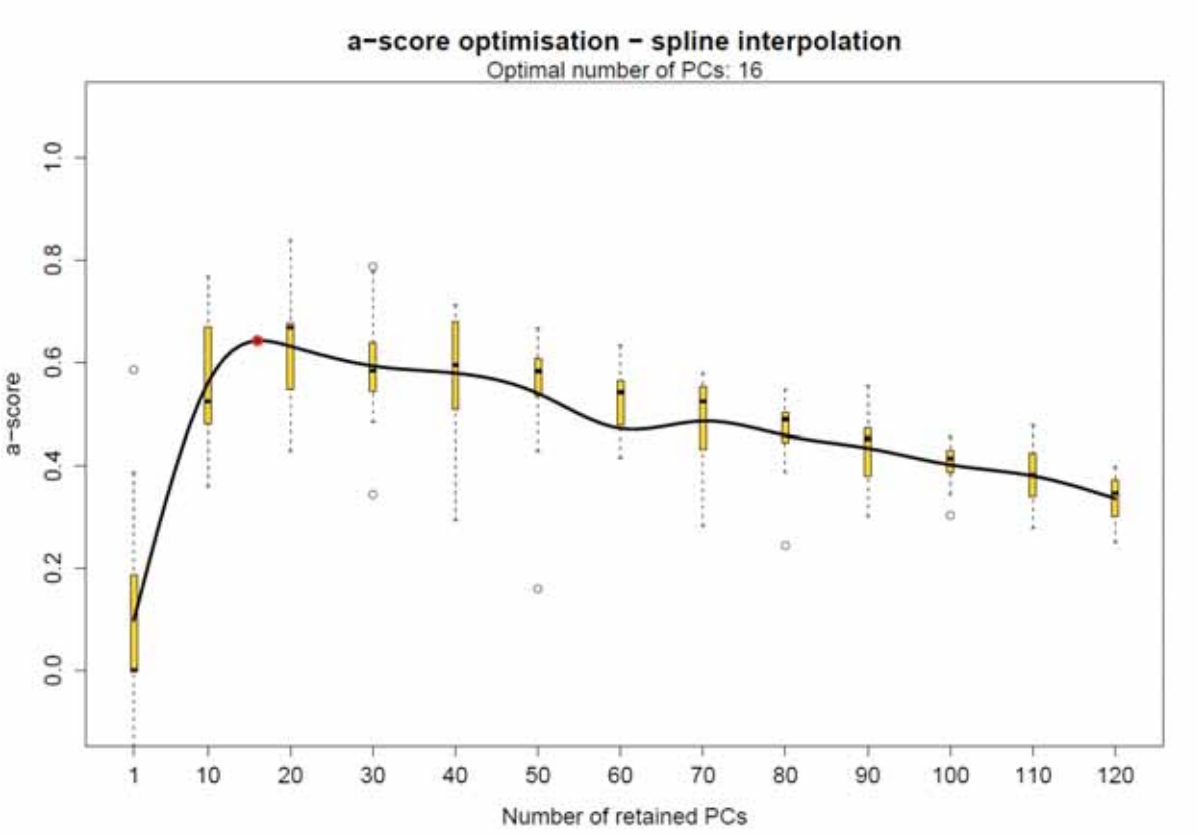
Appendix 3.2 Estimates of relationships between individuals within eleven Fijian populations of *P. margaritifera* with 4,123 SNP loci, using ML-RELATE (Kalinowski *et al.* 2006).

Population	Total # relationships tested	Unrelated relationships	Full sib relationships	Half sib relationships	Parent-offspring relationships
Ra (Farm)	1225	1223	1	1	0
Raviravi (Farm)	496	496	0	0	0
Lau (Wild)	1225	1220	0	5	0
Yasawa (Wild)	595	593	2	0	0
Udu Point (Wild)	153	153	0	0	0
Taveuni (Farm)	903	902	0	1	0
Kadavu, Galoa (Wild)	300	300	0	0	0
Kadavu, Ravitaki (Wild)	300	299	0	1	0
Savusavu, Vatubukulaca (Farm)	1225	1131	8	86	0
Savusavu, Wailevu (Farm)	1176	1174	1	1	0
Savusavu, Wailevu (Farm; hatchery)	1225	1026	83	116	0

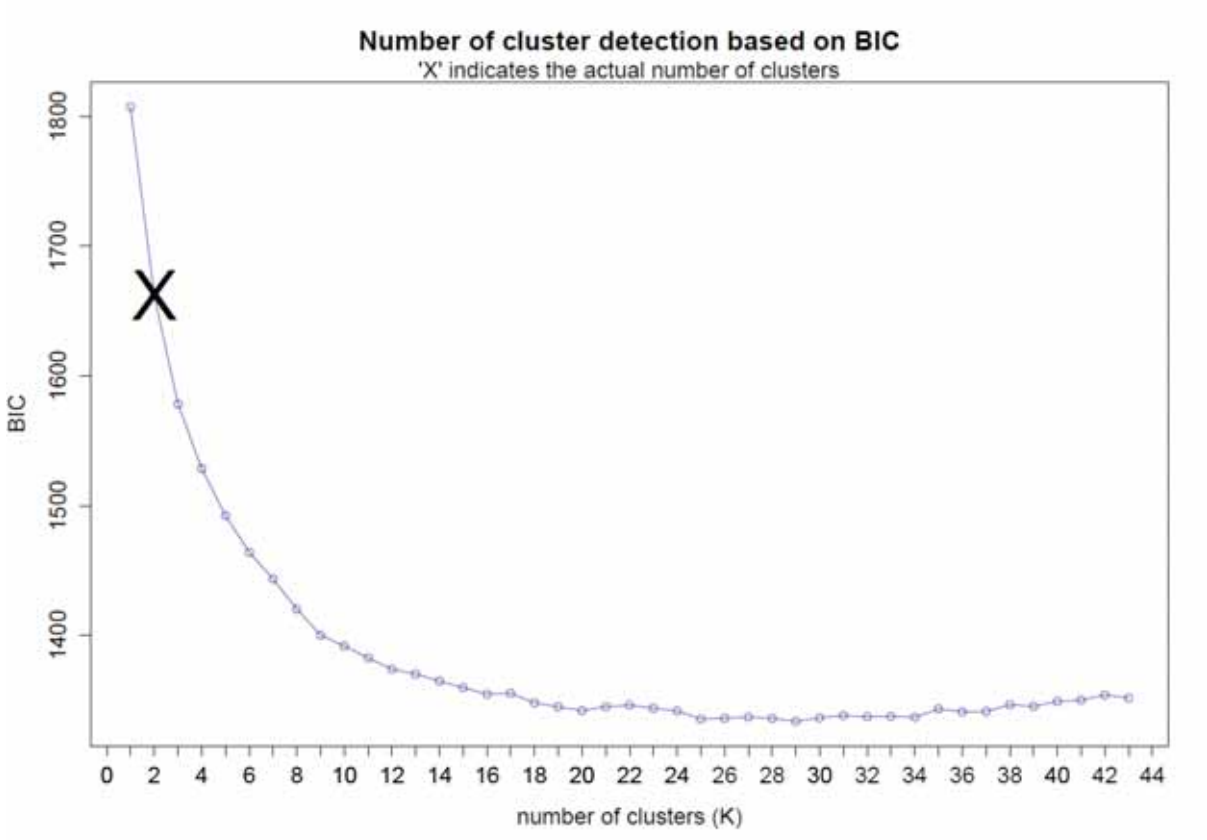
Appendix 3.3 Estimates of full-sib, half-sib and parent-offspring relationships. Estimates are provided between geographic regions sampled from eleven Fijian populations of *P. margaritifera* with 4,123 SNP loci using ML-RELATE (Kalinowski *et al.* 2006). All other between-region relationships examined indicated that individuals were unrelated.

	Yasawa-Lau	Yasawa-Raviravi	Yasawa-Taveuni	Yasawa-Udu Point	Kadavu-Udu Point	Kadavu-Taveuni	Kadavu-Lau	Kadavu-Yasawa	Kadavu-Raviravi	Kadavu-Ra	Taveuni-Ra	Taveuni-Raviravi
Full sib relationships	1	0	1	0	0	2	2	2	0	0	0	0
Half sib relationships	1	4	7	1	2	19	19	9	6	2	1	2
Parent-offspring relationships	0	0	0	0	0	0	0	0	0	0	0	0

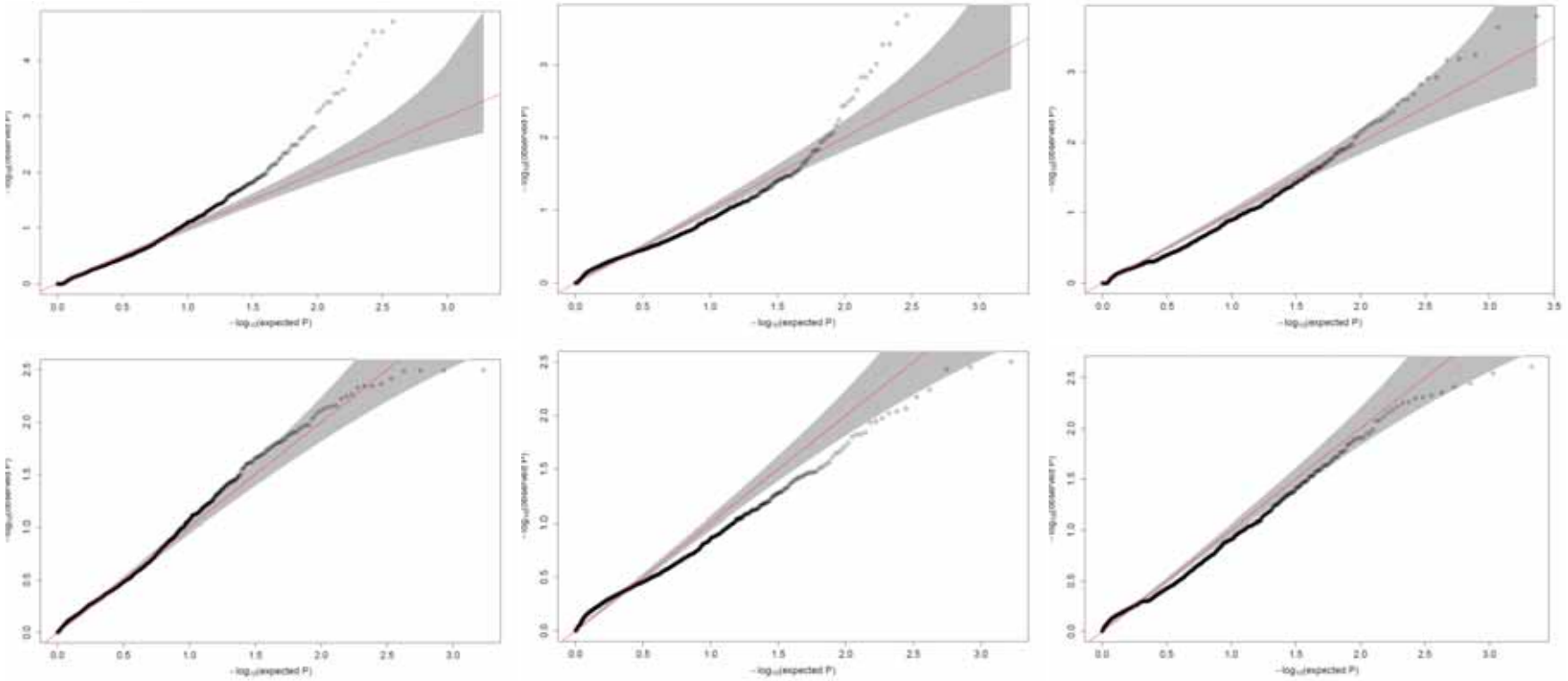
	Savusavu-Ra	Savusavu-Udu Point	Savusavu-Kadavu	Savusavu-Raviravi	Savusavu-Lau	Savusavu-Taveuni	Savusavu-Yasawa	Lau-Ra	Lau-Raviravi	Lau-Taveuni	Lau-Udu Point
Full sib relationships	0	0	4	0	25	4	4	0	0	0	0
Half sib relationships	5	6	24	14	73	37	17	3	4	5	1
Parent-offspring relationships	0	0	0	1	0	0	0	0	0	0	0



Appendix 3.4 α-score optimisation graph for generation of the Discriminant Analysis of Principal Components (DAPC) scatter plot. An optimal number of 16 principal components were suggested for retention using this analysis, based on 4,123 SNP loci in the *R* package *adegenet* (Jombart 2008; Jombart *et al.* 2010).



Appendix 3.5 Determination of the number of clusters following generation of the DAPC scatter plot using 4,123 SNP loci. An optimal number of k=2 was suggested based on the BIC method implemented in the find.clusters function of the *R* package *adegenet* (Jombart 2008; Jombart & Ahmed 2011).



Appendix 3.6 Verification of outlier loci detected for population pairwise comparisons using Quantile-Quantile plots (QQ plots) at an FDR of 0.01. Comparisons shown are for Savusavu-Lau (left), Udu Point-Kadavu (middle) and Yasawa-Lau (right). QQ plots are arranged in pairs with the top row displaying the p value distributions of all SNP loci while the bottom row displays the distribution when all outlier loci are removed. The red line indicates y=x linearity for conformity to a normal distribution, with the surrounding grey area approximating a 95 % confidence interval.

CHAPTER FOUR:

Appendix 4.1 Summary of temporal seed inputs for particle dispersal model.

Sampling site	Particle seed #1 start date (Spawning peak 1)	Particle seed #2 start date (Spawning peak 2)	References	Comment
Australia Abrolhos Islands.	October 1	None	(Rose 1990; Rose & Baker 1994; Saucedo & Southgate 2008; Tranter 1958)	
Taiwan	June 1	None	(Arjarasirikoon *et al.* 2004)	
Vietnam	June 1	November 1	(Arjarasirikoon *et al.* 2004)	
Indonesia	June 1	November 1	(Saucedo & Southgate 2008)	Inferred from data for spawning behaviour in documented from Vietnam and Indonesia.
Papua New Guinea	July 1	November 1	(Saucedo & Southgate 2008)	Inferred from data for spawning behaviour documented from the Great Barrier Reef.
Solomon Islands	July 1	November 1	(Saucedo & Southgate 2008)	Inferred from data for spawning behaviour documented from the Great Barrier Reef.
Australia Great Barrier Reef	July	November 1	(Tranter 1958)	
Fiji Islands	March 1	November 1	(Lal *et al.* 2016b; Vilisoni 2012)	
Tonga	March 1	November 1	(Saucedo & Southgate 2008)	Inferred from data for spawning behaviour documented from the Fiji Islands.
Cook Islands	May 1	November 1	(Saucedo & Southgate 2008)	Inferred from data for spawning behaviour documented from French Polynesia.
French Polynesia	May 1	November 1	(Pouvreau *et al.* 2000)	

Appendix 4.2 Numbers of putative directional and balancing F_{st} outlier loci discovered in *P. margaritifera*. Data are reported following testing of Pacific Ocean populations at six False Discovery Rate thresholds, using BayeScan 2.1 (Foll 2012) and LOSITAN (Antao *et al.* 2008). Jointly-identified loci were identified using both outlier detection platforms.

FDR level	Bayescan 2.1 total outliers	Bayescan 2.1 directional	Bayescan 2.1 balancing	LOSITAN directional	LOSITAN balancing	Jointly-identified directional	Jointly-identified balancing
0.001	310	310	0	513	434	87	0
0.005	396	396	0	254	116	55	0
0.01	429	429	0	241	124	45	0
0.05	571	517	54	418	211	89	37
0.10	700	585	115	518	405	112	90
0.20	950	701	249	592	1215	137	216

Appendix 4.3 Summary of numbers of both putatively balancing and directional SNPs detected. Loci are reported following testing of the entire dataset, to identify selectively-neutral SNPs.

FDR threshold	Number of outlier SNPs detected by LOSITAN	Number of outlier SNPs detected by Bayescan 2.1	Jointly identified SNPs
0.001	6403	620	423
0.005	4397	901	487
0.01	4843	1059	621
0.05	5778	1605	1085
0.1	6480	1956	1392
0.2	7285	2498	1871

Appendix 5.1 Published journal article reprints

Reprints of all published scientific journal articles produced within the PhD candidature:

Lal, M.M., Southgate, P.C., Jerry, D.R., Zenger, K.R. (2016) Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. *Marine Genomics* **25**. 57-68

Lal, M.M., Southgate P.C., Jerry D.R., Bosserelle C., Zenger K.R. (2016) A parallel population genomic and hydrodynamic approach to fishery management of highly-dispersive marine invertebrates: the case of the Fijian black-lip pearl oyster *Pinctada margaritifera*. *PloS ONE* **11**, e0161390.

Lal, M.M., Southgate P.C., Jerry D.R., Bosserelle C., Zenger K.R. (2017) Swept away: ocean currents and seascape features influence genetic structure across the 18,000 Km Indo-Pacific distribution of a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. *BMC Genomics* **18**, 66.

Marine Genomics 25 (2016) 57–68

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*☆,☆☆

Monal M. Lal *, Paul C. Southgate [1], Dean R. Jerry, Kyall R. Zenger

[a] Centre for Sustainable Tropical Fisheries and Aquaculture (CSTFA), James Cook University, Townsville Campus, Townsville, QLD 4811, Australia
[b] College of Marine and Environmental Sciences (CMES), James Cook University, Townsville Campus, Townsville, QLD 4811, Australia

ARTICLE INFO

Article history:
Received 30 September 2015
Received in revised form 27 October 2015
Accepted 27 October 2015
Available online 3 November 2015

Keywords:
Population genomics
Aquaculture
SNP
Fine-scale differentiation
Colour morphotype
Fiji Islands

ABSTRACT

Population genomic investigations on highly dispersive marine organisms typically require thousands of genome-wide SNP loci to resolve fine-scale population structure and detect signatures of selection. This information is important for species conservation efforts and stock management in both wild and captive populations, as well as genome mapping and genome wide association studies. Double digest Restriction site-Associated DNA Sequencing (ddRADseq) is a recent tool for delivering genome wide SNPs for non-model organisms. However, its application to marine invertebrate taxa has been limited, particularly given the complex and highly repetitive nature of many of these organisms' genomes. This study develops and evaluates an optimised ddRADseq technique together with associated analyses for generating genome-wide SNP data, and performs population genomic analyses to inform aquaculture and fishery management of a marine bivalve, the black-lip pearl oyster *Pinctada margaritifera*. A total of 5243 high-quality genome-wide SNP markers were detected, and used to assess population structure, genome diversity, detect F_{st} outliers and perform association testing in 156 individuals belonging to three wild and one hatchery produced populations from the Fiji Islands. Shallow but significant population structure was revealed among all wild populations (average pairwise $F_{st} = 0.046$) when visualised with DAPC and an individual network analysis (NetView P), with clear evidence of a genetic bottleneck in the hatchery population ($N_{eLD} = 6.1$), compared to wild populations ($N_{eLD} > 192.5$). F_{st} outlier detection revealed 42–62 highly differentiated SNPs ($p < 0.02$), while case–control association discovered up to 152 SNPs ($p < 0.001$). Both analyses were able to successfully differentiate individuals between the orange and black tissue colour morphotypes characteristic of this species. BLAST searches revealed that five of these SNPs were associated with a melanin biosynthesis pathway, demonstrating their biological relevance. This study has produced highly informative SNP and population genomic data in *P. margaritifera*, and using the same approach promises to be of substantial value to a range of other non-model, broadcast-spawning or marine invertebrate taxa.

Content has been removed
due to copyright restrictions

Content has been removed
due to copyright restrictions

Content has been removed
due to copyright restrictions

Content has been removed
due to copyright restrictions

Content has been removed
due to copyright restrictions

Content has been removed
due to copyright restrictions

Content has been removed
due to copyright restrictions

Content has been removed
due to copyright restrictions

Content has been removed
due to copyright restrictions

PLOS ONE

RESEARCH ARTICLE

A Parallel Population Genomic and Hydrodynamic Approach to Fishery Management of Highly-Dispersive Marine Invertebrates: The Case of the Fijian Black-Lip Pearl Oyster *Pinctada margaritifera*

Monal M. Lal[1,2]*, Paul C. Southgate[1,2,3], Dean R. Jerry[1,2], Cyprien Bosserelle[4], Kyall R. Zenger[1,2]

1 Centre for Sustainable Tropical Fisheries and Aquaculture, James Cook University, Townsville, Queensland, Australia, **2** College of Science and Engineering, James Cook University, Townsville, Queensland, Australia, **3** Australian Centre for Pacific Islands Research, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Maroochydore, Queensland, Australia, **4** Geoscience Division, Secretariat of the Pacific Community, Nabua, Suva, Fiji Islands

* monal.lal@my.jcu.edu.au

Citation: Lal MM, Southgate PC, Jerry DR, Bosserelle C, Zenger KR (2016) A Parallel Population Genomic and Hydrodynamic Approach to Fishery Management of Highly-Dispersive Marine Invertebrates: The Case of the Fijian Black-Lip Pearl Oyster *Pinctada margaritifera*. PLoS ONE 11(8): e0161390. doi:10.1371/journal.pone.0161390

Editor: Michael A. Russello, University of British Columbia Okanagan, CANADA

Received: May 15, 2016

Accepted: August 4, 2016

Published: August 25, 2016

Data Availability Statement: Additional data will be available from a Dryad repository at http://dx.doi.org/10.5061/dryad.802q5 These include genotypic data (4,123 SNPs) and sampling site information.

Funding: This study was conducted within the Australian Centre for International Agricultural Research (ACIAR) Project FIS/2009/057: "Pearl Industry Development in the Western Pacific" led by PCS at the University of the Sunshine Coast. The funding source (ACIAR) had no role in the study design, data collection and analysis, interpretation of

Abstract

Fishery management and conservation of marine species increasingly relies on genetic data to delineate biologically relevant stock boundaries. Unfortunately for high gene flow species which may display low, but statistically significant population structure, there is no clear consensus on the level of differentiation required to resolve distinct stocks. The use of fine-scale neutral and adaptive variation, considered together with environmental data can offer additional insights to this problem. Genome-wide genetic data (4,123 SNPs), together with an independent hydrodynamic particle dispersal model were used to inform farm and fishery management in the Fijian black-lip pearl oyster *Pinctada margaritifera*, where comprehensive fishery management is lacking, and the sustainability of exploitation uncertain. Weak fine-scale patterns of population structure were detected, indicative of broad-scale panmixia among wild oysters, while a hatchery-sourced farmed population exhibited a higher degree of genetic divergence (F_{st} = 0.0850–0.102). This hatchery-produced population had also experienced a bottleneck (N_{eLD} = 5.1; 95% C.I. = [5.1–5.3]); compared to infinite N_{eLD} estimates for all wild oysters. Simulation of larval transport pathways confirmed the existence of broad-scale mixture by surface ocean currents, correlating well with fine-scale patterns of population structuring. F_{st} outlier tests failed to detect large numbers of loci supportive of selection, with 2–5 directional outlier SNPs identified (average F_{st} = 0.116). The lack of biologically significant population genetic structure, absence of evidence for local adaptation and larval dispersal simulation, all indicate the existence of a single genetic stock of *P. margaritifera* in the Fiji Islands. This approach using independent genomic and oceanographic tools has allowed fundamental insights into stock structure in this species,

the data, decision to publish, or preparation of the manuscript. For further information please visit: http://aciar.gov.au/project/fis/2009/057.

Competing Interests: The authors have declared that no competing interests exist.

with transferability to other highly-dispersive marine taxa for their conservation and management.

Introduction

Sustainable management and conservation of marine species is faced with a number of challenges, among which is the wide distribution of taxa across diverse habitats and geopolitical jurisdictions, that make species-specific management plans difficult to design and implement. Many taxa also face high rates of exploitation, that in some cases has led to the collapse or abnormally slow recovery of wild fisheries, bringing into question whether current management strategies are effective or appropriate [1–3]. The need for accurate fishery management has resulted in the development of the stock concept for aquatic species, which can allow for targeted conservation efforts and informed exploitation, once stock boundaries have been defined [2,4]. Despite the usefulness and importance of the stock concept, there is currently no clear consensus on what constitutes a stock, and numerous definitions in the literature shift emphasis for defining stock boundaries between the degree of demographic homogeneity within stocks, and their reproductive isolation [5]. Since a stock is the fundamental unit used for fishery assessment and administration, it is imperative that the spatial boundaries delineated are also biologically meaningful, to ensure correct management action [3,6].

For assessment of a particular stock, it is important to determine the number and extent of populations being examined. However, the biological concept of a population has either ecological (demographic interactions of individuals), or evolutionary (genetic structuring) aspects [3,5]. Reiss et al. [3] make the observation that many fishery management and assessment tools focus primarily on the ecological aspects of populations (e.g. population growth and mortality rates), while overall management goals also include many evolutionary criteria such as the conservation of genetic diversity and maintenance of sustainable spawning stock biomass. This disconnect highlights the need for bridging the gap between fisheries management and population genetics, and particularly for characterising stock boundaries, identifying the level of divergence required to manage two populations together, or as separate entities [3–7].

A major problem posed for application of the stock concept in the marine environment is the relative absence of barriers to dispersal and migration compared to terrestrial systems, and the highly-dispersive larval stages of many species [2]. For species which are either highly mobile and/or broadcast spawners with prolonged pelagic larval duration (PLD), the potential for gene dispersal is high, often resulting in weak population differentiation that is evident over large geographic distances [6,8–10]. Furthermore, despite the presence of weak population structure, selective forces can produce fitness differences between populations through local adaptation [11].

For a large number of species that exhibit high levels of gene flow, low levels of genetic structure may be present, but difficult to detect [2,3]. The importance of detecting low, but biologically significant differentiation for understanding the ecology and evolution of these taxa, and implications for their conservation and management is discussed by André et al. [12], Gaggiotti et al. [7], Hauser and Carvalho [13], Palumbi [9, 14], Waples [2] and Waples and Gaggiotti [6]. It is clear from these studies that a common solution for delimiting population and stock boundaries in high gene flow species is not possible, but rather assessment on an individual basis is required, taking into consideration the biological, ecological and fishery management issues involved. Additionally, in situations where traditional stock assessment is not possible (e.g. due to logistical or financial reasons), genetic approaches examining fine-scale population

structure and functional differences (such as local adaptation), can be important for resolving stock boundaries.

A potential solution in recent years has been the use of genome-wide SNPs, which can reveal fine-scale patterns of population structure to highlight differences between populations, and also detect signatures of selection [15–17]; with much higher resolving power than traditional markers (e.g microsatellites and mtDNA). However, while genetic analyses by themselves are a powerful tool for investigating population connectivity and structure, consideration of other data for defining stocks such as phenotypic information, demographic data, or biophysical modelling should not be overlooked [3,18,19]. For broadcast spawning species with prolonged PLD, investigations considering independent environmental and molecular data together, can provide unrivaled insights into the biological and physical processes that organise and regulate population structure [4,20]. Hydrodynamic dispersal modelling is an analysis tool that relies on oceanographic data, and can be used for simulation and independent inference of larval dispersal from source to sink locations [20,21]. Because many marine species produce large quantities of very small larvae with variable PLD that makes tagging and tracking studies very difficult, highly realistic estimates of population connectivity can be achieved when hydrodynamic dispersal data are combined with genetic analyses [4,20,22–24].

Bivalve molluscs present a number of unique challenges for stock assessment, which include highly variable patterns of larval dispersal and recruitment. Additionally, traditional bivalve stock assessment surveys typically require extensive sampling to determine distribution and abundance, which in most situations can be costly and impractical. Because the adults of many taxa are sedentary and recruitment rates highly variable, a stock may occupy a discrete geographic region as large as an entire reef system, or as small as a single bivalve bed [25]. When coupled with the homogenising effects of larval exchange over large distances, accurate stock assessment can quickly become problematic. For many bivalves, and pearl oysters in particular, examination of morphological differences for stock assessment primarily relies on variable shell characters to elucidate differences between individuals, populations and species [26]. This can be a difficult exercise, particularly during early stages of development [27], as factors including phenotypic plasticity and environmental effects can confound measurements. In recent times, molecular methods have been increasingly relied upon to provide solutions to these problems [26,28].

In French Polynesia, the black-lip pearl oyster *Pinctada margaritifera* (Pteriidae) displays substantial genetic fragmentation, despite being a broadcast-spawner with an extended PLD of 26–30 days [29,30]. This has been related primarily to habitat heterogeneity, with significant genetic structure detected between open and closed atoll lagoon systems [31,32]. Here, detection of both fine-scale and broad-scale patterns of differentiation were identified as being biologically important for fishery and aquaculture management [33,34]. For the Fiji Islands, cultured pearl and pearl shell production from *P. margaritifera* is a valuable industry and substantial source of coastal community livelihoods. It produces a high-value, low-volume and non-perishable product with a comparatively smaller environmental footprint than most other forms of aquaculture, making it an ideal export commodity for developing Pacific island countries [35–37]. The industry is almost exclusively dependent on wild oysters for which there are currently no comprehensive fishery management guidelines, and therefore no information is available on the number of discrete populations present, their levels of genetic fitness and relatedness, or if domestic translocation of animals is suitable for the establishment of new pearl farms.

Two preliminary stock assessment surveys using traditional methods reported low abundances of *P. margaritifera* at all locations examined, and recommended immediate conservation efforts to increase population densities of wild oysters [38,39]. A previous study which

examined oysters sampled at four Fijian sites discovered a mixed pattern of population structure, and identified a need for comprehensive evaluation of additional populations to determine country-wide patterns of genetic structure and connectivity [17]. In this study, we assess the stock structure of *P. margaritifera* in the Fiji Islands for fishery and aquaculture management, using independent population genomic and hydrodynamic modelling approaches. This work provides valuable insights for the fishery management and aquaculture of this commercially important bivalve mollusc, and also demonstrates solutions for challenges that apply to stock assessment efforts in other broadcast-spawning marine taxa, that possess similar life history characteristics.

Methods and Materials

Specimen collection, tissue sampling and DNA extraction

Adult and juvenile *P. margaritifera* (n = 427) sized between 7–18 cm in dorso-ventral measurement (DVM), were collected from 11 sites in the Fiji Islands representing both farmed and wild populations country-wide, from December 2012 to October 2013 (Fig 1). Permission to

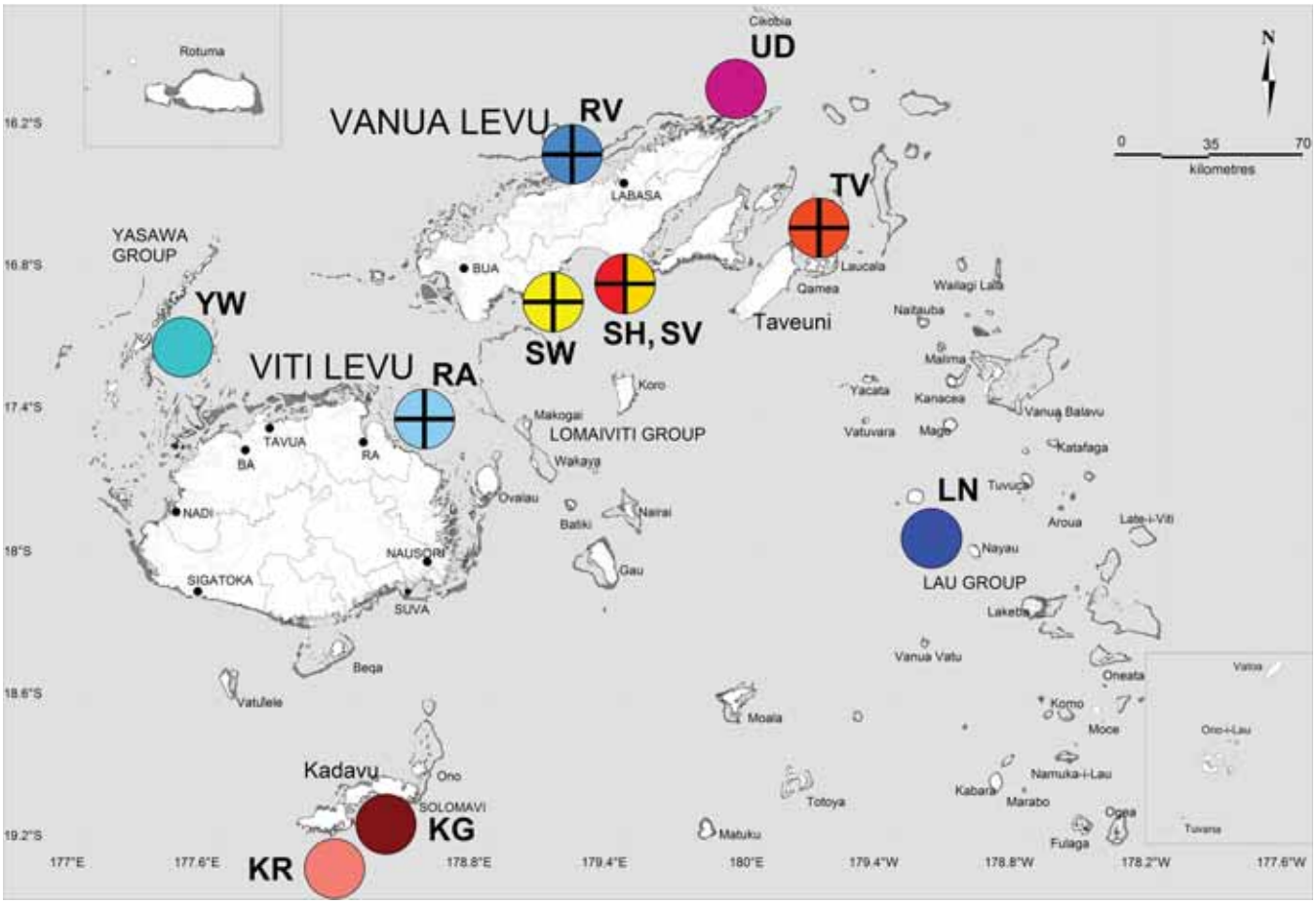


Fig 1. Map of sampling locations in the Fiji Islands adapted from Lal et al. [17], where wild and farmed *P. margaritifera* were collected. Reef outlines are presented in dark grey, and site colours correspond to population colour codes used for Figs 2 and 3. Solid circles represent wild oyster collection sites, while circles superimposed with a cross indicate farm locations. Site codes represent the following locations: YW, Naviti Island in the Yasawa group; RA, farm site at Namarai, Ra; SW, farm site at Wailevu, Savusavu; SH, farm site at Wailevu, Savusavu for hatchery produced oysters; SV, farm site at Vatubukulaca, Savusavu; RV, farm site at Raviravi; UD, Vunikodi, Udu Point; TV, farm site at Wailoa, Taveuni; LN, Nayau Island in the Lau group; KG, Galoa Island off Kadavu Island and KR, Ravitaki on Kadavu Island.

doi:10.1371/journal.pone.0161390.g001

sample wild sites was obtained from Fijian traditional fishing ground (*i qoliqoli*) custodians, while farm site access was permitted by farm owners. The vast majority of farmed oysters are collected as settling wild juveniles or spat, that recruit to dedicated settlement substrates deployed by farms. Additionally, limited numbers of individuals are propagated in a single hatchery, and are the progeny of wild-sourced broodstock. Oysters are grown in pocket panel nets that are suspended in the water column from long lines [40]. At all farm sites, wild populations are present in adjacent habitats. Farmed oysters were sampled at Ra (n = 50), Raviravi (n = 32), Taveuni (n = 43) and three locations in Savusavu: Vatubukulaca (n = 50); Wailevu (n = 49) and a hatchery-produced population also at Wailevu (n = 50). Oysters collected from all farms originated either from spat collectors [40], or were gleaned from adjacent coral reef habitats. Wild populations were sampled at two sites on the Island of Kadavu (Galoa Island; n = 25 and Ravitaki; n = 25), the Yasawa archipelago (Naviti Island; n = 35), Udu Point (n = 18) and the Lau archipelago (Nayau Island, n = 50). Two sites were sampled on Kadavu to detect any differentiation present between adjacent locations due to environmental heterogeneity (e.g. reef effects). Proximal mantle and adductor muscle tissues (1.5 and 1 cm respectively), were removed and transferred to tubes containing 20% salt saturated dimethyl sulfoxide (DMSO-salt) preservative [41]. All oysters were handled in accordance with James Cook University's animal ethics requirements and guidelines. Genomic DNA was extracted using a modified cetyl trimethyl ammonium bromide (CTAB, Amresco, cat. #0833-500G) chloroform/isoamyl alcohol protocol, with a warm (30°C) isopropanol precipitation [42]. To clean up all DNA extractions, a Sephadex G50 [43] spin column protocol was used, prior to quantification with a Nanodrop 1000 Spectrophotometer (Thermo Scientific).

ddRADseq library preparation and sequencing

Double digest restriction site-associated (ddRAD) libraries were prepared following the methods of Peterson et al. [44], with a number of modifications for *P. margaritifera* as described by Lal et al. [17]. Briefly, nine libraries in total were prepared (48 barcoded individuals per pool × nine unique Illumina TruSeq indices), from which four libraries were pooled at equimolar ratios for sequencing in one lane, while the remaining five libraries were pooled for a second lane. After cluster generation and amplification (HiSeq PE Cluster Kit V4 cBOT), 100 bp paired-end sequencing was performed on an Illumina HiSeq 2000 platform at the Australian Genome Research Facility (AGRF) in Melbourne, Victoria.

Sequence quality control, marker filtering and genotype calling

Raw reads obtained following sequencing were processed as described by Lal et al. [17], with all read filtering and SNP genotyping carried out using STACKs v.1.20 software [45,46]. From all available SNPs, only the most informative SNP per locus was selected for further analysis, as per Lal et al. [17]. Final genotypes were called at a Minor Allele Frequency (MAF) of 2% and minimum stack depth of 10, with the minimum proportions of loci allowed across individuals set at 20%, and across populations at 50% (-r and -p options respectively). In addition, each unique SNP was genotyped in at least 10 individuals within a population, and represented in a minimum of two populations across the whole dataset [47].

All loci were screened using Arlequin v.3.5.1.3 [48] for departure from Hardy-Weinberg Equilibrium (HWE), and removed if deviations were significant after FDR correction ($p<0.00001$), or loci were monomorphic across all populations [49,50]. All loci were also tested for genotypic linkage disequilibrium (LD) in Genepop v.4.3 [51], as per Lal et al. [17]. Additionally, all loci were compared with NCBI viral and bacterial sequence databases using Basic

Local Alignment Search Tool (BLAST) searches [52], to detect contamination which may have occurred during library preparation, with all matching loci excluded from the final dataset.

Evaluation of genetic diversity, inbreeding and population differentiation

For assessment of genetic diversity within and between populations, allelic diversity indices including average observed (H_o), and average expected heterozygosities corrected for population sample size ($H_{n.b.}$) were computed. Inbreeding coefficient (F_{is}) calculation and estimation of the effective population size based on the linkage disequilibrium method (N_{eLD}) was also carried out for each population, all using Genetix v.4.05.2 [53] and NeEstimator v.2.01 [54]. Furthermore, family relationships among all individuals within sampled populations were assessed with ML-RELATE [55], which allowed for the identification of any parent-offspring, full-sib or half-sib pairs present. Relationships between individuals from different regions were also evaluated by assessing all populations together, in order to detect migration.

High levels of genome-wide polymorphism characterise many bivalves and other marine invertebrates, which may affect RADseq-based genotyping approaches by disproportionately sampling the genome due to mutations in restriction enzyme cut sites [56,57]. As previously outlined by Lal et al. [17] for *P. margaritifera*, to ascertain the potential degree of bias, F_{is} and heterozygosity were calculated for the dataset during preliminary testing at a range of missing data thresholds from 80 to 20%. These parameters were also calculated at varying read depths per stack from 5 to 15 (in the STACKs 'populations' module), before performing final F_{is} and heterozygosity computations. Heterozygosity and F_{is} changed with increasing read depth per stack from 3 to 6, however, no substantial change occurred beyond a read depth of 7. Based on these results, a final read depth threshold of 10 was selected for generating final genotypes.

To investigate individual genomic levels of diversity, multi-locus heterozygosity was examined, with the standardised heterozygosity (SH) and internal relatedness (IR) computed for each population with the *R* package *Rhh* [58,59]. Furthermore, the average multi-locus heterozygosity (Av. MLH) per population was computed manually following Slate et al. [60], along with the proportion of rare alleles with a MAF <5%. To investigate levels of population structure between sampling locations, pairwise F_{st} estimates for each population were calculated using Arlequin v.3.5.1 [48] with 10,000 permutations, and broad-scale population structure visualised by performing a Discriminant Analysis of Principal Components (DAPC) in the *R* package *adegenet* 1.4.2 [59,61–63]. The DAPC was carried out for all loci, and an α-score optimisation used to determine the number of principal components to retain. Additionally, the 'find.clusters' function of *adegenet* was utilised to determine the optimal number of actual clusters using the Bayesian Information Criterion (BIC) method.

Resolution of fine-scale population structure

To reveal any fine-scale stratification between and among all populations, network analysis was carried out using the NetView P pipeline v.0.4.2.5 [64,65]. A population network was generated based on a shared allele 1- identity-by-similarity (IBS) distance matrix created in the PLINK v.1.07 toolset [66]. The network itself is constructed with the super-paramagnetic clustering (SPC) algorithm and Sorting Points Into Neighbourhoods (SPIN) software, which computes the maximum number of nearest neighbours for a given individual [64,65,67]. The network is then visualised and edited in the Cytoscape v.2.8.3 network construction package [68]. The IBS matrix and corresponding networks were constructed at various thresholds of the maximum number of nearest neighbour (k-NN) values between 5 and 40. Additionally, a hierarchical Analysis of Molecular Variance (AMOVA) was carried out in GenAlEx v.6.5 [69], to examine variation between farmed and wild groups of populations.

Examination of adaptive variation

To detect signatures of selection, all pairwise population combinations were considered for F_{st} outlier detection. Testing failed to detect any outlier loci (see results), with the exception of three population pairs. Two independent outlier detection methods were used to identify candidate loci under selection, comprising the BayeScan v.2.1 [70,71] and LOSITAN selection detection workbench [72] packages. BayeScan 2.1 and LOSITAN employ different analytical approaches, and their joint use increased the statistical confidence of F_{st} outlier detection [16,73,74]. Jointly identified loci at high probability using both methods were considered to be statistically true outliers.

BayeScan 2.1 analyses were performed on a 1:10 prior odds probability for the neutral model and commenced with 20 pilot runs consisting of 5,000 iterations each. This was followed by 100,000 iterations with a burn-in length of 50,000 iterations [70]. Once probabilities had been calculated for each locus, the BayeScan 2.1 function *plot_R* was used in the *R* v.3.2.0 statistical package to identify putative outlier loci at various False Discovery Rates (FDR). A range of FDR values from 0.01 to 0.10 were evaluated based on preliminary testing, and recommendations by Ball [75] and Hayes [76]. All LOSITAN outlier detection was computed within a 95% confidence interval under an infinite allele model, with 50,000 iterations also evaluating a range of FDR values from 0.01 to 0.10 to match the BayeScan 2.1 analyses. All other test parameters remained at their default settings, with the exception of the 'Neutral' mean F_{st} and 'Force mean F_{st}' options being enabled.

The results of the BayeScan 2.1 and LOSITAN analyses, together with the construction of pairs of Quantile-Quantile plots (QQ-plots) were used to assess the suitability of an FDR threshold for outlier detection between the two methods. The *R* package *GWASTools* v.1.14.0 [59,77] was used to construct all QQ-plots at all FDR levels examined. All loci were included in the first QQ plot constructed, to visualise deviation outside the bounds of a 95% confidence interval. If deviation was observed, a second plot was generated excluding all outlier loci. If all remaining loci were normally distributed, this was interpreted as confirmation that true outlier loci had been detected.

Particle dispersal simulation

To independently compare results of the population genomic analyses with environmental data and to simulate larval transport pathways between sampling locations, a particle dispersal model was developed, which is publicly available at https://github.com/CyprienBosserelle/DisperGPU. Larvae typically remain in the plankton for 26–30 days prior to settlement [29,30], and due to very limited motility, are largely dispersed by current advection and turbulent diffusion in the ocean surface (mixed) layer.

Hydrodynamic and dispersal numerical models

The particle dispersal model was driven by current velocity output from the global HYbrid Coordinate Ocean Model (HYCOM) data [78,79]. HYCOM is a global hydrodynamic model that simulates ocean surface heights, currents, salinity and temperature, both at the surface and at depth. The model is driven by meteorological forcing, and constantly constrained by the assimilation of global, remote and in-situ ocean observations. As the model simulates regional and global circulation, it does not include tidal or surface wind waves. HYCOM is highly useful for forecasting and simulation experiments, with public availability at https://hycom.org. The HYCOM model had a resolution of 1/12th of a degree and output every day. Although it simulates current movement in all three dimensions, only the surface layer was used to drive the dispersal model, as this is where larvae remain in the water column [80]. The particle model used a standard Lagrangian formulation [22,23], where particles have no physical representation, but rather track the displacement of neutrally buoyant small objects such as larvae (relative to

the model resolution), at the ocean surface. Particle displacement is expressed as:

$$\Delta x = u_p * \Delta t + K \quad (1)$$

Here x represents particle position (latitude and longitude), Δx is particle displacement during a time step Δt (which was set at 1 hour), and u_p is the surface current speed at the location of the particle. K is the eddy diffusivity which takes account of the random displacement of the particle, due to turbulent eddies at a scale smaller than the hydrodynamics model resolution. K is calculated after Viikmäe et al. [81] as follows:

$$K = \sqrt{-4E_h \Delta t \log(1 - R_{NA})} \cos(2\pi R_{NB}) \quad (2)$$

Here E_h is a horizontal turbulent diffusion coefficient, and R_{NA} with R_{NB} are normally distributed random numbers. The horizontal turbulent diffusion coefficient is unknown, but assumed to be 1 m^2s^{-1}. u_p is calculated by interpolating the velocity from the hydrodynamics model, both spatially and temporally. Gridded surface currents are first interpolated to the dispersal step, after which the current velocity at each particle position is calculated using a bi-linear interpolation of the gridded surface currents, where only surface currents are taken into account and vertical movements neglected [82]. The particle age is retained and increases with simulation progression.

Model configuration

Particles were seeded in eight locations broadly corresponding to locations from where oysters were sampled for genetic analyses (see Fig 4). Seeding locations were represented at scales larger than the sampling locations to factor in the extent of surrounding coral reef habitat and farm boundaries. All seed areas were also extended further offshore to account for the fact that the HYCOM model is not adapted for shallow water environments, and does not resolve fine-scale hydrodynamic patterns <10 km [83]. At each seed location, 25,600 particles were released once at the start of the simulation, which optimised the computational requirements for running the dispersal model.

The simulation was carried out using HYCOM data for February-April 2009 and 2010, based on observations of the peak spawning period for *P. margaritifera* in Fiji [84,85], and to test for circulation pattern differences over El Niño Southern Oscillation (ENSO) event years, (2009 recorded an El Niño). Selection of this timeframe was also based upon inference of when sampled oysters were likely to be completing larval development and undergoing settlement, using shell size to approximate age [86,87]. In this way, results of both the genetic and hydrodynamic analyses were restricted to the oysters sampled.

Particle positions were extracted at time intervals of 1, 15, 30 and 60 days post-seeding and no mortality or competency behaviour of the particles was simulated. Explicit, quantitative correlation of the genetic and hydrodynamic analyses was not possible, as this would have required genetic analysis of oysters at all potential source and sink locations with dense sampling coverage, and modelling of substantially more complex particle behaviour than computational resources permitted. Instead, an independent approach was adopted here, to examine congruency of results produced by the two analyses. Although the model is unsuitable for evaluation of recruitment rates, it does allow insights into possible connectivity between sampling locations.

Results

Genotyping and SNP discovery

Following sequencing, a total of 765,273,656 PE raw reads were obtained for all nine libraries across both lanes. Read filtering using the STACKs pipeline ('process_radtags' and 'ustacks'

modules) to discard low quality reads (Phred33 score <30; 5.25% discarded), ambiguous barcodes and overrepresented sequences, resulted in 725,064,036 high quality reads remaining. These reads were used to generate a locus catalogue in the 'cstacks' module containing 303,650 stacks (S1 Table). This catalogue was used to generate all genotypes, using a median number of 555,524 reads to assemble 33,738 stacks for each individual (average read depth per stack of 17.81). Subsequent filtering at a minimum read depth of 10 per stack and MAF>0.02 resulted in a total of 42,341 genome-wide SNPs being genotyped. The primary dataset of 42,341 SNPs was screened to retain only the single most informative SNP per locus, remove those loci significantly deviating from HWE ($p<0.00001$) and under LD ($p<0.0001$) across all populations, retain individuals/populations with maximum genotyping rates, and also remove loci generated from contaminant sequences. These steps generated a final dataset of 4,123 high quality, polymorphic, genome-wide SNPs for further population genomic analyses.

Population genomic diversity and differentiation

Observed heterozygosities were significantly lower ($p<0.05$) than expected heterozygosities for all populations (H_o: 0.0621–0.1461; $H_{n.b.}$: 0.2903–0.3449, see Table 1), and displayed similar trends to the proportions of rare alleles present in each population. The individual average

Table 1. Genetic diversity indices for the wild and farmed *P. margaritifera* populations examined.

Population	Origin	n	Proportion of rare alleles (MAF <5%)	N_{eLD}[95% C. I.]	H_o(± SD)	$H_{n.b.}$(± SD)	F_{is}(p<0.01)	MLH (± SD)	SH(± SD)	IR(± SD)
Ra (Namarai)	Farm (major island; Viti Levu)	50	11.3%	658.4 [534.8–854.9]	0.1338 (±0.1261)	0.2903 (±0.1443)	0.4639	0.1407(± 0.0189)	1.1226(± 0.1623)	0.5105(± 0.0667)
Taveuni (Wailoa)	Farm (offshore island)	43	10.9%	∞[∞—∞]	0.1054 (±0.1155)	0.2943 (±0.1507)	0.5513	0.1052(± 0.0699)	0.7383(± 0.3749)	0.6733(± 0.1780)
Raviravi	Farm (major island; Vanua Levu)	32	10.4%	∞[2422.5 - ∞]	0.1353 (±0.1325)	0.2950 (±0.1488)	0.4552	0.1465(± 0.0221)	1.1414(± 0.1290)	0.4943(± 0.0813)
Savusavu (Vatubukulaca)	Farm (major island; Vanua Levu)	50	6.5%	∞[∞—∞]	0.0922 (±0.1387)	0.3151 (±0.1414)	0.5239	0.1007(± 0.0469)	0.8249(± 0.4129)	0.6760(± 0.1511)
Savusavu (Wailevu)	Farm (major island; Vanua Levu)	49	8.6%	152.4 [142.0–164.3]	0.1258 (±0.1552)	0.3062 (±0.1430)	0.4903	0.1366(± 0.0149)	1.1138(± 0.1183)	0.5567(± 0.0537)
Savusavu (Wailevu, hatchery)	Farm (major island; Vanua Levu)	50	11.4%	5.2[5.1–5.3]	0.1380 (±0.1860)	0.3063 (±0.1540)	0.4370	0.1456(± 0.0228)	1.1690(± 0.1727)	0.5713(± 0.0702)
Lau (Nayau Island)	Wild (archipelago)	50	9.8%	∞[∞—∞]	0.1093 (±0.1176)	0.2975 (±0.1476)	0.5058	0.1111(± 0.0356)	0.8899(± 0.2815)	0.6189(± 0.1246)
Yasawa (Naviti Island)	Wild (archipelago)	35	7.0%	∞[∞—∞]	0.0653 (±0.0956)	0.3113 (±0.1453)	0.6423	0.0703(± 0.0343)	0.5514(± 0.2783)	0.7613(± 0.1229)
Udu Point (Vunikodi)	Wild (major island; Vanua Levu)	18	7.4%	∞[∞—∞]	0.1461 (±0.1535)	0.3169 (±0.1468)	0.4740	0.1522(± 0.0096)	1.1609(± 0.0708)	0.4972(± 0.0337)
Kadavu (Galoa Island)	Wild (archipelago)	25	3.8%	∞ [∞—∞]	0.0673 (±0.1322)	0.3449 (±0.1380)	0.6407	0.0695(± 0.0311)	0.5361(± 0.2510)	0.7897(± 0.0950)
Kadavu (Ravitaki)	Wild (archipelago)	25	3.8%	∞ [∞—∞]	0.0621 (±0.1131)	0.3444 (±0.1398)	0.6876	0.0687(± 0.0191)	0.5498(± 0.1564)	0.7907(± 0.0584)

The parameters calculated included proportion of rare alleles (<5%), effective population size by the linkage disequilibrium method (N_{eLD}; 95% confidence intervals indicated within brackets), observed heterozygosity (H_o), average expected heterozygosity corrected for population sample size($H_{n.b.}$), inbreeding coefficient values (F_{is}), average individual multi-locus heterozygosity (MLH), standardised heterozygosity (SH) and internal relatedness (IR).

doi:10.1371/journal.pone.0161390.t001

multi-locus heterozygosity (MLH) computations matched the trends in observed heterozygosity, with the Kadavu (Ravitaki, wild) and Udu Point (wild) populations having the lowest (0.0687) and highest (0.1522) values, respectively. Lower MLH values were observed for island archipelago populations, when compared with oysters sampled from locations neighbouring larger land masses; e.g. Yasawa and the two Kadavu sites (0.0703, 0.0695 and 0.0687 respectively), vs. Ra, Raviravi and Udu Point (0.1407, 0.1465 and 0.1522, respectively). Similar patterns were apparent in the standardised heterozygosity (SH) metrics (Table 1), with island archipelago population SH values ranging from 0.5361–0.8899 (Kadavu; Galoa to Lau), and mainland populations producing values between 0.8249–1.1609 (Savusavu; Vatubukulaca to Udu Point).

Inbreeding coefficient (F_{is}) values were variable across populations (Table 1), ranging from 0.4370 for the Savusavu hatchery population, to 0.6876 for the Kadavu (Ravitaki) wild population. Interestingly, the hatchery produced Savusavu oysters demonstrated the lowest F_{is} values, whereas several wild populations, such as Yasawa (0.6423) and Taveuni (0.5513), produced higher values. Generally, slightly higher F_{is} values were observed among populations sourced from island archipelagos, e.g. Taveuni, Yasawa and the two Kadavu sites (0.5513, 0.6423, 0.6407 and 0.6876, respectively). This contrasted with estimates for oysters collected from fringing reef systems connected with the major islands of Viti Levu and Vanua Levu; e.g. Raviravi, Ra, Udu Point and Wailevu at Savusavu (0.4552, 0.4639, 0.4740 and 0.4903, respectively). Internal relatedness (IR) was comparable to the F_{is} values calculated for each respective population. The highest IR values were observed for all island populations, ranging from 0.6189 (Lau) to 0.7907 (Kadavu, Ravitaki). Among the farmed populations, the Raviravi (0.4943), Ra (0.5105), Savusavu (Wailevu; 0.5567) and Savusavu (Wailevu hatchery; 0.5713) oysters exhibited intermediate IR values, while the highest IR was recorded for oysters sampled at Savusavu (Vatubukulaca; 0.6760).

Estimates of effective population sizes were infinite for all populations (Table 1), with the exception of the Ra (658.4; [95% CI: 534–854.9]), Savusavu (Wailevu; 152.4 [95% CI: 142–164.3]) and Savusavu hatchery oysters (5.2 [95% CI: 5.1–5.3]). Pearl oysters obtained from these locations were all farmed animals, and sourced from spat collector deployments adjacent to the farm sites. The only farm sites sampled which produced infinite N_{eLD} values were Taveuni and Ra, however, most of these animals had been directly collected from reef systems adjacent to the farms themselves. The Savusavu hatchery population was found to be bottlenecked with the lowest N_{eLD} of 5.2, most likely as a result of variable family survival and broodstock contributions.

Relatedness calculations between individuals revealed no parent-offspring pairs present in the dataset (S2 Table). However, full-sib and half-sib relationships were detected for the Savusavu (Vatubukulaca) farm population (with 8 full-sib and 86 half-sib pairs), and 83 full-sib and 116 half-sib pairs identified for the Savusavu hatchery-produced oysters. When between-region relationships were assessed by examining all populations together (S3 Table), the degree of relatedness declined with increasing geographic distance. The largest number of full-sib relationships was detected between Savusavu and Lau (25), with lower numbers between Savusavu and Kadavu, Taveuni and the Yasawa archipelago respectively, (four relationships each). Higher numbers of half-sib relationships between these regions were discovered, particularly between Savusavu and Lau, Taveuni, Kadavu, the Yasawa archipelago and Raviravi (73, 37, 24, 17 and 14 respectively). Between the most distant populations sampled, only 1–2 full-sib and 1–9 half-sib relationships were detected between the Yasawa and Lau, Taveuni and Kadavu populations, respectively. However, 19 half-sib relationships were evident between both Kadavu-Lau and Kadavu-Taveuni.

Resolution of population structure

Pairwise F_{st} estimates (S4 Table) did not significantly depart from zero across almost all populations (average overall F_{st} = 0.0028; p>0.05), except for the hatchery produced oysters (Savusavu, Wailevu), which showed weak, but significant separation (p<0.000001) from four other populations: Ra (farm), Raviravi (farm), Udu Point (wild) and Savusavu, Wailevu (farm). Evaluation of population structure with a DAPC following α-score optimisation to retain 16 informative principal components (S1 Fig), revealed differentiation across two separate clusters (Fig 2). The Savusavu hatchery oysters separated from all other populations, with all remaining

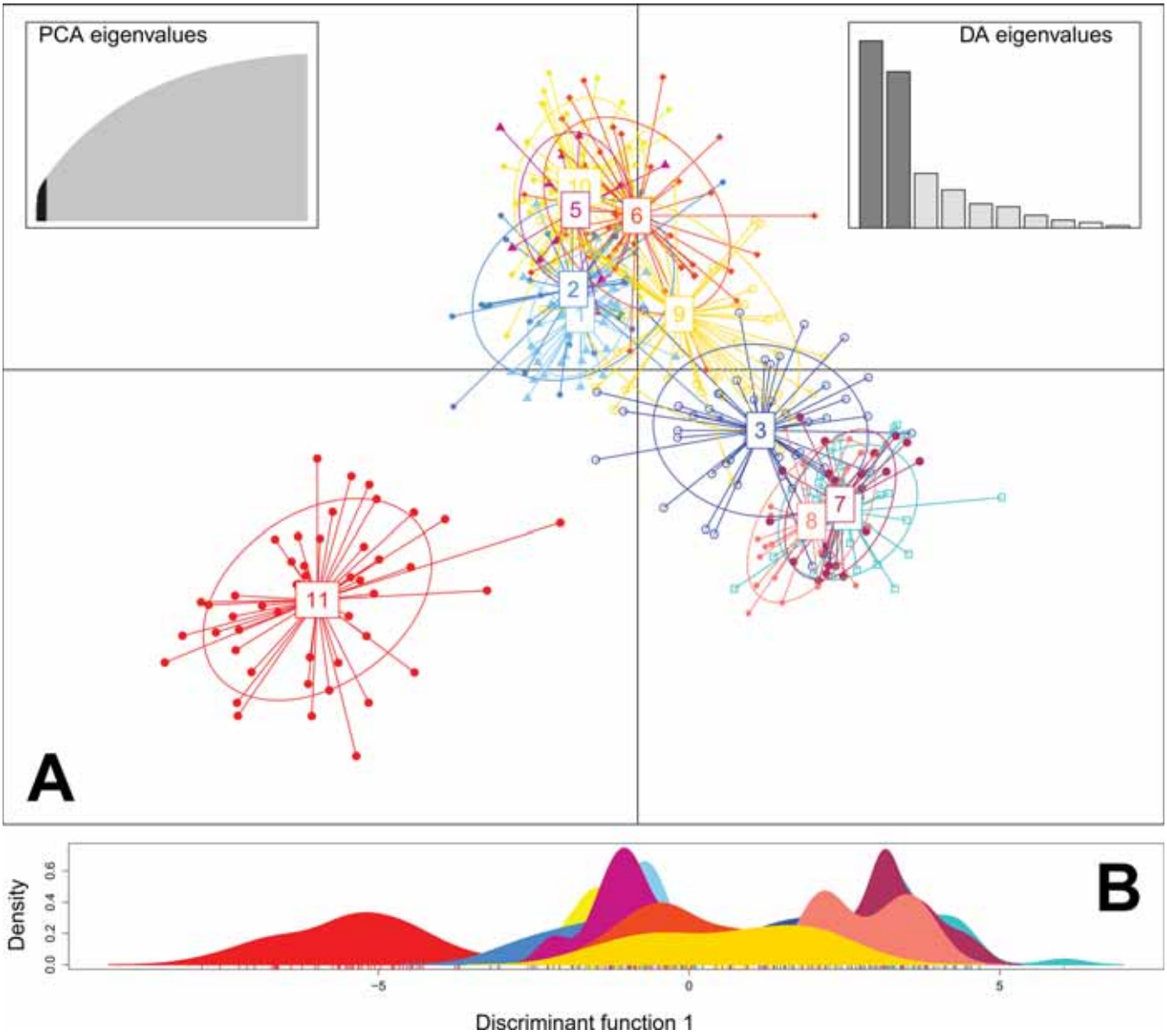


Fig 2. Discriminant Analysis of Principal Components (DAPC) scatter plot (A) and individual density plot on the first discriminant function (B), drawn across 427 *P. margaritifera* individuals in the *R* package *adegenet*. Dots represent individuals, with colours denoting sampling origin and inclusion of 95% inertia ellipses. Site colours correspond with Fig 1, and site numbers are as follows: (1) farm site at Namarai, Ra; (2) farm site at Raviravi; (3) Lau group; (4) Yasawa group; (5) Udu Point; (6) Taveuni; (7) Kadavu (Galoa Island); (8) Kadavu (Ravitaki); (9) farm site at Savusavu (Vatubukulaca); (10) farm site at Savusavu (Wailevu) and (11) farm site at Savusavu (Wailevu, hatchery produced oysters).

doi:10.1371/journal.pone.0161390.g002

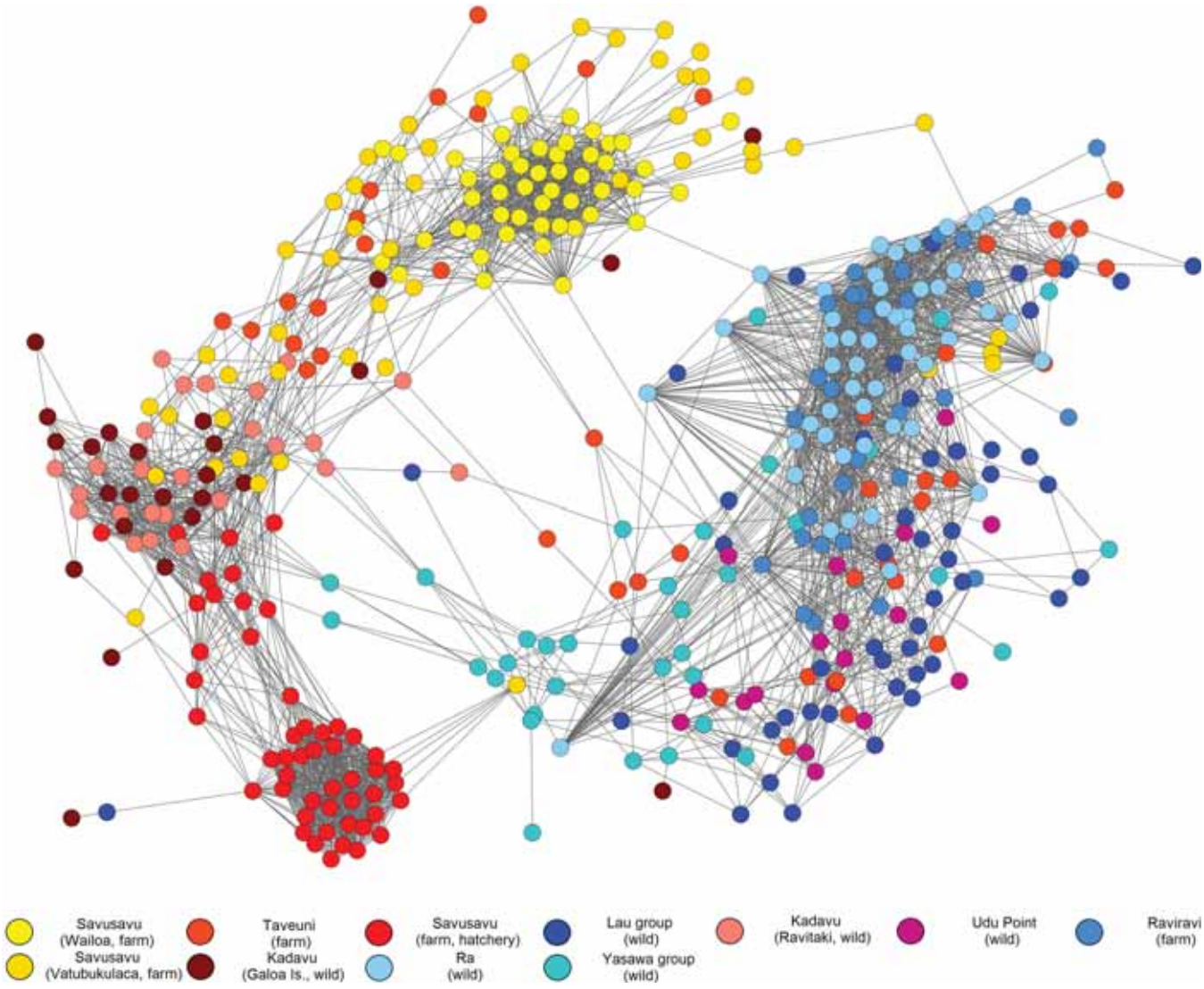


Fig 3. Population network of *P. margaritifera* individuals created using the Netview P v.0.4.2.5 pipeline after Steinig *et al*. [64]. The network has been visualised at a maximum number of nearest neighbour (k-NN) threshold of 40, using 4,123 SNPs and 427 individuals. Each dot represents a single individual, and population colours correspond with Figs 1 and 2.

doi:10.1371/journal.pone.0161390.g003

populations forming a single, diffuse cluster with overlapping 95% inertia ellipses. This separation was confirmed by testing for the actual number of discrete clusters, which was determined to be k = 2 (Bayesian Information Criterion (BIC) method; S2 Fig).

Examination of fine-scale population sub-structure using the NetView P network (Fig 3) revealed a similar pattern of separation to the DAPC analysis, although with a greater level of individual resolution. Two large genetic groups were resolved, one of which incorporated six populations, while the other comprised a diffuse assemblage of the remaining five populations. The first group included the Savusavu (Wailevu) and Savusavu hatchery oysters, which formed two distinct clusters and remained separate from all other groups. Located between these two clusters, the two Kadavu, as well as the Taveuni and Savusavu (Vatubukulaca) populations also grouped together. The second larger group contained the Ra and Raviravi populations which formed a tight assemblage, along with a less compact cluster containing the Yasawa, Lau and

Udu Point oysters. Connectivity between the two larger groups was limited to individuals belonging to the Yasawa, Taveuni, Savusavu (Vatubukulaca) and Lau populations. Identical trends were observed in networks constructed at lower k-NN values ranging from 5 to 35 (results not shown here), with the overall patterns of separation remaining consistent. Results of the hierarchical AMOVA were significant ($p<0.001$), and found that only 2% of the proportion of variation was attributable between wild and farm populations, whereas greater proportions were divided among individuals (68%), among populations (18%) and within individuals (12%).

Examination of adaptive variation

Testing failed to detect any outlier loci, with the exception of three population pairs. Detection of F_{st} outlier loci at three FDR thresholds of 0.01, 0.05 and 0.10 for each of the pairwise population comparisons discovered between two and nine directional outlier SNPs jointly identified by Bayescan 2.1 and LOSITAN (Table 2). These pairwise population comparisons were carried out between Savusavu (Wailevu) and Lau, Udu Point and Kadavu (both populations considered together), as well as the Yasawa archipelago and Lau. These sites were located at maximum geographic distances across the Fiji Islands, positioned across environmental gradients (offshore island vs. mainland island and fringing vs. barrier reef habitats), as well as at opposing points along the major larval transport pathway identified from the particle dispersal simulation analysis. All directional outliers detected by Bayescan were also detected by LOSITAN, and no outlier loci were detected by either platform when all populations were considered together. Bayescan 2.1 analyses failed to detect any balancing outlier loci (zero or negative alpha values) for all pairwise population comparisons, and hence all balancing outliers reported were from LOSITAN computations. LOSITAN runs detected between 43 and 278 balancing loci across all three FDR thresholds for each pairwise population comparison. In order to select an FDR threshold for accepting a final number of outlier loci for each comparison, QQ plots were constructed for each dataset at all three thresholds. A final stringent FDR threshold of 0.01 was selected on the basis of the QQ plots (S3 Fig), under which 5, 3 and 2 directional outlier loci were detected between the Savusavu (Wailevu)-Lau, Udu Point-Kadavu and Yasawa-Lau pairwise population comparisons, respectively.

To gauge the strength of the selection signal, the average F_{st} values for all directional and balancing outlier loci detected were examined at the selected FDR of 0.01. For the Savusavu (Wailevu)-Lau comparison, the average Bayescan 2.1 F_{st} value was 0.1168. Similarly, average F_{st} values of 0.1025 and 0.1496 were observed for the Yasawa-Lau, and Udu-Kadavu comparisons, respectively. The average LOSITAN F_{st} values for the balancing outliers detected

Table 2. Numbers of putative directional and balancing F_{st} outlier loci discovered. Tests were carried out at three False Discovery Rate (FDR) thresholds using BayeScan 2.1 [70] and LOSITAN [72]. Jointly-identified loci were identified using both outlier detection platforms.

		Directional			Balancing		
Populations compared	**FDR**	**BayeScan 2.1**	**LOSITAN**	**Jointly-identified**	**BayeScan 2.1**	**LOSITAN**	**Jointly-identified**
Savusavu, (Wailevu) and Lau	0.01	5	28	5	0	197	0
	0.05	8	46	8	0	206	0
	0.10	9	96	9	0	248	0
Udu Point and both Kadavu populations	0.01	3	21	3	0	43	0
	0.05	3	37	3	0	108	0
	0.10	4	56	4	0	84	0
Yasawa and Lau	0.01	2	18	2	0	201	0
	0.05	3	46	3	0	278	0
	0.10	4	61	4	0	241	0

doi:10.1371/journal.pone.0161390.t002

remained consistent for the Savusavu (Wailevu)-Lau, Udu-Kadavu and Yasawa-Lau comparisons, (-0.0343, -0.0464 and -0.0426, respectively). Given this set of results, it appears that any signatures of selection if present, are too weak to be detected and/or indecipherable from the background signal. This was supported by contruction of neighbour joining trees to visualise population structure using directional outlier loci identified for each pairwise population comparison, based on 1-proportion of shared allele distances (results not shown here). All trees failed to show any separation between populations.

Particle dispersal modelling

Simulation of larval transport pathways with the particle dispersal model demonstrated broad-scale mixture of larvae by surface ocean current systems operating within the Fiji Islands; (see Fig 4 for 2009 particle position outputs at 1, 15, 30 and 60 day time points and S4 Fig for an animation of the full dispersal simulation over 100 days. 2010 data were very similar to 2009 patterns and are not presented here). A singular dispersal corridor appears to initially drive

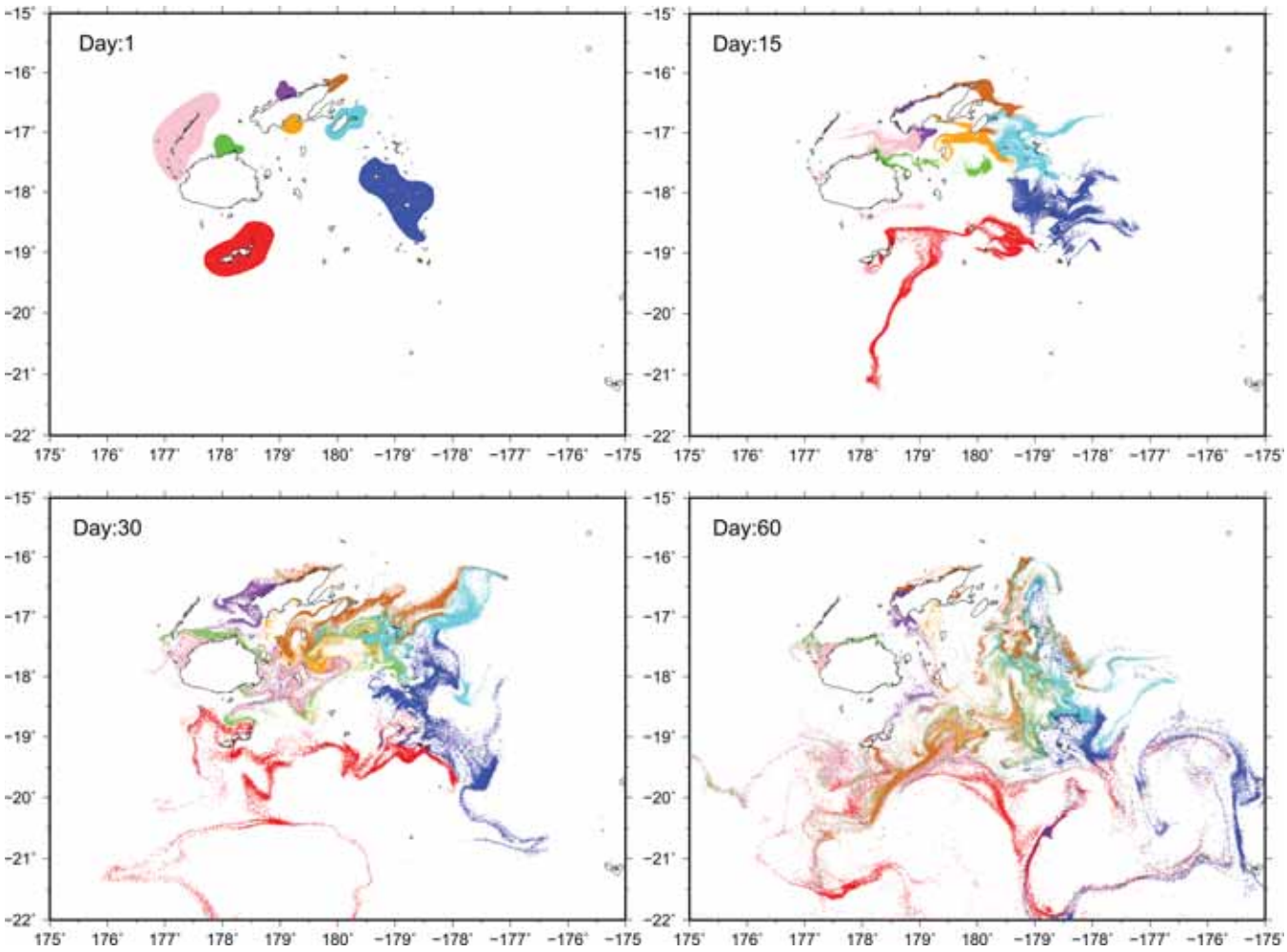


Fig 4. Results of 2009 particle dispersal simulation. Particle seed locations are shown in the day 1 position output, with the sampling regions colour coded as follows: Kadavu group (red), Yasawa group (pink), Ra (green), Raviravi (purple), Savusavu (orange), Udu Point (brown), Taveuni (light blue) and the central Lau group (dark blue). Simulated particle positions are shown at 15, 30 and 60 day outputs. An animation of dispersal simulation is provided as S4 Fig.

doi:10.1371/journal.pone.0161390.g004

larvae from all seed locations eastwards towards the Lau group of islands for a period of approximately 30 days; after which current movements oscillate across the centre of the Fiji group, while progressing in a southerly direction. Gene flow thus is likely to be homogenous between the Yasawa archipelago, Raviravi and Udu Point through the Bligh Water channel, towards sink locations in the Koro and Lau basins. Reef systems in the Lau group appear to receive recruits from all locations in Fiji, although varying degrees of self-recruitment are likely for the Udu Point, Raviravi and Yasawa populations, due to the prevailing current dynamics and architecture of the Great Sea Reef system north of Vanua Levu retaining larvae in those regions. Despite this, a portion of larvae originating in the Yasawa archipelago appear to recruit along the western coastline of Viti Levu and Ra. Similarly, larvae which are exported from Udu Point and Raviravi may mix with individuals from Savusavu and Taveuni. The lowest degree of mixing is likely to occur between populations located along a North-South axis (e.g. Udu Point and Kadavu), as the dominant dispersal pathway operates in a West to East direction. Interestingly, the simulation indicates that if larvae advected from Kadavu and Lau survive beyond 40 days post-hatching, it may be possible for a few individuals to recruit eastwards onto the reefs of Tongatapu in the Kingdom of Tonga, (approximate position -175° longitude; see Day 60 output in Fig 4).

Discussion

By independently evaluating population genomic analyses with hydrodynamic dispersal simulation, we identified that Fijian *P. margaritifera* display a very shallow pattern of population structure, and are highly likely to constitute a single, biologically significant stock for fishery management. While diffuse patterns of population differentiation are apparent given the resolution of 4,123 SNPs used, the overall pairwise F_{st} estimates are small and not statistically significant (average overall $F_{st} = 0.0028$; $p>0.05$). Given the largely homogenising larval mass transport pattern resolved using hydrodynamic dispersal simulation and the levels of relatedness between populations, the pattern of structure detected plausibly reflects fine-scale differentiation at the generational and family levels, together with small, isolated patches of localised recruitment [32]. Furthermore, examination of loci under selection failed to detect any signatures of local adaptation, suggesting that environmental differences among populations are insufficiently heterogeneous to drive selection at the spatial scale examined (<400 km). Additionally, if weak local adaptation is present, the very high levels of gene flow between populations would likely override discernible signatures of selection. These results demonstrate the utility of independent population genomic and biophysical datasets for providing insights into the biology and ecology of a broadcast spawning bivalve, and have great potential for application to other marine species with similar life histories, where patterns of genetic structure and connectivity may not be well understood.

Resolution of population structure, diversity and relatedness

A weak pattern of population structure with high levels of connectivity was evident among all populations sampled using both broad-scale (DAPC) and fine-scale (NetView P) methods, mirroring the results of a previous study in Fiji [17]. Investigations of *P. margaritifera* populations elsewhere have yielded similar results, including French Polynesia [31,32] and Japan [88]. Considering that *P. margaritifera* is a broadcast spawner with a relatively long PLD of between 26–30 days [29,30], the degree of larval mixing driven by surface ocean currents (as demonstrated by the hydrodynamic dispersal simulation), supports the finding that Fijian oysters from all 11 locations sampled may be classified as a singular genetic entity.

Population pairwise F_{st} estimates indicated shallow and non-significant levels of structure, with the hatchery-produced oysters being the only population demonstrating detectable

differentiation. This is not surprising considering that this population had undergone a genetic bottleneck through limited broodstock use, and differential larval mortality typical of hatchery rearing conditions. DAPC with BIC analysis, and NetView P network analysis both resolved similar cluster patterns, and overall patterns correlated well with F_{st} results and larval transport pathways inferred from particle dispersal simulation.

The levels of observed heterozygosity (H_o) detected were lower than expected across all populations (Table 1), keeping with the trend of heterozygote deficiency previously observed for *P. margaritifera* in Fiji [17], French Polynesia [31–34,89] and Japan [88]. Heterozygote deficits appear to be characteristic of a number of marine molluscs [90–92], and in the current study are also likely due to a technical artefact associated with RADseq-based genotyping approaches, where restriction enzyme cut site polymorphisms may cause allelic dropouts [56,57]. While stringent filtering measures were used to reduce the proportion of null alleles present in the final dataset, thorough testing of their effect on H_o, F_{is}, N_{eLD} and population differentiation estimates following the methods of Lal et al. [17] for *P. margaritifera*, revealed no impact on these metrics.

When assessing populations separately, estimates of individual average multi-locus heterozygosity (MLH), standardised heterozygosity (SH), inbreeding coefficient (F_{is}) and internal relatedness (IR) agreed with trends observed in H_o, which generally showed a lower diversity among pearl oysters sampled from island archipelago populations, compared to those from the larger land masses of Viti Levu and Vanua Levu (e.g. Av.MLH for the Kadavu (Galoa Island) and Raviravi (Vanua Levu) populations were 0.0695 cf. 0.1465 respectively). This observation may indicate higher rates of self-recruitment among island archipelago populations, and fits a growing body of evidence supporting significant self-recruitment for a number of broadcast spawning coral and reef fish species, with geographic setting strongly influencing the degree of larval retention within populations [93].

Patterns detected in the NetView P network, relatedness analyses and dispersal simulation indicate support for this observation, as geographically distant populations clustered separately (e.g. Kadavu and Taveuni island sites), and shared fewer pairwise family relationships than others with higher degrees of connectivity either through proximity (e.g. Ra and Raviravi), or position within the major current pathway (e.g. Yasawa and Lau). This was particularly evident between populations <150 Km apart containing 17–73 half-sibs, whereas populations situated farther apart held only 1–9. Examination of pairwise relationships between individuals within populations identified a larger number of full-sib and half-sib relationships for the bottlenecked hatchery produced population, as well as one farmed population sourced from spat collectors. For the latter, it is feasible that several individuals from one or more families remained poorly mixed in the plankton, and subsequently settled together on the spat collectors. This was suggested by Knutsen et al. [94] for their study on Atlantic cod, and similar variability has been observed in hatchery-produced *P. maxima* [90,95].

Assessments of N_{eLD} and individual pairwise relationships within populations indicated a generally high degree of connectivity between populations. However, reduced N_{eLD} was detected for three farmed populations, one of which was a hatchery-produced cohort that had experienced a genetic bottleneck as a result of standard hatchery spawning practices [17,88,90,95]. A possible explanation for the lower N_{eLD} observed for the two other populations may be differential settlement and survival on the spat collectors these oysters were collected from, as previous studies have shown highly variable settlement, survival and predation rates of newly settled *P. margaritifera* spat on collector gear [96–99].

The use of hydrodynamic modelling in parallel with genome-wide data for farmed and wild populations, adds fresh perspective for understanding the interaction of geographic and oceanographic influences contributing to population genetic structure in *P. margaritifera*. Studies on

the genetic stock structure of this species predominantly originate in French Polynesia, where oysters are found in three distinct types of reef environments [31,34]. These comprise high island lagoons with fringing and barrier reef systems with open oceanic circulation (similar to those found in Fiji), atoll lagoons also with open circulation, and closed atoll lagoons with highly reduced circulation [31,32,34]. Lemer and Planes [31] detected connectivity at both small (less than 500 km) and large (greater than 1500 km) spatial scales between French Polynesian archipelagos which had open oceanic circulation patterns, mirroring the results of our observations for Fijian populations, but also found significant genetic structure for oysters contained within closed atoll lagoons.

Examination of adaptive variation

Understanding levels of adaptive variation is critical for management of translocation, population supplementation and/or assisted migration, in order to avoid negative consequences such as outbreeding depression that may result from moving individuals into an environment they may be maladapted to [100,101]. This latter consideration is especially important for aquaculture, as productivity is heavily reliant on stock fitness [102–104]. Knutsen et al. [94] in their study on Atlantic cod also failed to detect signatures of selection, despite the species having an extensive North Atlantic natural distribution over known salinity and temperature clines. An explanation they offer for this finding is that their work examined a restricted geographical range, where environmental differences may be small, relative to conspecifics occupying more heterogeneous habitats over the broader species distribution. The situation may be similar for *P. margaritifera* in the present study, and examination of populations across larger spatial scales beyond the Fiji Islands should provide further insights.

The inability of F_{st} outlier testing to discern signatures of selection possibly indicates that the environments oysters were sampled from may be insufficiently heterogeneous to drive local adaptation at an easily detectable threshold. Further considerations include the type of trait under selection (polygenic or monogenic), as well as the opposing dynamics of gene flow against the strength of selection. That is, where local adaptation is present, it may be too weak to be detected by the SNP marker set used and lost to background noise. Nayfa and Zenger [11] examined three populations of the closely related silver-lip pearl oyster *P. maxima*, from Bali, West Papua and Aru in Indonesia, which were subject to a complex system of prevailing and seasonally reversing surface ocean currents. Evidence of directional selection was detected despite high levels of gene flow, causing divergence between oysters from Bali and West Papua against those from Aru, and the recommendation for aquaculture was to manage the Aru population separately from Bali and West Papua.

Particle dispersal modelling

Examination of larval dispersal patterns using hydrodynamic modelling alone has been used for a number of marine taxa [105,106], including *P. margaritifera* [107], but comparatively few studies have sought to combine larval dispersal data with genome-wide population information. Among studies which have coupled oceanographic and genetic methods are White et al. [108], Galindo et al. [21] and Dao et al. [24] using microsatellite loci, however, the limited number of these markers have provided finite information about fine-scale population structure and adaptive variation [109,110].

The discovery of homogenised surface ocean current movement towards the Lau archipelago is well supported by the results of population genomic analyses presented here, particularly regarding broad and fine-scale population differentiation, genetic diversity levels and lack of adaptive variation within and among populations. It is interesting that the major larval sink

location is situated in the Lau archipelago, which retains consistency across ENSO years. Further examination of fine-scale larval transport pathways is warranted to determine the degree of mixing within the Lau group, and to see if any settlement heterogeneity occurs there. Unfortunately, this was beyond the capability of the HYCOM hydrodynamic model used here, as the data is not captured at a resolution finer than a grid size of 10 km^2 [79,83]. The HYCOM model is the only hydrodynamic model available for the Fiji Islands, however, given the future availability of a finer resolution model, gaining these insights is possible.

For broadcast spawning marine taxa with extended PLD, the inclusion of hydrodynamic dispersal data to better understand population connectivity in the marine environment is indispensable, as assessment of the magnitude of larval movements, along with patterns of current-driven differential recruitment may become possible. Work by Thomas et al. [107] in French Polynesia on connectivity between populations discovered that larval sink and source locations for *P. margaritifera* accounted for 26% and 59% of the variation observed respectively, underscoring its importance for larval supply and management of farmed and wild pearl oysters.

Implications for fishery management

The persistent problem in stock assessment investigations of determining "biologically meaningful" genetic divergence between populations requires careful evaluation on a case by case basis, with respect to the biological questions being answered [3], fishery management goals and the characteristics of the organism(s) involved [4,94]. For high gene flow species where fine-resolution population genomic analyses detect weak divergence by examining neutral and adaptive variation, the use of independent environmental data provides important additional knowledge for informed fishery management decision making.

Given the findings of non-significant population differentiation and the absence of signatures of selection or apparent phenotypic differences among populations, these data support the existence of a singular, biological stock in the Fiji Islands. This suggests that fishery management of *P. margaritifera* in Fiji may be based upon treatment of all populations sampled here as one cohesive unit. Further evidence of this is found in the independent assessment of population connectivity by hydrodynamic dispersal simulation, which confirms broad scale panmixia across all populations. This finding is promising for developing aquaculture of this species in the country, as it may mean that spat collected in locations which freely exchange recruits can also be grown-out among them (e.g. Kadavu, Ra, Savusavu, Taveuni and Lau). For those populations which experience less connectivity (e.g. Yasawa, Raviravi and Udu Point), further investigation is required to determine if any negative consequences may result from either keeping these groups isolated, or opening them up to translocation.

The small spatial scale of the Fiji Islands and high levels of gene flow apparent for Fijian *P. margaritifera*, may actually facilitate uncomplicated fishery management and aquaculture development of this species in the country, compared to other locations such as French Polynesia, where oysters are distributed over larger scales and across heterogeneous habitats [31]. For French Polynesian populations, Lemer and Planes [34] and Arnaud-Haond et al. [33] reported that farmed populations originally sourced from genetically distinct wild oysters over a period of 20 years, had accumulated higher levels of genetic diversity than their progenitors, potentially providing a risk of outbreeding depression for wild oysters interbreeding with farmed individuals. While it is unlikely that a similar situation could occur for Fijian *P. margaritifera*, there are important lessons to be learnt from the French Polynesian experience. If hatchery production of spat outpaces the collection of wild spat as the primary source of oysters for grow out in the future, any potentially negative consequences as a result of genetic pollution effects could be minimised by careful selection of broodstock to maintain levels of genetic fitness.

Conclusions

The use of genome-wide SNP data and hydrodynamic particle dispersal modelling have provided valuable insights into the population structure and connectivity of the black-lip pearl oyster in the Fiji Islands, filling a substantial knowledge gap on the stock structure of this species in the country. Simulation of larval transport with hydrodynamic dispersal modelling confirmed the existence of broad-scale connectivity by surface ocean current systems, correlating very well with patterns of differentiation, heterozygosity and adaptive variation discovered in the genetic data. There is strong support for the existence of a singular stock structure in the Fiji Islands, which is promising for developing aquaculture of this species in the country, as it indicates that germplasm transfer is possible between locations which freely exchange recruits. The combined use of both selectively neutral and loci under selection to elucidate fine-scale population variability (or the lack thereof), has high utility for stock assessment in high gene flow species, where biologically meaningful levels of divergence are not immediately apparent. Furthermore, independent assessment of connectivity using environmental data such as particle dispersal simulation, can provide valuable additional information for making fishery management decisions, when patterns in genetic data don't easily lend themselves to the identification of stock boundaries. Our study highlights the value of using both genomic and hydrodynamic data, for a comprehensive understanding of population structure and connectivity in broadcast-spawning marine taxa, and utilising the information collectively for aquaculture and sustainable fishery management.

Supporting Information

S1 Fig. α-score optimisation graph for generation of the Discriminant Analysis of Principal Components (DAPC) scatter plot. An optimal number of 16 principal components were suggested for retention using this analysis, based on 4,123 SNP loci in the *R* package *adegenet* [59,62,63].
(PDF)

S2 Fig. Determination of the number of clusters following generation of the DAPC scatter plot using 4,123 SNP loci. An optimal number of k = 2 was suggested based on the BIC method implemented in the find.clusters function of the *R* package *adegenet* [59,61,62].
(PDF)

S3 Fig. Verification of outlier loci detected for population pairwise comparisons using Quantile-Quantile plots (QQ plots) at an FDR of 0.01. Comparisons shown are for Savusavu-Lau (left), Udu Point-Kadavu (middle) and Yasawa-Lau (right). QQ plots are arranged in pairs with the top row displaying the p value distributions of all SNP loci while the bottom row displays the distribution when all outlier loci are removed. The red line indicates y = x linearity for conformity to a normal distribution, with the surrounding grey area approximating a 95% confidence interval.
(TIF)

S4 Fig. Animation of full particle dispersal model simulation run for 2009 over 100 days. Particle seed location colour codes are identical to those described in Fig 4. [See.GIF file. Please note that the.GIF file needs to be opened in a web browser to display correctly.]
(GIF)

S1 Table. Sequencing recovery rates and SNP identification at each filtering step in the STACKs 1.20 pipeline.
(DOCX)

S2 Table. Estimates of relationships. Values are provided for relationships between individuals within eleven Fijian populations of *P. margaritifera*, with 4,123 SNP loci using ML-RELATE [55]
(DOCX)

S3 Table. Estimates of full-sib, half-sib and parent-offspring relationships. Estimates are provided between geographic regions sampled from eleven Fijian populations of *P. margaritifera* with 4,123 SNP loci using ML-RELATE [55]. All other between-region relationships examined indicated that individuals were unrelated.
(DOCX)

S4 Table. Population pairwise F_{st} estimates. Estimates were computed using Arlequin [48] (Weir and Cockerham 1984 unbiased method), for 4,123 SNP loci in *P. margaritifera* from 11 Fijian populations. Significantly different values at $p<0.000001$ following 10,000 permutations are indicated with an asterisk.
(DOCX)

Acknowledgments

This study was conducted within the Australian Centre for International Agricultural Research (ACIAR) Project FIS/2009/057: "Pearl Industry Development in the Western Pacific" led by the University of the Sunshine Coast. The research was carried out during a John Allwright Fellowship awarded to MML. We wish to thank Eike Steinig, Shannon Kjeldsen, Maria Nayfa and Roger Huerlimann for advice on various statistical analyses including the Netview P pipeline, Linux scripting and outlier analyses. We also thank two anonymous reviewers for their evaluation of the manuscript, Carolyn Smith-Keune and Georgia McDougall for their assistance with optimisation of the ddRADseq methodology, as well as Litia Gaunavou for generation of the sampling site map. For either providing or assisting with collection of oyster tissue samples, our gratitude also extends to Laisiasa Cavakiqali, Aisea Titoko, Emosi Titoko, Cherie Morris, Shirleen Bala, Justin Hunter, Chris O'Keefe, Sachin Deo, Pranesh Kishore, Adi Dionani Salaivanua, Kelly Brown, Jerome Taoi, Epeli Loganimoce, Albert Whippy, Bai Whippy, Toga Whippy, Isimeli Loganimoce, Marilyn Vilisoni, Babitu Rarawa, Ilitomasi Nuku, Samisoni Rakai, Patrick Fong, Nepoci Raleve and Claude Prévost. Logistical support for fieldwork in the Fiji Islands was provided by project partners the Secretariat of the Pacific Community (SPC) and the University of the South Pacific (USP).

Author Contributions

Conceptualization: MML PCS KRZ.

Data curation: MML CB KRZ.

Formal analysis: MML CB KRZ.

Funding acquisition: PCS.

Investigation: MML CB.

Methodology: MML CB KRZ.

Project administration: MML PCS KRZ.

Resources: MML PCS DRJ CB KRZ.

Software: MML CB.

Supervision: PCS DRJ KRZ.

Validation: MML CB KRZ.

Visualization: MML CB.

Writing – original draft: MML.

Writing – review & editing: MML PCS DRJ CB KRZ.

References

1. Ward RD (2000) Genetics in fisheries management. Hydrobiologia 420: 191–201.
2. Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. Journal of Heredity 89: 438–450.
3. Reiss H, Hoarau G, Dickey-Collas M, Wolff WJ (2009) Genetic population structure of marine fish: mismatch between biological and fisheries management units. Fish and Fisheries 10: 361–395.
4. Waples RS, Punt AE, Cope JM (2008) Integrating genetic data into management of marine resources: how can we do it better? Fish and Fisheries 9: 423–449.
5. Carvalho GR, Hauser L (1994) Molecular genetics and the stock concept in fisheries. Reviews in Fish Biology and Fisheries 4: 326–350.
6. Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular Ecology 15: 1419–1439. PMID: 16629801
7. Gaggiotti OE, Bekkevold D, Jørgensen HBH, Foll M, Carvalho GR, Andre C, et al. (2009) Disentangling the Effects of Evolutionary, Demographic, and Environmental Factors Influencing Genetic Structure of Natural Populations: Atlantic Herring as a Case Study. Evolution 63: 2939–2951. doi: 10.1111/j.1558-5646.2009.00779.x PMID: 19624724
8. Hauser L, Baird M, Hilborn RAY, Seeb LW, Seeb JE (2011) An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. Molecular Ecology Resources 11: 150–161. doi: 10.1111/j.1755-0998.2010.02961.x PMID: 21429171
9. Palumbi SR (2003) Population Genetics, Demographic Connectivity, and the Design of Marine Reserves. Ecological Applications 13: S146–S158.
10. Weersing K, Toonen RJ (2009) Population genetics, larval dispersal, and connectivity in marine systems. Marine Ecology Progress Series 393: 1–12.
11. Nayfa MG, Zenger KR (2016) Unravelling the effects of gene flow and selection in highly connected populations of the silver-lip pearl oyster (*Pinctada maxima*). Marine Genomics.
12. André C, Larsson LC, Laikre L, Bekkevold D, Brigham J, Carvalho GR, et al. (2011) Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. Heredity 106: 270–280. doi: 10.1038/hdy.2010.71 PMID: 20551979
13. Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries 9: 333–362.
14. Palumbi SR (1994) Genetic Divergence, Reproductive Isolation, and Marine Speciation. Annual Review of Ecology and Systematics 25: 547–572.
15. Limborg MT, Helyar SJ, De Bruyn M, Taylor MI, Nielsen EE, Ogden ROB, et al. (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). Molecular Ecology 21: 3686–3703. doi: 10.1111/j.1365-294X.2012.05639.x PMID: 22694661
16. Pujolar JM, Jacobsen MW, Als TD, Frydenberg J, Munch K, Jónsson B, et al. (2014) Genome-wide single-generation signatures of local selection in the panmictic European eel. Molecular Ecology 23: 2514–2528. doi: 10.1111/mec.12753 PMID: 24750353
17. Lal MM, Southgate PC, Jerry DR, Zenger KR (2016) Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. Marine Genomics 25: 57–68. doi: 10.1016/j.margen.2015.10.010 PMID: 26545807
18. Liggins L, Treml EA, Riginos C (2013) Taking the Plunge: An Introduction to Undertaking Seascape Genetic Studies and using Biophysical Models. Geography Compass 7: 173–196.

19. Cannuel R, Beninger PG, McCombie H, Boudry P (2009) Gill Development and Its Functional and Evolutionary Implications in the Blue Mussel *Mytilus edulis* (Bivalvia: Mytilidae). Biological Bulletin 217: 173–188. PMID: 19875822

20. Berry O, England P, Marriott RJ, Burridge CP, Newman SJ (2012) Understanding age-specific dispersal in fishes through hydrodynamic modelling, genetic simulations and microsatellite DNA analysis. Molecular Ecology 21: 2145–2159. doi: 10.1111/j.1365-294X.2012.05520.x PMID: 22417082

21. Galindo HM, Olson DB, Palumbi SR (2006) Seascape Genetics: A Coupled Oceanographic-Genetic Model Predicts Population Structure of Caribbean Corals. Current Biology 16: 1622–1626. PMID: 16920623

22. Siegel DA, Kinlan BP, Gaylord B, Gaines SD (2003) Lagrangian descriptions of marine larval dispersion. Marine Ecology Progress Series 260: 83–96.

23. Siegel DA, Mitarai S, Costello CJ, Gaines SD, Kendall BE, Warner RR, et al. (2008) The stochastic nature of larval connectivity among nearshore marine populations. Proceedings of the National Academy of Sciences of the United States of America 105: 8974–8979. doi: 10.1073/pnas.0802544105 PMID: 18577590

24. Dao HT, Smith-Keune C, Wolanski E, Jones CM, Jerry DR (2015) Oceanographic Currents and Local Ecological Knowledge Indicate, and Genetics Does Not Refute, a Contemporary Pattern of Larval Dispersal for The Ornate Spiny Lobster, *Panulirus ornatus* in the South-East Asian Archipelago. PLoS One 10.

25. Gosling EM (2015) Fisheries and management of natural populations. In: Gosling EM, editor. Marine Bivalve Molluscs. West Sussex, United Kingdom: Wiley Blackwell. pp. 270–324.

26. Wada KT, Tëmkin I (2008) Taxonomy and Phylogeny: Commercial Species. In: Southgate PC, Lucas JS, editors. The Pearl Oyster. Amsterdam, the Netherlands: Elsevier. pp. 54–57.

27. Hare PM, Palumbi RS, Butman AC (2000) Single-step species identification of bivalve larvae using multiplex polymerase chain reaction. Marine Biology 137: 953–961.

28. Wada KT, Jerry DR (2008) Population Genetics and Stock Improvement. In: Southgate PC, Lucas JS, editors. The Pearl Oyster. Amsterdam, the Netherlands: Elsevier. pp. 437–471.

29. Alagarswami K, Dharmaraj S, Chellam A, Velayudhan TS (1989) Larval and juvenile rearing of blacklip pearl oyster, *Pinctada margaritifera* (Linnaeus). Aquaculture 76: 43–56.

30. Doroudi MS, Southgate PC (2003) Embryonic and larval development of *Pinctada margaritifera* (Linnaeus, 1758). Molluscan Research 23: 101–107.

31. Lemer S, Planes S (2014) Effects of habitat fragmentation on the genetic structure and connectivity of the black-lipped pearl oyster *Pinctada margaritifera* populations in French Polynesia. Marine Biology 161: 2035–2049.

32. Arnaud-Haond S, Vonau V, Rouxel C, Bonhomme F, Prou J, Goyard E, et al. (2008) Genetic structure at different spatial scales in the pearl oyster (*Pinctada margaritifera cumingii*) in French Polynesian lagoons: beware of sampling strategy and genetic patchiness. Marine Biology 155: 147–157.

33. Arnaud-Haond S, Vonau V, Bonhomme F, Boudry P, Prou J, Seaman T, et al. (2003) Spat collection of the pearl oyster (*Pinctada margaritifera cumingii*) in French Polynesia: an evaluation of the potential impact on genetic variability of wild and farmed populations after 20 years of commercial exploitation. Aquaculture 219: 181–192.

34. Lemer S, Planes S (2012) Translocation of wild populations: conservation implications for the genetic diversity of the black-lipped pearl oyster *Pinctada margaritifera*. Molecular Ecology 21: 2949–2962. doi: 10.1111/j.1365-294X.2012.05588.x PMID: 22548374

35. SPC (2003) Profiles of high interest aquaculture commodities for Pacific Islands countries. Noumea, New Caledonia: Secretariat of the Pacific Community. 71 p.

36. SPC (2007) SPC Aquaculture Action Plan 2007. Noumea, New Caledonia: Secretariat of the Pacific Community. 31 p.

37. Southgate PC, Strack E, Hart A, Wada KT, Monteforte M, Cariño M, et al. (2008) Exploitation and Culture of Major Commercial Species. In: Southgate PC, Lucas JS, editors. The Pearl Oyster. Amsterdam, the Netherlands: Elsevier. pp. 303–355.

38. Passfield K (1995) Report of a pearl oyster survey of W. Vanua Levu, Beqa, Totoya and Makogai, in the Fiji Islands (Draft). Suva, Fiji Islands: Fiji Fisheries Division. 1–24 p.

39. Friedman K, Kronen M, Vunisea A, Pinca S, Pakoa K, Magron F, et al. (2010) Fiji Islands Country Report: Profiles and Results from Survey Work at Dromuna, Muaivuso, Mali and Lakeba. (September to November 2002, April to June 2003, June and July 2007, and February 2009). Noumea, New Caledonia: Secretariat of the Pacific Community. 978-982-00-0399-6 978-982-00-0399-6. 503 p.

40. Southgate PC (2008) Pearl Oyster Culture. In: Southgate PC, Lucas JS, editors. The Pearl Oyster. Amsterdam, the Netherlands: Elsevier. pp. 231–272.

41. Dawson MN, Raskoff KA, Jacobs DK (1998) Field preservation of marine invertebrate tissue for DNA analyses. Mol Mar Biol Biotechnol 7: 145–152. PMID: 11541322

42. Adamkewicz SL, Harasewych MG (1996) Systematics and biogeography of the genus *Donax* (Bivalvia: Donacidae) in eastern North America. American Malacological Bulletin 13: 97–103.

43. GE (2007) Illustra AutoSeq G-50 and AutoSeq 96 dye terminator removal. Data file 28-9175-28. Illustra AutoSeq G-50 documents. Buckinghamshire, United Kingdom: GE Healthcare UK Limited. pp. 1–4.

44. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLoS ONE 7: e37135. doi: 10.1371/journal.pone.0037135 PMID: 22675423

45. Catchen J, Bassham S, Wilson T, Currey M, O'Brien C, Yeates Q, et al. (2013) The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. Molecular Ecology 22: 2864–2883. doi: 10.1111/mec.12330 PMID: 23718143

46. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. G3: Genes, Genomes, Genetics 1: 171–182.

47. Huang H, Knowles LL (2014) Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation Study of RAD Sequences. Systematic Biology.

48. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47–50.

49. Zenger KR, Khatkar MS, Cavanagh JA, Hawken RJ, Raadsma HW (2007) Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into Australian and global population variability, including impact of selection. Animal Genetics 38: 7–14. PMID: 17257182

50. Zenger KR, Khatkar MS, Tier B, Hobbs M, Cavanagh JAL, Solkner J, et al. QC analyses of SNP array data: experience from a large population of dairy sires with 23.8 million data points Proceedings of the Association for the Advancement of Animal Breeding and Genetics 2007; 17th Association for the Advancement of Animal Breeding and Genetics Conference 2007, 23–26 September 2007, Armidale, NSW, Australia. Association for the Advancement of Animal Breeding and Genetics pp. 123–126.

51. Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources 8: 103–106. doi: 10.1111/j.1471-8286.2007.01931.x PMID: 21585727

52. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL (2008) NCBI BLAST: a better web interface. Nucleic Acids Research 36: W5–W9. doi: 10.1093/nar/gkn201 PMID: 18440982

53. Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Université de Montpellier II, Montpellier, France: Laboratoire Génome, Populations, Interactions, CNRS UMR 5171.

54. Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular Ecology Resources 14: 209–214. doi: 10.1111/1755-0998.12157 PMID: 23992227

55. Kalinowski ST, Wagner AP, Taper ML (2006) ML-RELATE: a computer program for maximum likelihood estimation of relatedness and relationship. Molecular Ecology Notes 6: 576–579.

56. Andrews KR, Luikart G (2014) Recent novel approaches for population genomics data analysis. Molecular Ecology 23: 1661–1667. doi: 10.1111/mec.12686 PMID: 24495199

57. Puritz JB, Matz MV, Toonen RJ, Weber JN, Bolnick DI, Bird CE (2014) Demystifying the RAD fad. Molecular Ecology: n/a-n/a.

58. Alho JS, Välimäki K, Merilä J (2010) Rhh: an R extension for estimating multilocus heterozygosity and heterozygosity–heterozygosity correlation. Molecular Ecology Resources 10: 720–722. doi: 10.1111/j.1755-0998.2010.02830.x PMID: 21565077

59. Team RC (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015.

60. Slate J, David P, Dodds KG, Veenvliet BA, Glass BC, Broad TE, et al. (2004) Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Heredity 93: 255–265. PMID: 15254488

61. Jombart T, Ahmed I (2011) adegenet 1.3–1: new tools for the analysis of genome-wide SNP data. Bioinformatics.

62. Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403–1405. doi: 10.1093/bioinformatics/btn129 PMID: 18397895

63. Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11: 94. doi: 10.1186/1471-2156-11-94 PMID: 20950446

64. Steinig EJ, Neuditschko M, Khatkar MS, Raadsma HW, Zenger KR (2016) NetView P: A network visualization tool to unravel complex population structure using genome-wide SNPs. Molecular Ecology Resources: 1–12.

65. Neuditschko M, Khatkar MS, Raadsma HW (2012) NETVIEW: A High-Definition Network-Visualization Approach to Detect Fine-Scale Population Structures from Genome-Wide Patterns of Variation. PLoS ONE 7: e48375. doi: 10.1371/journal.pone.0048375 PMID: 23152744

66. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81: 559–575. PMID: 17701901

67. Tsafrir D, Tsafrir I, Ein-Dor L, Zuk O, Notterman DA, Domany E (2005) Sorting points into neighborhoods (SPIN): data analysis and visualization by ordering distance matrices. Bioinformatics 21: 2301–2308. PMID: 15722375

68. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27: 431–432. doi: 10.1093/bioinformatics/btq675 PMID: 21149340

69. Peakall ROD, Smouse PE (2006) GENEALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288–295.

70. Foll M (2012) BayeScan v2.1 User Manual. Ecology 20: 1450–1462.

71. Foll M, Gaggiotti O (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics 180: 977–993. doi: 10.1534/genetics.108.092221 PMID: 18780740

72. Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a F_{st}-outlier method. BMC Bioinformatics 9: 323. doi: 10.1186/1471-2105-9-323 PMID: 18662398

73. White TA, Stamford J, Rus Hoelzel A (2010) Local selection and population structure in a deep-sea fish, the roundnose grenadier (*Coryphaenoides rupestris*). Molecular Ecology 19: 216–226. doi: 10.1111/j.1365-294X.2009.04446.x PMID: 20002604

74. Kovach RP, Gharrett AJ, Tallmon DA (2012) Genetic change for earlier migration timing in a pink salmon population. Proceedings of the Royal Society B: Biological Sciences 279: 3870–3878. doi: 10.1098/rspb.2012.1158 PMID: 22787027

75. Ball RD (2013) Designing a GWAS: Power, Sample Size, and Data Structure. In: Gondro C, van der Werf J, Hayes B, editors. Genome-Wide Association Studies and Genomic Prediction: Humana Press. pp. 37–97.

76. Hayes B (2013) Overview of Statistical Methods for Genome-Wide Association Studies (GWAS). In: Gondro C, van der Werf J, Hayes B, editors. Genome-Wide Association Studies and Genomic Prediction: Humana Press. pp. 149–170.

77. Gogarten SM, Bhangale T, Conomos MP, Laurie CA, McHugh CP, Painter I, et al. (2012) GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. Bioinformatics 28: 3329–3331. doi: 10.1093/bioinformatics/bts610 PMID: 23052040

78. Cummings JA (2005) Operational multivariate ocean data assimilation. Quarterly Journal of the Royal Meteorological Society 131: 3583–3604.

79. Chassignet EP, Hurlburt HE, Smedstad OM, Halliwell GR, Hogan PJ, Wallcraft AJ, et al. (2007) The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system. Journal of Marine Systems 65: 60–83.

80. Beer AC, Southgate PC (2000) Collection of pearl oyster (family Pteriidae) spat at Orpheus Island, Great Barrier Reef (Australia). Journal of Shellfish Research 19: 821–826.

81. Viikmäe B, Torsvik T, Soomere T (2013) Impact of horizontal eddy diffusivity on Lagrangian statistics for coastal pollution from a major marine fairway. Ocean Dynamics 63: 589–597.

82. Markey KL, Abdo DA, Evans SN, Bosserelle C (2016) Keeping It Local: Dispersal Limitations of Coral Larvae to the High Latitude Coral Reefs of the Houtman Abrolhos Islands. PLoS ONE 11.

83. Halliwell GR (2004) Evaluation of vertical coordinate and vertical mixing algorithms in the HYbrid-Coordinate Ocean Model (HYCOM). Ocean Modelling 7: 285–322.

84. Vilisoni MTJ (2012) Recruitment patterns of molluscs in Savusavu Bay, Fiji with emphasis on the Blacklip Pearl Oyster, Pinctada margaritifera (Linnaeus, 1758) [Unpublished thesis]. Suva, Fiji Islands: University of the South Pacific.

85. Saucedo PE, Southgate PC (2008) Reproduction, Development and Growth. In: Southgate PC, Lucas JS, editors. The Pearl Oyster. Amsterdam, the Netherlands: Elsevier. pp. 133–186.

86. Pouvreau S, Prasil V (2001) Growth of the black-lip pearl oyster, *Pinctada margaritifera*, at nine culture sites of French Polynesia: synthesis of several sampling designs conducted between 1994 and 1999. Aquatic Living Resources 14: 155–163.

87. Pouvreau S, Tiapari J, Gangnery A, Lagarde F, Garnier M, Teissier H, et al. (2000) Growth of the black-lip pearl oyster, *Pinctada margaritifera*, in suspended culture under hydrobiological conditions of Takapoto lagoon (French Polynesia). Aquaculture 184: 133–154.

88. Durand P, Wada KT, Blanc F (1993) Genetic variation in wild and hatchery stocks of the black pearl oyster, *Pinctada margaritifera*, from Japan. Aquaculture 110: 27–40.

89. Lemer S, Rochel E, Planes S (2011) Correction Method for Null Alleles in Species with Variable Microsatellite Flanking Regions, A Case Study of the Black-Lipped Pearl Oyster *Pinctada margaritifera*. Journal of Heredity 102: 243–246. doi: 10.1093/jhered/esq123 PMID: 21220742

90. Lind CE, Evans BS, Knauer J, Taylor JJU, Jerry DR (2009) Decreased genetic diversity and a reduced effective population size in cultured silver-lipped pearl oysters (*Pinctada maxima*). Aquaculture 286: 12–19.

91. Miller AD, Versace VL, Matthews TG, Montgomery S, Bowie KC (2013) Ocean currents influence the genetic structure of an intertidal mollusc in southeastern Australia–implications for predicting the movement of passive dispersers across a marine biogeographic barrier. Ecology and Evolution 3: 1248–1261. doi: 10.1002/ece3.535 PMID: 23762511

92. Peñaloza C, Bishop SC, Toro J, Houston RD (2014) RAD Sequencing reveals genome-wide heterozygote deficiency in pair crosses of the Chilean mussel Mytilus spp. 10th World Congress on Genetics Applied to Livestock Production. Vancouver, British Columbia, Canada. pp. 1–3.

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

94. Knutsen H, Olsen EM, Jorde PE, Espeland SH (2011) Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. Molecular ecology 20: 768. doi: 10.1111/j.1365-294X.2010.04979.x PMID: 21199035

95. Lind CE, Evans BS, Taylor JJU, Jerry DR (2010) The consequences of differential family survival rates and equalizing maternal contributions on the effective population size (Ne) of cultured silver-lipped pearl oysters, *Pinctada maxima*. Aquaculture Research 41: 1229–1242.

96. Pit JH, Southgate PC (2003) Fouling and predation; how do they affect growth and survival of the blacklip pearl oyster, *Pinctada margaritifera*, during nursery culture? Aquaculture International 11: 545–555.

97. Doroudi MS, Southgate PC (2002) The effect of chemical cues on settlement behaviour of blacklip pearl oyster (*Pinctada margaritifera*) larvae. Aquaculture 209: 117–124.

98. Friedman KJ, Bell JD (2000) Shorter immersion times increase yields of the blacklip pearl oyster, *Pinctada margaritifera* (Linne.), from spat collectors in Solomon Islands. Aquaculture 187: 299–313.

99. Friedman KJ, Southgate PC (1999) Growout of Blacklip Pearl Oysters, *Pinctada margaritifera* collected as wild spat in the Solomon Islands. Journal of Shellfish Research 18: 159–167.

100. Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. Trends in Ecology & Evolution 27: 489–496.

101. Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. Molecular Ecology 18: 375–402. doi: 10.1111/j.1365-294X.2008.03946.x PMID: 19143936

102. Jerry DR, Kvingedal R, Lind CE, Evans BS, Taylor JJU, Safari AE (2012) Donor-oyster derived heritability estimates and the effect of genotype × environment interaction on the production of pearl quality traits in the silver-lip pearl oyster, *Pinctada maxima*. Aquaculture 338– 341: 66–71.

103. Kvingedal R, Evans BS, Lind CE, Taylor JJU, Dupont-Nivet M, Jerry DR (2010) Population and family growth response to different rearing location, heritability estimates and genotype × environment interaction in the silver-lip pearl oyster (*Pinctada maxima*). Aquaculture 304: 1–6.

104. Kvingedal R, Evans BS, Taylor JJU, Knauer J, Jerry DR (2008) Family by environment interactions in shell size of 43-day old silver-lip pearl oyster (*Pinctada maxima*), five families reared under different nursery conditions. Aquaculture 279: 23–28.

105. Neo ML, Erftemeijer PLA, Beek KL, Maren DS, Teo SLM, Todd PA (2013) Recruitment constraints in Singapore's fluted giant clam (*Tridacna squamosa*) population—A dispersal model approach. PLoS One 8.

106. Wood S, Paris CB, Ridgwell A, Hendy EJ (2014) Modelling dispersal and connectivity of broadcast spawning corals at the global scale. Global Ecology and Biogeography 23: 1–11.

107. Thomas Y, Dumas F, Andréfouët S (2014) Larval Dispersal Modeling of Pearl Oyster *Pinctada margaritifera* following Realistic Environmental and Biological Forcing in Ahe Atoll Lagoon. PLoS ONE 9: e95050. doi: 10.1371/journal.pone.0095050 PMID: 24740288

108. White C, Selkoe KA, Watson J, Siegel DA, Zacherl DC, Toonen RJ (2010) Ocean currents help explain population genetic structure.

109. Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, et al. (2010) Adaptation genomics: the next generation. Trends in Ecology and Evolution 25: 705–712. doi: 10.1016/j.tree.2010.09.002 PMID: 20952088

110. Zarraonaindia I, Iriondo M, Albaina A, Pardo MA (2012) Multiple SNP markers reveal fine-scale population and deep phylogeographic structure in European anchovy (*Engraulis encrasicolus* L.). PloS one 7: e42201. doi: 10.1371/journal.pone.0042201 PMID: 22860082

Lal et al. BMC Genomics (2017) 18:66
DOI 10.1186/s12864-016-3410-y

BMC Genomics

RESEARCH ARTICLE Open Access

Swept away: ocean currents and seascape features influence genetic structure across the 18,000 Km Indo-Pacific distribution of a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*

Monal M. Lal[1*], Paul C. Southgate[1,2], Dean R. Jerry[1], Cyprien Bosserelle[3] and Kyall R. Zenger[1]

Abstract

Background: Genetic structure in many widely-distributed broadcast spawning marine invertebrates remains poorly understood, posing substantial challenges for their fishery management, conservation and aquaculture. Under the Core-Periphery Hypothesis (CPH), genetic diversity is expected to be highest at the centre of a species' distribution, progressively decreasing with increased differentiation towards outer range limits, as populations become increasingly isolated, fragmented and locally adapted. The unique life history characteristics of many marine invertebrates such as high dispersal rates, stochastic survival and variable recruitment are also likely to influence how populations are organised. To examine the microevolutionary forces influencing population structure, connectivity and adaptive variation in a highly-dispersive bivalve, populations of the black-lip pearl oyster *Pinctada margaritifera* were examined across its ~18,000 km Indo-Pacific distribution.

Results: Analyses utilising 9,624 genome-wide SNPs and 580 oysters, discovered differing patterns of significant and substantial broad-scale genetic structure between the Indian and Pacific Ocean basins. Indian Ocean populations were markedly divergent (F_{st} = 0.2534–0.4177, $p < 0.001$), compared to Pacific Ocean oysters, where basin-wide gene flow was much higher (F_{st} = 0.0007–0.1090, $p < 0.001$). Partitioning of genetic diversity (hierarchical AMOVA) attributed 18.1% of variance between ocean basins, whereas greater proportions were resolved within samples and populations (45.8% and 35.7% respectively). Visualisation of population structure at selectively neutral loci resolved three and five discrete genetic clusters for the Indian and Pacific Oceans respectively. Evaluation of genetic structure at adaptive loci for Pacific populations (89 SNPs under directional selection; F_{st} = 0.1012–0.4371, FDR = 0.05), revealed five clusters identical to those detected at neutral SNPs, suggesting environmental heterogeneity within the Pacific. Patterns of structure and connectivity were supported by Mantel tests of isolation by distance (IBD) and independent hydrodynamic particle dispersal simulations.

(Continued on next page)

* Correspondence: monal.lal@my.jcu.edu.au
[1]Centre for Sustainable Tropical Fisheries and Aquaculture, and College of Science and Engineering, James Cook University, Townsville QLD 4811, QLD, Australia
Full list of author information is available at the end of the article

(Continued from previous page)

Conclusions: It is evident that genetic structure and connectivity across the natural range of *P. margaritifera* is highly complex, and produced by the interaction of ocean currents, IBD and seascape features at a broad scale, together with habitat geomorphology and local adaptation at regional levels. Overall population organisation is far more elaborate than generalised CPH predictions, however valuable insights for regional fishery management, and a greater understanding of range-wide genetic structure in a highly-dispersive marine invertebrate have been gained.

Keywords: Population genomics, Aquaculture, Core-periphery hypothesis, SNP, Hydrodynamic dispersal, Species distribution

Background

Understanding the patterns and processes shaping population genetic structure across the extent of a species' distribution is an important prerequisite for biological conservation and management efforts, as well as studies of speciation [1]. For marine taxa, regional fishery management and aquaculture practices also rely on biologically meaningful population structure to delineate discrete stocks [2–4]. The ability to quantify genetic variation across the geographical limits of a species may shed light on why species might demonstrate stable range boundaries, and also permit assessment of the conservation value of central (*C*) versus marginal (*M*) populations [1, 5, 6]. Several studies (reviewed by Eckert et al. [5] and Sexton et al. [6]), have investigated the central-marginal (*C-M*) hypothesis, also known as the core-periphery hypothesis (CPH; [5, 7, 8]). While many comparisons between taxa have revealed a general decline in genetic diversity and increased differentiation towards range margins, others show no clear patterns [1].

It is expected that the interplay of microevolutionary forces, (namely natural selection, genetic drift and gene flow), will largely determine the magnitude and extent of population structure and connectivity, although the spatial distribution and demographic characteristics of the species could also exert strong influences [5, 6]. The CPH provides a model for interpreting how microevolutionary forces may shape genetic divergence patterns throughout a species' range. Under this model, a species which colonises a geographical gradient of environmental conditions, is over time expected to exhibit maximised abundance (highest survival, reproduction and growth rates) around a central point where conditions are optimal, while populations become smaller, more fragmented, increasingly divergent and influenced by selective forces towards the periphery [5, 7, 9]. However, exactly how the abundant centre distribution relates to the partitioning of genetic diversity, patterns of differentiation and adaptive differences across the *C-M* cline, remains a contentious topic [5, 9]. One explanation offered suggests that both effective population size (N_e) and gene flow (m) should be highest at the centre, and lowest at range margins. Consequently, central populations are expected to be less genetically differentiated and possess higher levels of genetic diversity, than those existing at range margins [5, 7]. Furthermore, due to environmental heterogeneity across a *C-M* cline, local adaptation may be observed between populations existing at the core and range peripheries.

While several studies have examined *C-M* genetic patterns in terrestrial taxa [5, 10], comparatively few investigations have involved marine species [8], and marine invertebrates in particular [11]. Marine systems present several challenges for range-wide studies, as >70% of invertebrates and many vertebrates are characterised by large population sizes, high fecundity, external fertilisation and larvae that typically remain in the plankton for several weeks, although this may vary anywhere from a few minutes to years [12–16]. Consequently, *C-M* patterns compared to terrestrial taxa may differ from expectations under the CPH, as the homogenising influence of gene flow may maintain high connectivity across the *C-M* cline [8]. Furthermore, divergence and local adaptation may not be as apparent if populations remain highly connected, and environmental gradients are shallow.

Among marine invertebrates, species which are either completely sessile as adults (e.g. barnacles, sponges and ascidians), or possess very limited mobility (e.g. sea urchins, bivalves, gastropods), present additional challenges for assessment of *C-M* trends [17, 18]. As larvae undergo pelagic dispersal and recruitment, differential selective pressures and survival rates pre- and post-settlement, and also between the plankton and benthos may strongly influence the genetic composition of populations [19, 20]. Furthermore, the spatial distribution of a population may be limited to isolated biodiversity hotspots (e.g. single bivalve beds), or an entire reef shelf [21, 22].

Given the complex nature of the biological and environmental influences at play, it is important to consider multiple sources of information for range-wide investigations in the marine environment, particularly when the species being examined is extensively distributed across heterogenous habitats. Considerations that have been highlighted in previous analyses of *C-M* patterns involving terrestrial taxa, include examination of the geographical direction of the periphery studied, latitudinal effects,

the effects of species-range geometry (e.g. shape and size), as well as sampling strategy [1, 5, 10]. While not all of these may apply to marine scenarios, for taxa that employ a broadcast spawning reproductive strategy, consideration of the extent of ocean current-mediated larval dispersal addresses many of these points [4, 23–26].

Incorporation of environmental data such as dispersal modelling into range-wide studies is capable of offering unprecedented insights into larval dispersal limits [4, 25, 27–29], and when considered together with both neutral and adaptive patterns of population structure, permit a holistic assessment of concordance with the CPH, or other models of range-wide structuring. The advantage of using independent datasets also includes the potential to reveal and/or corroborate previously undiscovered or poorly understood biogeographic barriers to dispersal, cryptic speciation and regional local adaptation [30–33].

The black-lip pearl oyster *Pinctada margaritifera* (Pteriidae), is a marine bivalve mollusc that has a broad Indo-Pacific distribution (Fig. 1), and is highly valued for cultured pearl and pearl shell production [34, 35]. Aquaculture of this species comprises a valuable industry and important source of coastal community livelihood across almost the entire extent of its distribution [34, 36]. While analyses to examine population structure and connectivity have previously been carried out, these have produced mixed findings, incorporated a range of different marker types (allozymes, mtDNA and microsatellites), and never examined the entirety of the species distribution [19, 37–43]. The current species description includes a total of six sub-species [35, 44, 45], that are described exclusively on the basis of variable morphological characters [46]. In the Pacific basin, Hawaiian populations are known as *P. margaritifera* var. *galstoffi* (Bartsch, 1931), Cook Islands and French Polynesian individuals as *P. m.* var. *cummingi* (Reeve, 1857), and all Central and Western Pacific specimens as *P. m.* var. *typica* (Linnaeus, 1758). Indian Ocean populations are represented by *P. m.* var. *persica* (Jameson, 1901; Persian Gulf), *P. m.* var. *erythraensis* (Jameson, 1901; Red Sea) and *P. m.* var. *zanzibarensis* (Jameson, 1901; East Africa, Madagascar and Seychelle Islands [44]).

Significant genetic heterogeneity has been reported for *P. margaritifera* at nuclear markers (allozymes, anDNA markers and microsatellite loci), at various sites in the Western and Central Pacific [37, 42, 47], while contrastingly mitochondrial markers did not [37]. More recent

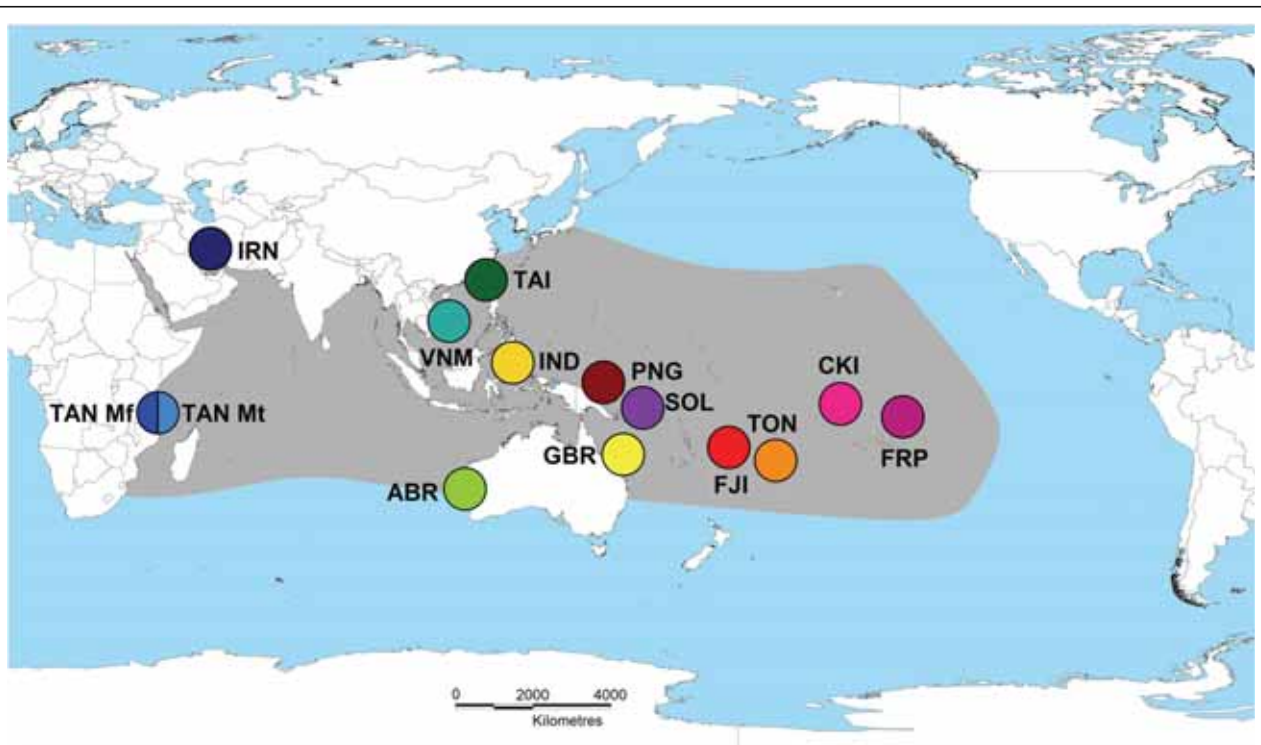


Fig. 1 Map of global sampling locations from where 580 individuals of *P. margaritifera* were collected. The approximate known distribution and range of the species is presented in grey, and adapted from Wada and Tëmkin [35]. Site codes represent the following locations: TAN Mf: Mafia Island, Tanzania (dark blue); TAN Mt: Mtwara, Tanzania (light blue); IRN: Hendorabi Island, Iran; TAI: Checheng, Taiwan; VNM: Nha Trang, Vietnam; IND: Manado, Indonesia; AU Abr: Abrolhos Islands, Australia; AU GBR: Great Barrier Reef, Australia; PNG: Kavieng, Papua New Guinea; SOL: Gizo Island, Solomon Islands; FJI: Kadavu, Savusavu, Lau and the Yasawa group, Fiji Islands; TON: Tongatapu, Tonga; CKI: Manihiki Atoll, Cook Islands and FRP: Arutua, French Polynesia

work also using microsatellite loci, discovered significant genetic structure both within and between French Polynesian island archipelagos, attributed to "open" and "closed" atoll lagoon hydromorphologies restricting patterns of gene flow [39]. Since then, genome-wide SNPs have been developed and characterised [48], and used to investigate stock structure for fishery management and aquaculture in the Fiji Islands [4], where a single genetic stock was identified.

Previous studies of range-wide genetic structuring in Pteriid pearl oysters have produced mixed results. Lind et al. [49] reported a reduction in genetic diversity towards the range periphery of the silver-lip pearl oyster, *P. maxima*, which is consistent with CPH assumptions. However, the natural distribution of this species is considerably less extensive than that of *P. margaritifera* [34, 35]. A bivalve which has a range similar to that of *P. margaritifera* is the Akoya pearl oyster, currently recognised as the *P. fucata*/*martensii*/*radiata*/*imbricata* species complex [35, 50]. While the population genetic structure of this taxon is pending resolution, it is thought that it may comprise one cosmopolitan, circum-globally distributed species, possessing a very high degree of intraspecific variation across its range [34, 35, 51].

Larval development of *P. margaritifera* occurs over 26–30 days in captivity [52, 53], however, time to settlement may be prolonged if conditions are unfavourable [54]. The high dispersal potential (and thus gene flow) in this species suggests that CPH trends may not be easily identifiable across the broader species range, except perhaps in situations where larval dispersal is restricted by seascape features (e.g. closed atoll lagoons or current gyres), or at the very limits of the species distribution where favourable habitat is limited, impacting fitness and population growth. Here, we assess populations of *P. margaritifera* across the extent of its Indo-Pacific distribution spanning over 18,000 km, and compare our observations with expectations under the CPH and regional morphological subdivisions. Independent population genomic and hydrodynamic approaches were utilised to assess population genetic structure, adaptive variation and larval connectivity. Through the use of independent biological and environmental datasets, this work sheds light on the links between genetic structure, ecology and oceanography, to reveal how populations of a broadcast spawner can be organised and maintained in the marine environment.

Methods

Specimen collection, tissue sampling and DNA extraction

Adult and juvenile *P. margaritifera* ($n = 580$) between 5 and 18 cm in dorso-ventral measurement (DVM) were collected from 14 sites across the species distribution (Fig. 1). All oysters were handled in accordance with James Cook University's animal ethics requirements and guidelines, with permission to collect tissues obtained from local authorities. In the Indian Ocean, oysters were collected from two sites in Tanzania (Mafia Island and Mtwara, $n = 35$ and $n = 20$ respectively), the Persian Gulf (Hendorabi Island, Iran; $n = 49$) and Post Office Island in the Abrolhos Islands group, Western Australia ($n = 50$). All Indian Ocean samples consisted of wild individuals with the exception of the Abrolhos Islands collection, where oysters were hatchery-produced from wild-caught broodstock. In the Western Pacific, oysters were sampled from Checheng, Taiwan ($n = 24$), Nha Trang, Vietnam ($n = 47$) and Manado, Indonesia ($n = 48$). Central Pacific locations were represented by Kavieng, Papua New Guinea ($n = 38$), Gizo Island in the Solomon Islands ($n = 50$), the Great Barrier Reef, Australia ($n = 35$), the Fiji Islands ($n = 61$) and Tonga ($n = 28$). In the Eastern Pacific, oysters were collected from Manihiki Atoll in the Cook Islands ($n = 45$), and Arutua, French Polynesia ($n = 50$). All Pacific Ocean samples consisted of wild oysters, with the exception of the Cook Islands and French Polynesian samples that were sourced from pearl farm stocks.

Proximal mantle and adductor muscle tissues (3 and 6 cm, respectively) were removed and transferred to tubes containing 20% salt saturated dimethyl sulfoxide (DMSO-salt) preservative [55]. Genomic DNA was extracted using a modified cetyl trimethyl ammonium bromide (CTAB, Amresco, cat. #0833-500G) chloroform/isoamyl alcohol protocol with a warm (30 °C) isopropanol precipitation [56]. To clean up all DNA extractions, a Sephadex G50 [57] spin column protocol was used prior to quantification with a Nanodrop 1000 Spectrophotometer (Thermo Scientific). All samples were subsequently normalised at 100 ng/μL in a 50 μL final volume, and submitted for DArTseq™ 1.0 genotyping at Diversity Arrays Technology PL, Canberra, ACT, Australia.

DArTseq™ 1.0 library preparation and sequencing

Diversity Arrays Technology (DArT PL) proprietary genotyping by sequencing (DArTseq™) reduced-representation libraries were prepared as described by Kilian et al. [58] and Sansaloni et al. [59], with a number of modifications for *P. margaritifera*. Briefly, genome complexity reduction was achieved with a double restriction digest, using a *Pst*I and *Sph*I methylation-sensitive restriction enzyme (RE) combination, in a joint digestion-ligation reaction at 37 °C for 2 h with 150–200 ng gDNA. Because *P. margaritifera* like other bivalve species is highly polymorphic [48, 60], highly repetitive genomic regions were avoided and low copy regions more efficiently targeted for sequence capture with the use of methylation-sensitive REs [61].

Custom proprietary barcoded adapters (6–9 bp) were ligated to RE cut-site overhangs as per Kilian et al. [58],

with the adapters designed to modify RE cut sites following ligation, to prevent insert fragment re-digestion. The *Pst*I-compatible (forward) adapter incorporated an Illumina flowcell attachment region, sequencing primer sequence and a varying length barcode region [58, 62]. The reverse adapter also contained a flowcell attachment region, and was compatible with the *Sph*I cut-site overhang. Samples were processed in batches of 94, with 15% of all samples in a batch randomly selected for replication, to provide a basis for assessing region recovery and genotyping reproducibility. Target "mixed" fragments [62], containing both *Sph*I and *Nla*III cut-sites were selectively amplified using custom designed primers for each sample, under the following PCR conditions: initial denaturation at 94 °C for 1 min, then 30 cycles of 94 °C for 20s, 58 °C for 30s and 72 °C for 45 s, followed by a final extension step at 72 °C for 7 min. Amplified samples were subsequently cleaned using a GenElute PCR Clean-up Kit (Sigma-Aldrich, cat.# NA1020-1KT), on a TECAN Freedom EVO150 automated liquid handler.

To examine fragment size concordance and digestion efficiency, all samples were visualised on a 0.8% agarose gel stained with EtBr, and quantified using the ImageJ software package [63]. Samples which did not appear to have undergone complete digestion and/or amplification were removed from downstream library preparation. A total of 580 samples were each normalised and pooled using an automated liquid handler (TECAN, Freedom EVO150), at equimolar ratios for sequencing on the Illumina HiSeq 2500 platform. After cluster generation and amplification (HiSeq SR Cluster Kit V4 cBOT, cat.# GD-401-4001), 77 bp single-end sequencing was performed at the DArT PL facility in Canberra, Australia.

Sequence quality control, marker filtering and genotype calling at DArT PL

Raw reads obtained following sequencing were processed using Illumina CASAVA v.1.8.2 software for initial assessment of read quality, sequence representation and generation of FASTQ files. Filtered FASTQ files were then supplied to the DArT PL proprietary software pipeline DArTtoolbox, which performed further filtering, variant calling and generated final genotypes in sequential primary and secondary workflows [64]. Within DArTtoolbox, the primary workflow first involved the package DArTsoft14 to remove reads with a quality score <25 from further processing, and apply stringent filtering to the barcode region of all sequences to increase confidence in genomic region recovery. Individual samples were then de-multiplexed by barcode, and subsequently aligned and matched to catalogued sequences in both NCBI GenBank and DArTdb custom databases to check for viral and bacterial contamination, with any matches removed from further processing.

The secondary workflow employed the DArTsoft14 and KD Compute packages along with the DArTdb database, to identify polymorphisms by aligning identical reads to create clusters across all individuals sequenced. These clusters were then catalogued in DArTdb, and matched against each other to create reduced-representation loci (RRL), based on their degree of similarity and size. SNP and reference allele loci were identified within clusters and assigned the following DArT scores: "0" = reference allele homozygote, "1" = SNP allele homozygote and "2" = heterozygote, based on their frequency of occurrence. To ensure robust variant calling, all monomorphic clusters were removed, SNP loci had to be present in both allelic states (homozygous and heterozygous), and a genetic similarity matrix was produced using the first 10,000 SNPs called to assess technical replication error [65], and exclude clusters containing tri-allelic or aberrant SNPs and overrepresented sequences.

Once SNP markers had been confidently identified, each locus was assessed in the KD Compute package for homozygote and heterozygote call rate, frequency, polymorphic information content (PIC), average SNP count, read depth and repeatability, before final genotype scores were supplied by DArT PL. Following the receipt of genotype data from DArT PL, the dataset was further filtered to retain only a single, highly informative SNP at each genomic locus. This was achieved by filtering out duplicate SNPs (possessing identical Clone IDs), according to call rate and Minor Allele Frequency (MAF). Subsequently, loci were screened for call rate, average Polymorphic Information Content (PIC), MAF and average repeatability, to retain SNPs suitable for population genomic analyses. All loci were then tested for departure from Hardy-Weinberg Equilibrium (HWE) using Arlequin v.3.5.1.3 [66], using an exact test with 10,000 steps in the Markov Chain and 100,000 dememorisations. Additionally, all loci were tested for genotypic linkage disequilibrium (LD) in Genepop v.4.3 [67], as per Lal et al. [48]. Two separate datasets were then created, one which contained selectively neutral loci, and the other which included loci putatively under selection. Bayescan v.2.1 and LOSITAN software were used to detect loci under selection, and further details are provided under that section of the methods.

Evaluation of genomic diversity, inbreeding and population differentiation

For assessment of genomic diversity within and between populations, allelic diversity indices including average observed (H_o) and average expected heterozygosities corrected for population sample size ($H_{n.b.}$) were computed. Inbreeding coefficient (F_{is}) calculations and estimation of effective population size based on the linkage disequilibrium method (N_{eLD}), were also carried out for

each population, all using Genetix v.4.05.2 [68] and NeEstimator v.2.01 [69]. Average homozygosity by locus (HL), standardised heterozygosity (SH) and internal relatedness (IR) were also computed per individual, with the *R* package *Rhh* [70]. In addition, the average multi-locus heterozygosity (Av. MLH) per population was determined after Slate et al. [71], along with the mean number of alleles per locus (*A*) using the *diveRsity* [72] *R* package. The number of private alleles (A_p) was computed using HP-RARE v.1.0 [73], according to population groups identified from Netview P and DAPC analyses (see results), due to the levels of genetic divergence observed. Furthermore, rare allelic richness (*Ar*, <5% MAF) was computed manually for each population.

Resolution of broad and fine-scale population structure and connectivity

Pairwise F_{st} estimates for each population were calculated using Arlequin v.3.5.1.3 with 10,000 permutations [66], along with a hierarchical Analysis of Molecular Variance (AMOVA) in the *R* package *Poppr* [74]. The AMOVA examined variation between individuals, populations and regions (Pacific vs. Indian Ocean basins). To assess an isolation by distance (IBD) model of gene flow among populations, Mantel tests were carried out using GenAlEx v.6.5 [75], based on pairwise F_{st} and straight-line geographic distance matrices over 10,000 permutations. Mantel tests were performed considering populations within each ocean basin together, separately, and also within Pacific Ocean population clusters identified by DAPC and NetView P analyses. Nei's (1978) standard genetic distances (D_S) between populations were also computed in Genetix v.4.05.2 with 10,000 permutations [68], and broad-scale population structure visualised by performing a Discriminant Analysis of Principal Components (DAPC) in the *R* package *adegenet* 1.4.2 [76–78]. The DAPC was carried out for all loci, and α-score optimisation used to determine the number of principal components to retain. To reveal any fine-scale stratification between and among all populations, network analysis was carried out using the NetView P pipeline v.0.4.2.5 [79, 80]. To further investigate the direction and magnitude of migration between populations, migration networks were generated using the *divMigrate* function of the *R* package *diveRsity*, utilising the Nei's G_{st} method [72, 81].

Examination of adaptive variation

To first create a selectively neutral dataset for population genomic analyses, a filtered dataset containing 10,683 SNP loci was used as the starting point for this step. Both BayeScan v.2.1 [82, 83] and LOSITAN selection detection workbench [84] software packages were employed to identify candidate loci under selection, at FDRs = 0.001, 0.005, 0.01, 0.05 and 0.1 and 0.2. The numbers of loci detected are summarised in Additional file 3, and verification of these loci was carried out using QQ plots (data not shown). The intended approach was to select loci jointly identified by both Bayescan 2.1 and LOSITAN, at the appropriate FDR threshold determined by QQ plot distribution. As these software packages employ different analytical approaches, their joint use generally increases the statistical confidence of F_{st} outlier detection [85–87]. Candidate loci identified with high probability using both methods were to be considered as true outliers, and representative of putative selection impacting the populations examined. However, given the tendency of LOSITAN to overestimate the numbers of loci under selection [32, 48, 88], and disagreement on an appropriate FDR threshold to apply using both methods, a conservative approach was taken where LOSITAN results were disregarded, and the Bayescan 2.1 results at an FDR = 0.01 considered. This indicated that a total of 1,059 putatively balancing and directional loci were present in the dataset, and following their removal, a selectively neutral dataset containing 9,624 SNPs remained.

Further population-specific F_{st} outlier tests were used to detect local adaptation, with population pairs tested at FDRs of 0.001, 0.005, 0.01, 0.05 and 0.1 and 0.2. However, testing for F_{st} outliers was restricted to populations sampled from the Pacific Ocean basin, as they were the least differentiated amongst themselves (i.e. lowest neutral F_{st} levels <0.11; see results), while all Indian Ocean populations were significantly more divergent. Comprehensive descriptions of the settings used for both software packages were as per Lal et al. [4, 48]. Results of the Bayescan 2.1 and LOSITAN analyses, together with the construction of pairs of Quantile-Quantile plots (QQ-plots), were used to assess the suitability of an FDR threshold for outlier detection between the two methods. The *R* package *GWASTools* v.1.14.0 [89] was used to construct all QQ-plots at all FDR levels examined. All loci were included in the first QQ plot constructed to visualise deviation outside the bounds of a 95% confidence interval. If deviation was observed, a second plot was generated excluding all outlier loci. If all remaining loci were normally distributed, this was interpreted as confirmation that outlier loci had been identified with high probability.

Particle dispersal simulation

To independently evaluate larval connectivity using oceanographic data for comparison with population genomic analyses, larval transport pathways between sampling locations were simulated using the particle dispersal modelling software DisperGPU (https://github.com/CyprienBosserelle/DisperGPU). Larvae of *P. margaritifera* remain in the plankton for 26–30 days prior to settlement [52, 53], and due to very

limited motility, are largely dispersed by current advection and turbulent diffusion in the ocean surface (mixed) layer.

Hydrodynamic and dispersal numerical models

The particle dispersal model was driven by current velocity output from the global HYbrid Coordinate Ocean Model (HYCOM) data [90, 91]. HYCOM is a global hydrodynamic model that simulates ocean surface heights, currents, salinity and temperature, both at the surface and at depth. The model is driven by meteorological forcing, and constantly constrained by the assimilation of global, remote and in-situ ocean observations. As the model simulates regional and global circulation, it does not include tidal or surface wind waves. HYCOM is highly useful for forecasting and simulation experiments, with public availability at https://hycom.org. The HYCOM model had a resolution of 1/12th of a degree and output every day. The particle model used a standard Lagrangian formulation [92, 93], where particles have no physical representation, but rather track the displacement of neutrally buoyant small objects such as larvae (relative to the model resolution), at the ocean surface. Particle displacement is expressed as:

$$\Delta x = u_p * \Delta t + K \quad (1)$$

Here x represents particle position (latitude and longitude), Δx is particle displacement during a time step Δt (which was set at 1 h), and u_p is the surface current speed at the location of the particle. K is the eddy diffusivity which takes account of the random displacement of the particle, due to turbulent eddies at a scale smaller than the hydrodynamics model resolution. K is calculated after Viikmäe et al. [94] as follows:

$$K = \sqrt{-4E_h \Delta t \log(1-R_{NA})}\cos(2\pi R_{NB}) \quad (2)$$

Here E_h is a horizontal turbulent diffusion coefficient, and R_{NA} with R_{NB} are normally distributed random numbers. The horizontal turbulent diffusion coefficient is unknown, but assumed to be 5 m^2s^{-1} [94] and u_p (in Eq. 1) is calculated by interpolating the velocity from the hydrodynamic model, both spatially and temporally. Gridded surface currents are first interpolated to the dispersal step, after which the current velocity at each particle position is calculated using a bi-linear interpolation of the gridded surface currents, where only surface currents are taken into account and vertical movements neglected [95]. The particle age is retained and increases with simulation progression.

Model configuration

Particles were seeded in 11 locations corresponding to locations from where oysters were sampled for genetic analyses (see Fig. 5), which were represented at scales larger than the precise sampling locations to factor in the extent of surrounding coral reef habitat, as per Lal et al. [4]. All seed areas were also extended farther offshore to account for the fact that the HYCOM model is not adapted for shallow water environments, and does not resolve fine-scale hydrodynamic patterns <10 km [96]. Dispersal simulations for the Tanzanian and Iranian sites were not explored, due to the considerable distances between locations, and preliminary examination of circulation patterns that predicted a lack of particle admixture.

Within the Pacific basin, *P. margaritifera* is known to have two reproductive events per year, with peaks and duration of spawning events varying by location. In the Indian Ocean, spawning appears to be restricted to a single season [97]. A summary of the number and duration of spawning seasons for each sampling location was compiled from literature, to replicate larval supply over the year (see Additional file 1). At each seed location, 25,600 particles (see Lal et al., [4]) were released per day for 14 days, corresponding to documented spawning peaks for the species, and the model run forward in time for 90 and 60 days for the first and second spawning periods respectively, within a single calendar year. Simulations were run separately for each of the two spawning periods using HYCOM data for 2015 and 2014, which were selected as these corresponded to an El Niño Southern Oscillation event (ENSO), [98, 99]. This permitted evaluation of any changes in dispersal patterns due to ENSO events over the 2014–2015 time scale.

Particle positions were extracted at time intervals of 60 and 90 days post-seeding for the first and second spawning seasons respectively, per year, and particle displacement visualised using the Generic Mapping Tools package [100]. Explicit, quantitative correlation of the genetic and hydrodynamic analyses was not possible, as this would have required genetic analysis of oysters at all potential source and sink locations with dense sampling coverage, and modelling of substantially more complex particle competency behaviour than computational resources permitted. Instead, an independent approach was adopted here, to examine congruency of results produced by the two analyses. No mortality or competency behaviour of the particles was simulated.

Results

SNP filtering

The raw dataset contained a total of 19,666 SNPs genotyped across all 580 individuals, at call rates ranging from 20 to 100%. The first filtering step undertaken to remove duplicate (clone) SNPs at genomic loci resulted in the removal of 8,079 SNPs (41% loss), after which the dataset was filtered for call rate (65%), average PIC (1%),

MAF (2%) and average repeatability (95%). A total of 7 loci were found to deviate from HWE ($p < 0.009$), and 99 loci were monomorphic across all 14 populations, which were subsequently removed together with 107 loci under significant LD ($p < 0.0001$). These steps collectively resulted in the retention of 10,683 SNPs (Additional file 6). Testing of this filtered dataset for F_{st} outlier loci detected 1,059 SNPs determined to be putatively under balancing and directional selection (Bayescan 2.1 results at FDR = 0.01; Additional file 3), and their removal generated a final neutral dataset of 9,624 SNPs (Additional file 5). This dataset was used for performing all population genomic analyses, while the original filtered dataset (10,683 SNPs) was retained for investigating adaptive variation.

Population genomic diversity and differentiation

Patterns observed in the mean numbers of alleles per locus (*A*) and rare allelic richness (*Ar*, <5% MAF) were similar, and appeared to vary by Ocean basin (Table 1). Values of *A* for Pacific Ocean populations ranged from 1.6256 (Cook Islands) to 1.8067 (Indonesia), whereas Indian Ocean populations produced values of 1.3934–1.5649 (Tanzania, Mtwara to Abrolhos Islands, Australia). Trends in the total numbers of private alleles (A_p) reflected the divergence between ocean basins and support very limited inter-basin gene flow, with more than 25% of total SNPs genotyped containing private alleles within each basin; (2,672 and 2,508 for Indian and Pacific Oceans respectively). Within ocean basins, little difference (~2% of total SNPs) was seen among Pacific populations (A_p range of 188–205), while greater differences (~3–5% total SNPs) were observed among the Abrolhos Islands, both Tanzanian, and Iranian sites (290, 354 and 458 respectively).

Average observed heterozygosities were significantly lower ($p < 0.05$) than average expected heterozygosities for all populations), and displayed similar variability with the trends observed for *A* and *Ar* values. Pacific Ocean populations displayed generally higher values (H_o: 0.0718–0.0929; $H_{n.b.}$: 0.1722–0.2060), than did Indian Ocean populations (H_o: 0.0371–0.0748; $H_{n.b.}$: 0.1187–0.1655). These patterns also extended to individual average multi-locus heterozygosity (MLH) computations, and measurements of standardised heterozygosity (SH). Average MLH was relatively uniform within Pacific Ocean populations, ranging from 0.0844 (French Polynesia) to 0.1030 (Fiji Islands), which was mirrored in the SH results of 0.9777–1.2189 for the same populations respectively. Within Indian Ocean samples, oysters collected from Tanzanian and Iranian sites showed lower values (MLH: 0.0520–0.0557; SH: 0.5830–0.6206), than animals sampled from the Abrolhos Islands (MLH = 0.0914; SH = 1.0682).

Inbreeding coefficient (F_{is}) values displayed a similar partitioning by region, with values for Pacific Ocean populations ranging from 0.5372 (Fiji Islands) to 0.6433 (Taiwan), while Indian Ocean animals (with the exception of Abrolhos Islands oysters; $F_{is} = 0.5542$), returned higher values from 0.6795 (Tanzania, Mtwara) to 0.7008 (Iran). Very similar patterns were evident in related homozygosity by locus (HL) and internal relatedness (IR) multi-locus metrics (see Table 1). Estimates of effective population size were robust, however, they varied considerably across all sampling locations. Several populations returned infinite N_{eLD} values, including oysters sampled from the GBR, Taiwan and the two Tanzanian locations. Estimates from Solomon Islands samples were at the low end of the range (119.8; [95% CI = 118.9–120.8]), while Cook Islands individuals produced higher values (1,684.7; [95% CI = 1,475.1–1,963.3]). The lowest estimates were obtained from Abrolhos Islands oysters (9.3; [95% CI = 9.3–9.4]), indicating a possible bottleneck, as these animals were F_1 hatchery-produced offspring of wild-caught parents.

Resolution of population structure and migration

Pairwise F_{st} estimates (Table 2) were highly significant ($p < 0.001$) for all population comparisons, with the exception of the two Tanzanian sites (0.0007), and PNG with the Solomon Islands (0.0059). A clear separation in population structure between ocean basins is evident, with pairwise estimates between sites all >0.25, ranging from Tanzania, Mtwara and Indonesia (0.2894), to Iran and the Cook Islands (0.4684). Within the Pacific, populations appear to be isolated by geographic separation, e.g. pairwise estimates for the GBR and Solomon Islands (0.0078) indicate greater homogeneity than more distant population pairs, such as the Cook Islands and Taiwan (0.1090). Higher degrees of separation are apparent within Indian Ocean populations, with pairwise estimates between Iran, and Mafia Islands with Mtwara being 0.2444 and 0.2534 respectively. The greatest level of differentiation among Indian Ocean sites was detected between the Abrolhos Islands and Iran (0.4177), with oysters from the Abrolhos Islands demonstrating greater similarity with Pacific populations (Abrolhos Islands and GBR pairwise $F_{st} = 0.1311$).

Pairwise Nei's standard genetic distances (D_S) described a similar pattern to the pairwise F_{st} estimates (Table 2), with the Iranian and two Tanzanian populations displaying marked separation from all other populations (0.214–0.306; $p < 0.05$). Partitioning between these populations however, was less evident, with D_S = 0.071 and 0.074 respectively (Iran with Mafia Islands and Mtwara). Distances between all Pacific Ocean populations conversely indicated greater homogeneity, ranging from 0.005 (PNG, GBR and Solomon Islands pairwise comparisons), to 0.044 (Cook Islands with

Table 1 Genetic diversity indices for the *P. margaritifera* populations sampled

Population	n	Source	N_{eLD} [95% C.I.]	*A*	*Ar* (<5%)	H_o (± SD)	$H_{n.b.}$ (± SD)	F_{is} (p < 0.01)	Av. MLH (± SD)	HL (± SD)	SH (± SD)	IR (± SD)
Australia Abrolhos Island	50	Farm (hatchery)	9.3 [9.3 – 9.4]	1.5649	0.5446	0.0748 (±0.1244)	0.1655 (±0.1924)	0.5542	0.0914 (±0.0115)	0.8592 (±0.0174)	1.0682 (±0.1457)	0.5765 (±0.0516)
Australia GBR	35	Wild	∞ [∞ – ∞]	1.7603	0.3822	0.0762 (±0.0995)	0.2005 (±0.1771)	0.6265	0.0877 (±0.0044)	0.8618 (±0.0073)	1.0189 (±0.0567)	0.5737 (±0.0222)
Cook Islands	45	Farm (wild origin)	1684.7 [1475.1 – 1963.3]	1.6256	0.4984	0.0728 (±0.1092)	0.1722 (±0.1854)	0.5830	0.0868 (±0.0114)	0.8655 (±0.0179)	1.0066 (±0.1398)	0.5888 (±0.0523)
Fiji Islands	61	Farm (wild origin)	232.4 [229.9 – 234.9]	1.7934	0.3895	0.0929 (±0.1151)	0.1991 (±0.1758)	0.5372	0.1030 (±0.0306)	0.8370 (±0.0475)	1.2189 (±0.3905)	0.5050 (±0.1366)
French Polynesia	50	Farm	299.5 [293.4 – 305.9]	1.7208	0.4334	0.0718 (±0.1002)	0.1883 (±0.1814)	0.6236	0.0844 (±0.0091)	0.8687 (±0.0145)	0.9777 (±0.1132)	0.5965 (±0.0416)
Indonesia	48	Wild	1036.3 [972.6 – 1108.9]	1.8067	0.3635	0.0806 (±0.1027)	0.2054 (±0.1739)	0.6121	0.0925 (±0.0137)	0.8543 (±0.0215)	1.0816 (±0.1730)	0.5568 (±0.0633)
Iran	49	Wild	767.8 [693.1 – 860.3]	1.4402	0.7757	0.0371 (±0.0858)	0.1187 (±0.1795)	0.7008	0.0520 (±0.0039)	0.9378 (±0.0056)	0.5830 (±0.0445)	0.8127 (±0.0145)
Papua New Guinea	38	Wild	199.9 [196.3 – 203.8]	1.7632	0.3774	0.0732 (±0.0967)	0.2007 (±0.1769)	0.6410	0.0847 (±0.0034)	0.8661 (±0.0061)	0.9800 (±0.0399)	0.5884 (±0.0152)
Solomon Islands	50	Wild	119.8 [118.9 – 120.8]	1.8001	0.3748	0.0859 (±0.1077)	0.2019 (±0.1739)	0.5790	0.0965 (±0.0211)	0.8471 (±0.0336)	1.1374 (±0.2709)	0.5323 (±0.0964)
Taiwan	24	Wild	∞ [∞ – ∞]	1.7098	0.3830	0.0741 (±0.1035)	0.2021 (±0.1830)	0.6433	0.0859 (±0.0050)	0.8643 (±0.0080)	0.9947 (±0.0648)	0.5864 (±0.0230)
Tanzania (Mafia Island)	35	Wild	∞ [∞ – ∞]	1.4462	0.6369	0.0410 (±0.0878)	0.1296 (±0.1840)	0.6964	0.0553 (±0.0060)	0.9290 (±0.0083)	0.6149 (±0.0715)	0.7910 (±0.0215)
Tanzania (Mtwara)	20	Wild	∞ [∞ – ∞]	1.3934	0.6485	0.0427 (±0.0951)	0.1256 (±0.1871)	0.6795	0.0557 (±0.0083)	0.9285 (±0.0108)	0.6206 (±0.0994)	0.7898 (±0.0299)
Tonga	28	Wild	120.8 [118.7 – 122.8]	1.6995	0.4062	0.0775 (±0.1076)	0.1954 (±0.1828)	0.6119	0.0889 (±0.0104)	0.8594 (±0.0171)	1.0404 (±0.1326)	0.5714 (±0.0493)
Vietnam	47	Wild	681.5 [651.7 – 714.2]	1.8016	0.3587	0.0775 (±0.0994)	0.2060 (±0.1737)	0.6281	0.0892 (±0.0135)	0.8592 (±0.0215)	1.0378 (±0.1706)	0.5723 (±0.0616)

The parameters calculated include the effective population size by the linkage disequilibrium method (N_{eLD}; 95% confidence intervals indicated within brackets), mean number of alleles per locus (*A*), rare allelic richness at (*Ar*, MAF <5%), observed heterozygosity (H_o), average expected heterozygosity corrected for population sample size ($H_{n.b.}$), inbreeding coefficient (F_{is}), average individual multi-locus heterozygosity (Av. MLH), homozygosity by locus (HL), standardised heterozygosity (SH) and internal relatedness (IR)

Table 2 Population differentiation estimates for 14 *P. margaritifera* populations sampled

	Australia Abrolhos Islands	Australia GBR	Cook Islands	Fiji Islands	French Polynesia	Indonesia	Iran	Papua New Guinea	Solomon Islands	Taiwan	Tanzania (Mafia Island)	Tanzania (Mtwara)	Tonga	Vietnam
Australia Abrolhos Islands		0.056	0.095	0.068	0.082	0.053	0.264	0.056	0.057	0.053	0.236	0.238	0.069	0.051
Australia GBR	0.1311		0.033	0.009	0.021	0.009	0.256	**0.005**	0.005	0.009	0.234	0.236	0.011	0.008
Cook Islands	0.2173	0.0816		0.023	0.020	0.044	0.306	0.035	0.035	0.044	0.289	0.291	0.027	0.043
Fiji Islands	0.1526	0.0194	0.0537		0.011	0.018	0.273	0.011	0.010	0.017	0.253	0.255	0.006	0.016
French Polynesia	0.1892	0.0490	0.0502	0.0221		0.031	0.292	0.022	0.021	0.030	0.275	0.277	0.014	0.028
Indonesia	0.1209	0.0211	0.1084	0.0459	0.0759		0.243	0.008	0.009	**0.006**	0.217	0.219	0.020	0.006
Iran	0.4177	0.4145	0.4684	0.3903	0.4438	0.3711		0.255	0.257	0.241	0.071	0.074	0.276	0.239
Papua New Guinea	0.1297	0.0079	0.0862	0.0227	0.0532	0.0166	0.40860		0.005	**0.008**	0.232	0.233	0.013	0.007
Solomon Islands	0.1297	0.0071	0.0835	0.0248	0.0499	0.0208	0.38462	**0.0056**		0.008	0.234	0.237	0.012	0.008
Taiwan	0.1196	0.0172	0.1090	0.0375	0.0739	0.0100	0.41411	0.0148	0.0128		0.217	0.218	0.020	**0.006**
Tanzania (Mafia Island)	0.3508	0.3494	0.4185	0.3374	0.3951	0.3038	0.24438	0.3444	0.3236	0.3394		0.013	0.257	0.214
Tanzania (Mtwara)	0.3402	0.3323	0.4043	0.3235	0.3804	0.2894	0.25340	0.3280	0.3084	0.3200	**0.0069**		0.259	0.216
Tonga	0.1607	0.0235	0.0628	0.0099	0.0294	0.0494	0.44413	0.0279	0.0254	0.0443	0.3846	0.3644		0.019
Vietnam	0.1128	0.0185	0.1062	0.0407	0.0732	0.0124	0.37119	0.0161	0.0174	0.0088	0.3043	0.2906	0.0469	

Population pairwise F_{st} estimates computed in Arlequin v.3.5.1.3. are shown below the diagonal, while Nei's (1978) standard genetic distances (D_S) computed in Genetix v.4.05.2 with 10,000 permutations are reported above. All F_{st} values were significant at $p < 0.001$ following 1,000 permutations. Non-significant F_{st} and D_S values ($p > 0.05$) are presented in bold type

Indonesia and Taiwan pairwise comparisons). Oysters collected from the Abrolhos Islands were similarly differentiated, with $D_S = 0.056$ when compared to GBR individuals, and up to $D_S = 0.082$ with French Polynesian animals.

Results of the hierarchical AMOVA carried out between Indian vs. Pacific Ocean basins and populations indicated that 18.11% of the variance originated between ocean basins, with the greatest proportions of variance attributed to within-sample variation (45.79%), and between samples within populations (35.74%). Variation between populations within ocean basins was estimated at just 0.36%, indicating that genotypic variability at the individual oyster level accounted for the majority of the observed variation. Mantel tests indicated isolation by distance dispersal patterns both within each ocean basin ($R^2 = 0.939$, $p = 0.041$ and $R^2 = 0.464$, $p = 0.000$ for Indian and Pacific oceans respectively), as well as for all populations considered together ($R^2 = 0.613$, $p = 0.000$), although additional sampling within each region is needed to confirm the strength of these results. Further Mantel tests within the two largest Pacific Ocean population groupings did not detect significant IBD patterns ($p > 0.05$). Visualisation of population structure with a DAPC (α-score optimised to retain 22 PCs), revealed clear differentiation between all Pacific Ocean, and both Tanzanian and Iranian populations (Fig. 2a and b), when all individuals were analysed together. Further DAPC analyses involving separation of populations into their respective ocean basins further clarified the patterns observed. Analysis of all populations from the Pacific Ocean (Fig. 2c) revealed clear partitioning of the French Polynesian and Cook Islands oysters from all other populations, while animals sampled from Fiji and Tonga formed a single cluster. Similarly, individuals collected from PNG, Solomon Islands and the GBR formed a single cohesive group, as did oysters sampled from Indonesia, Taiwan and Vietnam. This pattern of separation was confirmed by testing for the actual number of discrete clusters using the BIC method, which was determined to be k = 8.

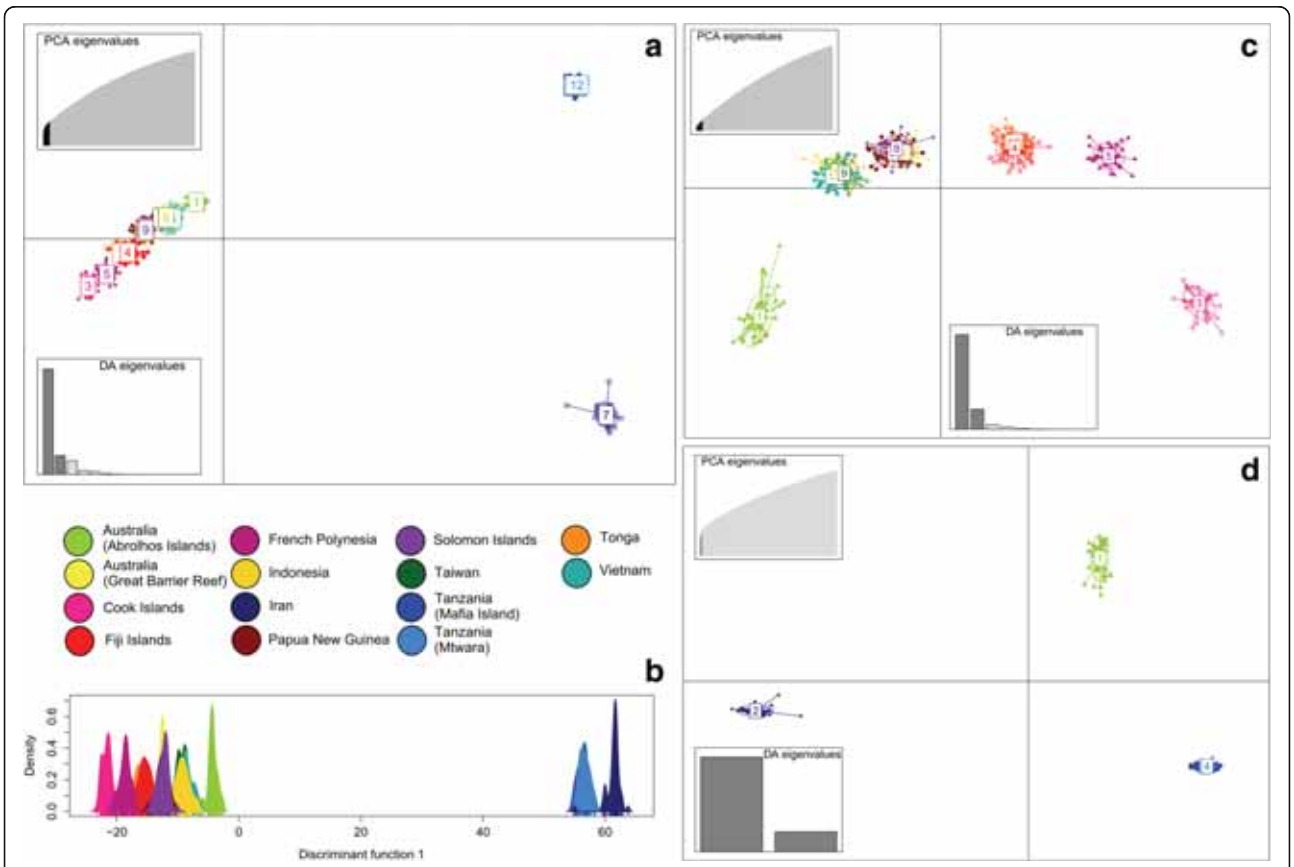


Fig. 2 Discriminant Analyses of Principal Components (DAPC) carried out using the *R* package *adegenet* to illustrate broad-scale patterns of population structure. Dots on scatterplots represent individuals, with colours denoting sampling origin and inclusion of 95% inertia ellipses. Scatterplot (**a**) was constructed among all 580 individuals collected from both the Pacific and Indian Ocean sites, while (**b**) is an individual density plot on the first discriminant function for this dataset. Scatterplots (**c**) and (**d**) were constructed on individuals sampled from Pacific Ocean (**c**) and Indian Ocean (**d**) sites only, to clearly identify regional differentiation

Examination of fine-scale population structure using Netview P (Fig. 3a and b) resolved similar patterns of differentiation to the DAPC, but offered greater resolution at the individual oyster level between several population pairs. In particular, when an organic network topology was used (k-NN = 40; Fig. 3a), it highlighted the degree of connectivity between the two broad clusters comprising oysters collected from Indonesia, Vietnam and Taiwan, along with individuals sampled from the GBR, Solomon Islands and PNG respectively. Analysis using a circular network topology (k-NN = 10; Fig. 3b) made this especially clear, as all individuals from these six locations collapsed into a single cluster. Interestingly, oysters collected from the Abrolhos Islands split into two subclusters (Fig. 3b), potentially indicating the presence of family groups, given that all individuals were sampled as a hatchery-produced cohort. Similarly, a closer relationship was apparent between French Polynesian, and Fijian-Tongan samples than with Cook Islands individuals, despite the greater geographic distance separating these populations. This may be due to prevailing ocean current patterns, which ensure greater connectivity through directional larval dispersal. Networks constructed at lower and higher k-NN thresholds all showed identical differentiation patterns.

Assessment of migration patterns and gene flow (Fig. 4) using *divMigrate* networks demonstrated nearly identical patterns of population structure between Indian (Fig. 4a) and Pacific (Fig. 4b) Ocean basins, when compared to the DAPC and Netview P networks. These similarities extended to closer examinations of Pacific Ocean populations by sub-region (Fig. 4c-e). Among Indian Ocean populations, directional migration between both Tanzanian sites was the strongest, but with very little connectivity between these two locations, Iran and the Abrolhos Islands. Connectivity within the Pacific region however, was substantially higher, with only the Cook Islands and French Polynesian populations remaining relatively isolated (Fig. 4b, e). Directional migration between Western Pacific sites (Vietnam, Indonesia, Taiwan, PNG, Solomon Islands and GBR) was found to be the strongest (Fig. 4c, e), followed by connectivity between the Fiji Islands and Tonga (Fig. 4d, e). Despite the geographic proximity of the Cook Islands to the Fiji Islands and Tonga, migration between both these locations and French Polynesia was considerably higher.

Examination of adaptive variation

F_{st} outlier tests discovered between 45 and 137 putatively directional, and 37–216 putatively balancing outlier loci jointly-identified by Bayescan 2.1 and LOSITAN, at six FDR thresholds for Pacific Ocean populations (Additional files 2 and 3). Both platforms failed to detect loci under balancing selection below an FDR = 0.01, and based on verification of loci detected at all FDR thresholds using QQ plots, a final stringent FDR threshold of 0.05 was selected. At this FDR, 89 directional and 37 balancing loci were jointly-identified, and used to construct NJ trees to visualise population structure at loci putatively under selection (Fig. 3c, d and e).

Weak population structure observed at selectively neutral and balancing loci (Fig. 3e and d respectively), correlated well with pairwise F_{st} and D_S comparisons. At directional loci however, clear divergence was evident between populations, which corresponded exactly with the five clusters identified by DAPC and Netview P networks in the Pacific Ocean. To gauge the strength of the selection signal, average Bayescan 2.1 F_{st} values among the 89 directional loci were examined, and found to equal 0.1915 (range = 0.1012 to 0.4371). Among the 37 balancing loci, average $F_{st} = -0.0066$ (range = −0.0114 to −0.0031), demonstrating that diffuse population structure (NJ trees Fig. 3e and d), becomes apparent when considering these and selectively neutral loci. These results indicate the likely presence of local adaptation acting on the populations examined, which is likely due to the heterogenous habitats occupied by *P. margaritifera* across the Pacific Ocean.

Particle dispersal modelling

Simulations of larval transport revealed a high degree of admixture by surface ocean currents within the Pacific basin over both 2014 and 2015 datasets, (Fig. 5 and see Additional files 4 a, b, c and d for animations of the full dispersal simulations). Interestingly, differences in the direction and extent of dispersal were observed between spawning seasons within either year, than between peak ENSO activity (2014 recorded an El Niño event, which dissipated in 2015). In particular, particles originating in both Taiwan and Vietnam were advected north towards Japan and the Ogasawara Islands archipelago during the first spawning seasons of both 2014 and 2015 (Additional file 1 a, c), while these current patterns reversed during the second spawning seasons, directing particles south across the Vietnamese coastline towards Malaysia (Additional file 1 b, d).

Overall patterns of population structure inferred from DAPC, Netview P and *divMigrate* analyses were highly concordant with simulated dispersal patterns for both ocean basins. At a broad scale, connectivity between the GBR, Solomon Islands, PNG, Indonesia, Vietnam and Taiwan was particularly obvious, together with the Fiji Islands and Tonga. Dispersal patterns for Indian Ocean sampling sites was limited to the Abrolhos Islands, where larval output is likely to spread northwards over much of the Western Australian seaboard (Fig. 5a and c). While providing unprecedented insights into the larval connectivity of *P. margaritifera*, these results should not be interpreted as reflecting actual recruitment over the

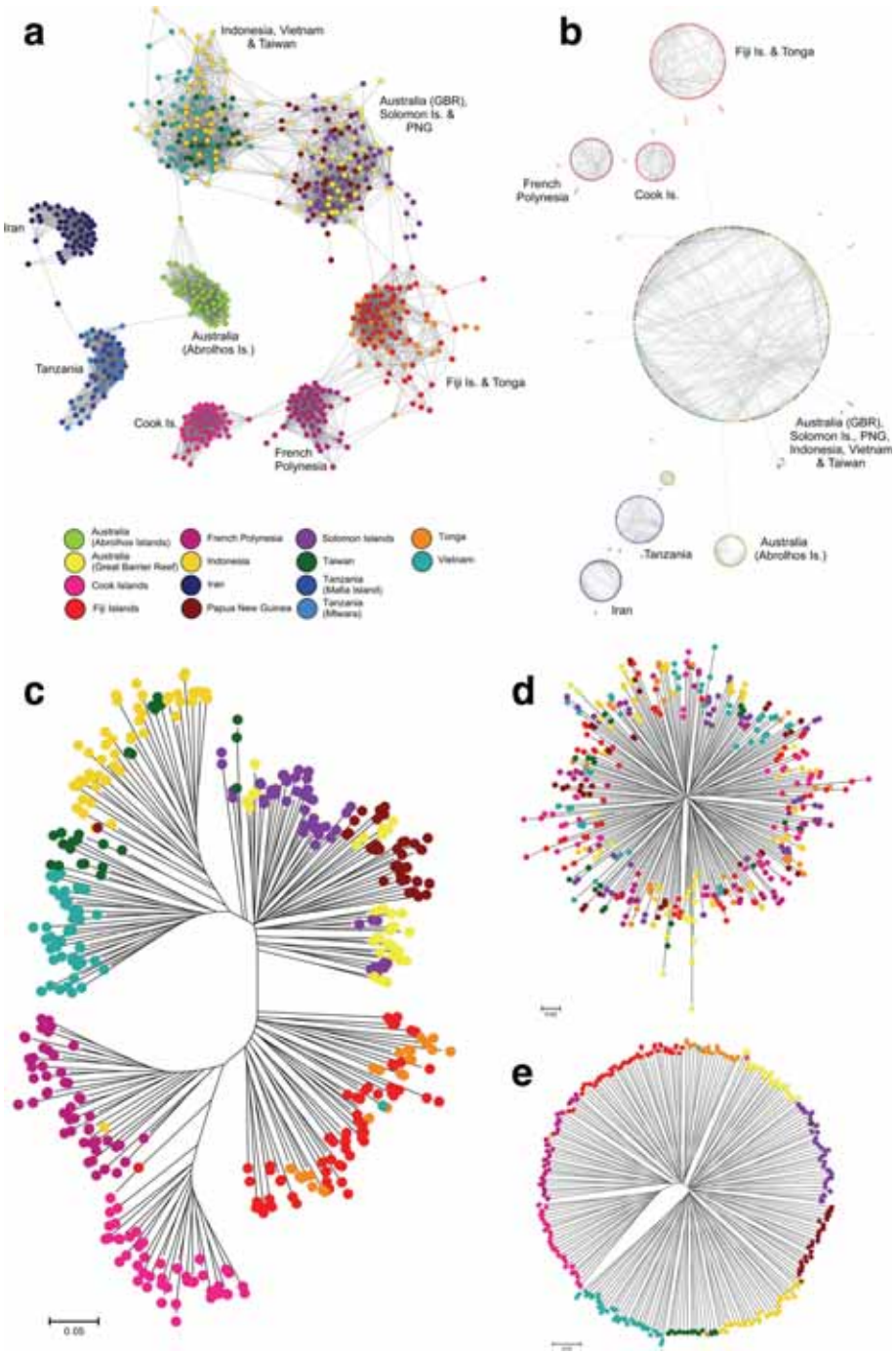


Fig. 3 Visualisation of population structure among 580 *P. margaritifera* individuals sampled. Fine-scale population networks constructed using the Netview P v.0.4.2.5 pipeline and selectively-neutral loci are shown in (**a**) organic; k-NN = 40 and (**b**) circular; k-NN = 10 topologies, with each dot representing a single individual. Oysters sampled from the Pacific Ocean had sufficiently low neutral F_{st} levels to permit testing for outlier loci, and Neighbour-Joining trees generated based on 1-psa distance matrices for these individuals are shown in (**c**) and (**d**). The tree displayed in (**c**) was drawn using 89 putatively directional outlier loci detected by both Bayescan 2.1 and LOSITAN at an FDR = 0.05, while (**d**) was generated using 37 also jointly-identified putatively balancing loci, at an FDR = 0.05. **e** Shows the arrangement of population structure in these same individuals, but with all loci (9,624 SNPs). The scale bars for (**c**), (**d**) and (**e**) indicate 1-psa genetic distance

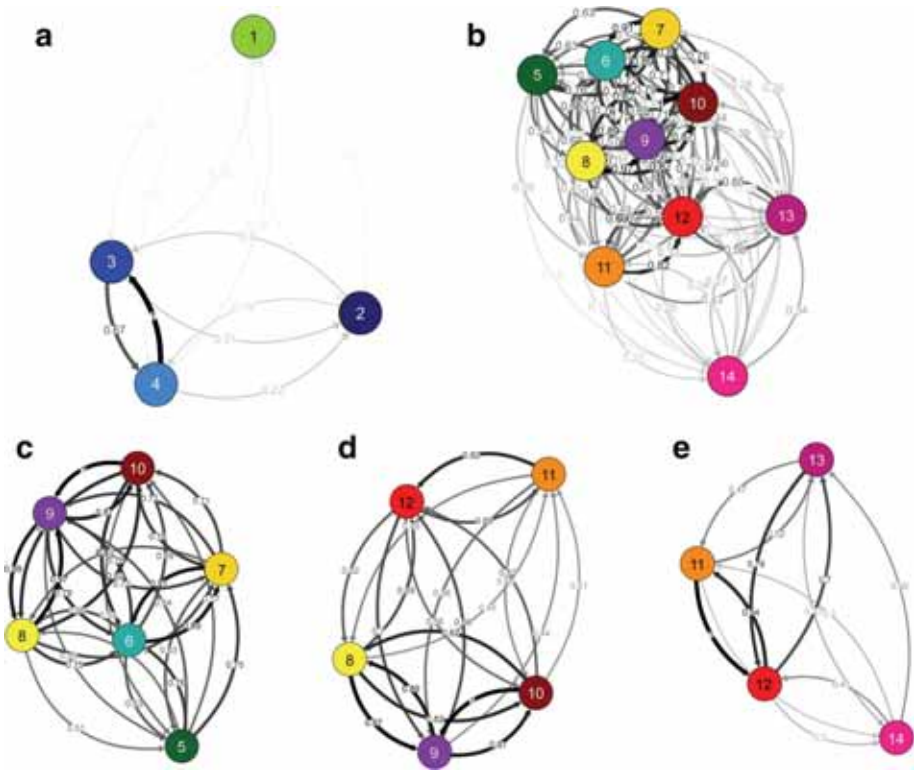


Fig. 4 Migration networks for *P. margaritifera* populations generated using the *divMigrate* function in *diveRsity* [72]. Circles represent populations, while arrows indicate the direction and magnitude (arrow edge values) of relative migration levels using Nei's G_{st} method [67, 81]. Darker arrows indicate stronger migration relationships compared to lighter arrows. Separate networks are shown for all Indian Ocean populations (**a**) and all Pacific Ocean populations (**b**) sampled. To better visualise separation between all Pacific Ocean populations, further networks have been generated for population groups located in the Western Pacific (**c**), Western and Central Pacific (**d**) and the Central and Eastern Pacific (**e**). All networks were generated following 1,000 bootstraps and all pairwise relationships are significant ($p < 0.01$). Population colour codes correspond to Figs. 1, 2 and 3, and have been numbered as follows. 1: Australia (Abrolhos Is.), 2: Iran; 3: Tanzania (Mafia Is.), 4: Tanzania (Mtwara), 5: Taiwan, 6: Vietnam, 7: Indonesia, 8: Australia (GBR), 9: Solomon Is., 10: Papua New Guinea, 11: Tonga, 12: Fiji Is., 13: French Polynesia and 14: Cook Is

limits of final particle positions. For example, because larval competency behaviour was not modelled, particles originating from the GBR transported into the South Tasman Sea are unlikely to survive due to unfavourable water temperatures in that region.

Discussion

This study examined range-wide population genetic structure and connectivity in the black-lip pearl oyster, over its ~18,000 km natural distribution. Assessments of differentiation at both neutral and adaptive markers, together with an independent particle dispersal simulation indicate that the evolutionary and physical processes organising population genetic structure are highly complex. At broad and regional scales, surface ocean currents, geographic distance and habitat geomorphology play important roles in regulating connectivity. At sub-regional and local scales, seascape features such as coral atolls, shoals and straits may impede gene flow, and the presence of environmental heterogeneity result in adaptive differences between populations.

In the Pacific Ocean, our observations do not lend support for a strong CPH model, where *P. margaritifera* is expected to exhibit reduced diversity and increased differentiation towards its range limits. However, this does not imply that CPH trends are absent, as very high levels of gene flow may conceal *C-M* gradients and sampling may not have detected the true range limits. The presence of local adaptation in habitat sub-regions also supports the presence of hetereogenous environments. Conversely in the Indian Ocean, clear divergence between the marginal populations sampled suggests the presence of *C-M* clines cannot be discounted, and requires further investigation at higher sampling densities, with particular attention to central populations. It is apparent that the mechanisms underlying range-wide genetic structure in *P. margaritifera* are quite complex, and require closer examination to better understand the evolutionary, ecological and physical factors at work.

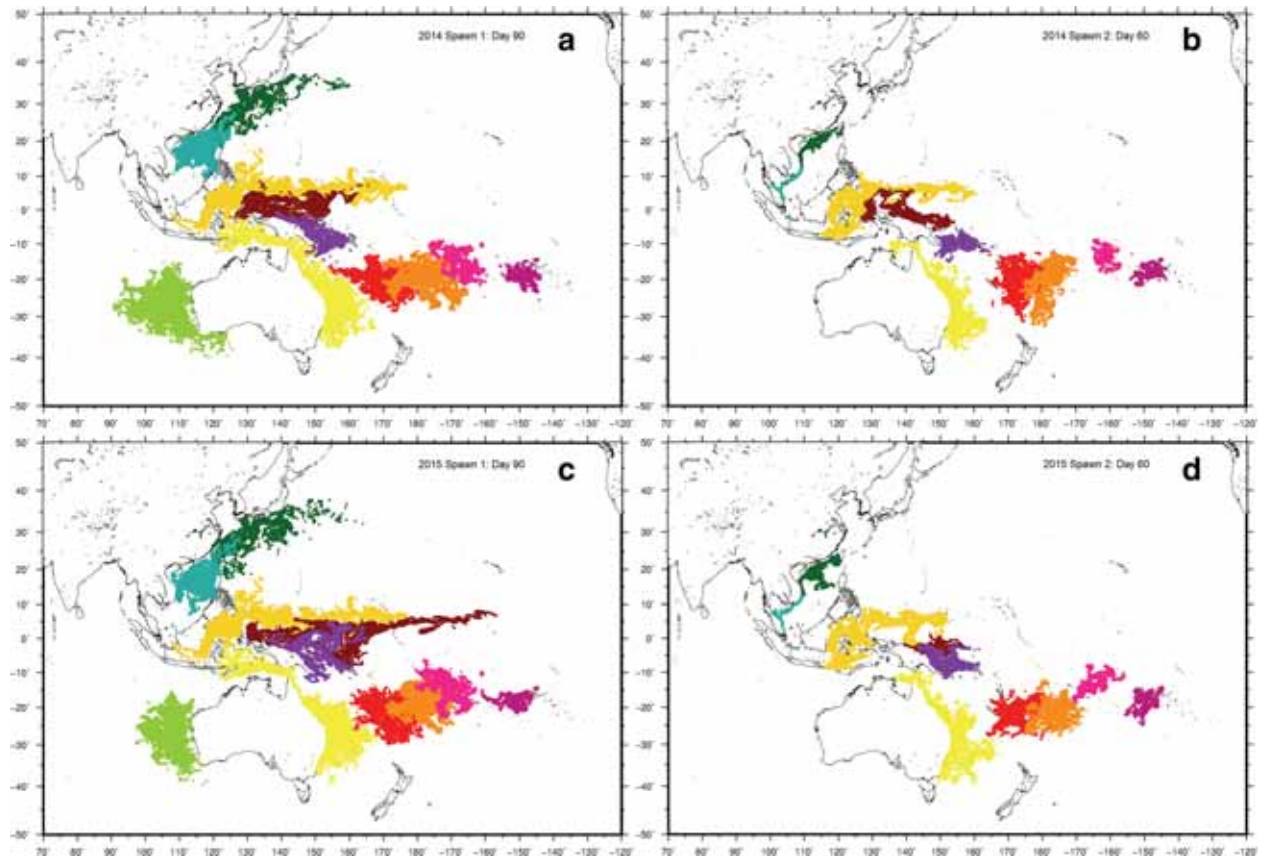


Fig. 5 Results of particle dispersal simulation for 11 sampling sites. Particle positions are displayed for the following simulations: spawning season 1 for 2014 (**a**), season 2 for 2014 (**b**), season 1 for 2015 (**c**) and season 2 for 2015 (**d**). All season 1 simulations were run for 90 days, and season 2 simulations over 60 days. Sampling site colour codes correspond with Figs. 1, 2, 3 and 4

Basin-wide population structure and connectivity

At a broad scale, *P. margaritifera* populations in the Indian and Pacific Oceans displayed substantial and significant divergence (pairwise F_{st} estimates = 0.2894–0.4684, $p < 0.001$). Strong population structure was evident within and between both ocean basins, however, due to the relative isolation of populations between these regions, each is discussed separately.

Pacific Ocean

Gene flow among Pacific Ocean populations appears to occur at a basin-wide scale, with pairwise F_{st} estimates reaching a maximum of 0.1090 (Cook Islands and Taiwan), over a distance of approximately 9,900 km. Despite the high degree of admixture among populations, visualisation of population structure (Figs. 2, 3, and 4) resolved five distinct genetic groups. When dispersal simulation data (Fig. 5 and Additional file 4 a-d) are compared to genetic differentiation patterns, the physical limits of simulated larval dispersal closely match population groupings. This observation suggests that while surface ocean currents permit sufficient gene flow across the Pacific Ocean to ensure populations retain a high degree of connectivity, circulation patterns and IBD may also facilitate regional larval retention, that stabilises population genetic structure. Because even low levels of gene flow [101, 102] are able to prevent population divergence, it is conceivable that standing genetic diversity and structure are maintained by a "founder takes all" density-dependent effect [103], where individuals arriving after an initial colonisation event may be "blocked" by established conspecifics [11, 103].

For the present study, at the geographical limits of the species distribution in the Pacific, decreased differentiation between Taiwan and French Polynesia (F_{st} = 0.0739) is evident despite the considerable distance involved (~11,000 km). This observation does not support generalised CPH predictions, and is likely a result of greater connectivity of this population pair through ocean current circulation [8, 104]. This is corroborated by dispersal simulation data (Additional file 4 a and c), and supported by pairwise migration analyses (Fig. 4). Larval competency following an extended pelagic dispersal phase is also expected to play a role in recruitment

success or failure, as individuals may have greater fitness as a result of shorter and potentially less stressful larval development [105, 106]. Here, ocean currents may impact recruitment rates by permitting increased larval fitness through reduced transport times, meaning that a population pair separated by greater physical distance may share higher connectivity, compared to a neighbouring population pair where larval plumes are vectored in mutually opposite directions or via circuitous pathways [12, 107].

Another factor influencing population structure and connectivity is habitat geomorphology, which is particularly evident in the Western Pacific, where long-range larval dispersal is restricted by the presence of numerous shoals, straits, islands, reefs and semi-enclosed seas [25]. This is reflected in the segregation of Taiwanese, Vietnamese and Indonesian individuals, from oysters collected in PNG, the Solomon Islands and the GBR (Figs. 2, 3 and 4). Similar patterns have been documented in several highly-dispersive marine taxa, ranging from a diatom [108] and limpet [109], to giant clam [110] and mullet [31].

Signatures of selection in the Pacific basin

Similarities in the patterns of population structure obtained at loci under directional selection (Fig. 3c-e), to spatial arrangements generated by DAPC and Netview P networks at selectively neutral loci (Figs. 2 and 3a-b), reinforce stock boundaries identified for *P. margaritifera* in the Pacific basin. The seascape of the Pacific region has been shaped by complex geological processes, giving rise to considerable habitat heterogeneity [111, 112]. Given the large extent of the species distribution sampled (>11,000 km), it is feasible that the selective differences observed may originate from distinct habitat sub-regions present within the Pacific basin [27, 113, 114].

For range-wide investigations of genetic structure in broadcast spawning marine species, consideration of adaptive variation can be important for uncovering functional differences between populations that might otherwise go undetected. As an example, adaptive divergence in the Atlantic cod related to temperature and salinity clines across the species distribution was detected by Nielsen et al. [33], but not evident within a restricted portion of its range [115], where environmental differences were predicted to be similar. Similarly, our previous study of *P. margaritifera* in the Fiji Islands failed to detect signatures of selection between and among populations [4]; however, results presented here indicate that detectable selection is evident only at the scale of Fijian and Tongan populations considered together.

In certain situations, adaptive differences in the face of high gene flow are the only discriminating factor through which concise fishery management is possible, by disentangling the effects of selection from demographic history, migration and genetic drift [24, 116, 117]. For example, Nayfa and Zenger [32] detected divergent selection between three Indonesian populations of the silver-lip pearl oyster *P. maxima* over ~2,000 km, where functional differences had manifested themselves in commercial fitness trait differences (namely growth rate and shell size [118]). Because the complex life histories of marine taxa may result in greater vulnerability to pre- and post-settlement selective forces [106, 119], the ability to detect these effects on the genetic composition of populations is critical for informing management for aquaculture, translocation, population supplementation and assisted migration [115, 120–122].

Indian Ocean

Populations sampled from the Indian Ocean displayed substantial vicariance, with the magnitude of separation between the three distinct genetic groups potentially indicating the presence of distinct ESUs, based on D_S estimates (Table 2; [123–125]). Work by Ranjbar et al. [126] and Cunha et al. [45], suggest that *Pinctada margaritifera* may in fact be a species complex, with populations in the Persian Gulf comprising a distinct ESU. Restriction of gene flow into the Persian Gulf from the greater Indian Ocean by the Strait of Hormuz likely isolates these individuals, and while the current study provides an initial assessment of basin-wide population differentiation for Indian Ocean *P. margaritifera*, further hierarchical sampling is required to determine regional patterns of evolutionary and contemporary genetic structure.

Particular attention to core populations from the central Indian Ocean (Maldives), Madagascar, Arabian Sea, Bay of Bengal, Andaman Sea and Sumatra may resolve these questions, and potentially ascertain the presence of a genetic break between the Indian and Pacific Oceans. Pairwise F_{st} estimates and visualisation of genetic structure between the closest marginal populations from the Western Pacific in the current dataset suggest this is a possibility (see Table 2 and Figs. 2, 3), as similar observations have been recorded for other invertebrate taxa [113, 127–129].

Patterns across the species' distribution

The CPH predicts that genetic diversity and connectivity should be highest at the centre of a species' range and decrease towards the periphery, however, our data indicate the presence of patterns which are substantially more complex than generalised CPH predictions. For Pacific populations, easily discernable *C-M* trends were absent, and may mean that the homogenising influence of basin-wide current circulation patterns disrupts any obvious patterns. However, ocean currents together with

isolation by geographic distance are also likely to maintain sub-regional population structure (e.g. Miller et al. [130] for the surf clam *Donax deltoides*).

Sample collection for the current study was organised according to the published theoretical distribution of *P. margaritifera* [35], and therefore it is possible that the true species distribution limit may not have been sampled, if it in fact extends beyond the current known range. If edge effects of decreased genetic diversity and marked differentiation are present, further sampling and analysis at the periphery of the species distribution in the Pacific Ocean may detect them. The levels of divergence between Indian Ocean oysters could reflect edge effects, considering that individuals were sampled from the ocean basin margins, however, as no central populations were able to be sampled, this observation cannot be substantiated. In addition to the CPH, other theoretical models for describing population organisation such as source-sink interactions, and range edge disequilibrium [6] warrant consideration. This is because for many species, range margins are often mobile with expansions and contractions over time, and are the result of numerous biotic and abiotic mechanisms [1, 5, 6].

Drivers of genetic structure and implications for fishery management

It is evident that the biological and physical processes governing population structure and genetic diversity in *P. margaritifera* are complex. In the Pacific Ocean, our data indicate that ocean currents, seascape features and geographic distances are major influences on population connectivity which both disrupts *C-M* clines, and simultaneously stabilises population structure according to basin sub-regions [27]. Broad-scale habitat geomorphology also plays an important role in differentiating populations, by restricting gene flow and influencing sub-regional natural selection. While our sampling scope in the Indian Ocean was insufficiently dense to determine the existence of *C-M* trends, ocean currents may play a large role in maintaining divergent populations. It is possible that a genetic break between the Indian and Pacific Oceans may exist at the South-East Asian archipelago, and further investigation of these populations could provide answers to this question, as it has for other marine invertebrates [127, 129]. Gauging the importance of oceanic circulation for driving population genetic structure and connectivity for *P. margaritifera* would not have been possible without simulations of larval dispersal, and we suggest that oceanographic and/or ecological modelling data is an indispensable component of range-wide investigations of genetic structure in marine organisms, which possess passively dispersing planktonic larvae [131, 132].

Data presented here do not support *P. m.* var. *typica* and *P. m.* var. *cummingi* as sub-species classifications in the Pacific Ocean, given the level of broad-scale admixture detected and absence of evidence for distinct ESUs. Unfortunately, as Hawaiian populations could not be sampled, no conclusion as to the status of *P. m.* var. *galstoffi* may be drawn. However, given the ability of larvae to disperse across the Pacific basin over the span of several generations, it is possible that Hawaiian populations may not be as divergent as previously thought [133]. Conversely, *P. m.* var. *zanzibarensis* and *P. m.* var. *persica* in the Indian Ocean may constitute distinct ESUs, given their substantial divergence from all other populations, although denser basin-wide sampling is required for verification. A comprehensive range-wide phylogenetic analysis of *P. margaritifera* is also needed to assess how many ESUs may be present, and to determine if the black-lip pearl oyster represents a true species complex. Because there are discernable regional morphological differences within *P. margaritifera*, there may be parallels with the Akoya species complex, which also displays morphological variability, high levels of gene flow and has a similarly extensive Indo-Pacific distribution [35, 50].

Conclusions

Our findings hold regional fishery management implications for Pacific populations of *P. margaritifera*, with the discovery of five distinct genetic stocks in the region. Given the economic importance of pearl oyster aquaculture for several Pacific Island nations [34, 134], this data provides a benchmark for further evaluation of fine-scale population structure at the level of individual countries and territories, to inform localised fishery management policies. Results presented here are also important for fishery management and aquaculture development in other broadcast spawning marine taxa, as an informed approach for designating stock boundaries relies on robust datasets comprising ecological, evolutionary and physical information.

Additional files

Additional file 1: Summary of temporal seed inputs for particle dispersal model [97, 135–140]. (DOC 32 kb)

Additional file 2: Numbers of putative directional and balancing F_{st} outlier loci discovered in *P. margaritifera*. Data are reported following testing of Pacific Ocean populations at six False Discovery Rate thresholds, using BayeScan 2.1 [82] and LOSITAN [84]. Jointly-identified loci were identified using both outlier detection platforms. (DOC 33 kb)

Additional file 3: Summary of numbers of both putatively balancing and directional SNPs detected. Loci are reported following testing of the entire dataset, to identify selectively-neutral SNPs. (DOC 29 kb)

Additional file 4: a. Animation of particle dispersal model simulation using 2014 HYCOM data for spawning season 1. Particle seed location colour codes for 11 populations are identical to those described in Fig. 1. **b.** Animation of particle dispersal model simulation using 2014 HYCOM

data for spawning season 2. Particle seed location colour codes for 10 populations are identical to those described in Fig. 1. **c.** Animation of particle dispersal model simulation using 2015 HYCOM data for spawning season 1. Particle seed location colour codes for 11 populations are identical to those described in Fig. 1. **d.** Animation of particle dispersal model simulation using 2015 HYCOM data for spawning season 2. Particle seed location colour codes for 10 populations are identical to those described in Fig. 1. (ZIP 13273 kb)

Additional file 5: Genotypic data. Genotypes of 580 individuals of *P. margaritifera* at 9,624 selectively neutral genome-wide SNPs are included in a standard STRUCTURE format. (ZIP 2057 kb)

Additional file 6: Genotypic data. Genotypes of 580 individuals of *P. margaritifera* at 10,683 adaptive and selectively neutral genome-wide SNPs are included in a standard STRUCTURE format. (ZIP 2276 kb)

Abbreviations

AMOVA: Analysis of molecular variance; BIC: Bayesian information criterion; CPH: Core-periphery hypothesis; DAPC: Discriminant analysis of principal components; DArT PL: Diversity arrays technology Ltd; DMSO: Dimethyl sulfoxide, C_2H_6OS; DNA: Deoxyribonucleic acid; DVM: Dorsoventral measurement; ENSO: El Niño Southern Oscillation; ESU: Evolutionary significant unit; EtBr: Ethidium bromide, $C_{21}H_{20}BrN_3$; FDR: False discovery rate; GBR: Great barrier reef; gDNA: Genomic deoxyribonucleic acid; HL: Homozygosity by locus; HWE: Hardy-Weinberg equilibrium; HYCOM: Hybrid coordinate ocean model; IBD: Isolation by distance; IR: Internal relatedness; JCU: James Cook University, Australia; k-NN: Number of nearest neighbours k-threshold; LD: Linkage disequilibrium; MAF: Minor allele frequency; MLH: Multi-locus heterozygosity; mtDNA: Mitochondrial deoxyribonucleic acid; MU: Management unit; NCBI: National centre for biotechnology information (U.S.A.); NJ: Neighbour-joining; PCR: Polymerase chain reaction; PIC: Polymorphic information content; PNG: Papua New Guinea; QQ-plot: Quantile-Quantile plot; RE: Restriction enzyme; SH: Standardised heterozygosity; SNP: Single nucleotide polymorphism

Acknowledgements

We wish to thank Shannon Kjeldsen, Eike Steinig, Maria Nayfa and Roger Huerlimann for advice on various statistical analyses including the Netview_P pipeline, provision of computing hardware, Linux scripting and outlier analyses. We also thank anonymous reviewers for their evaluation of the manuscript, and Litia Gaunavou for generation of the sampling site map. For either providing or assisting with collection of oyster tissue samples, our gratitude also extends to Gustaf Mamangkey, Naomi Gardiner, Ismail Saidi, Samantha Nowland, Rowan McIntyre, Steve Warden, Jo Buckee, Tina Weier, Georgia Langdon, Tevainui Frogier-Ellis, Hoc Tan Dao, Hứa Thái Tuyến, Samad Jahangard, Hossein Rameshi, Mehdi Doroudi, Max Wingfield, Laura Simmons, Gregory Bennett, Philippa Cohen, Waghon Lalao, Yu Wen Chiu, Eric Gan, Cherie Morris, Shirleen Bala, Justin Hunter, Pranesh Kishore, Adi Dionani Salaivanua, Kelly Brown, Jerome Taoi, Epeli Loganimoce, Albert Whippy, Bai Whippy, Toga Whippy, Isimeli Loganimoce, Marilyn Vilisoni, Babitu Rarawa, Ilitomasi Nuku, Samisoni Rakai, Patrick Fong, Nepoci Raleve and Claude Prévost. Logistical support for fieldwork in the Fiji Islands was provided by project partners the Secretariat of the Pacific Community (SPC) and the University of the South Pacific (USP).

Funding

This study was conducted within the Australian Centre for International Agricultural Research (ACIAR) Project FIS/2009/057: "Pearl Industry Development in the Western Pacific" led by the University of the Sunshine Coast. The research was carried out during a John Allwright Fellowship awarded to MML. The funding source (ACIAR) had no direct involvement in the study design, collection, analysis and interpretation of the data, nor the decision to submit this article for publication.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its Additional files.

Authors' contributions

MML carried out all tissue collections, laboratory bench work, participated in the investigation design and conceptualisation, developed modifications to the genotyping protocol, performed all data analyses and drafted the manuscript. PCS developed the broad project concept, participated in the investigation design, provided supervisory support, all project funding, advice on pearl oyster biology and ecology, and edited the manuscript. DRJ participated in the investigation design, provided supervisory support and edited the manuscript. CB developed the particle dispersal software and refined it for this investigation, participated in the investigation design, and edited the manuscript. KRZ participated in design and conceptualisation of the project, provided statistical advice and technical input on investigation design, provided supervisory support and edited the manuscript. Note for use of the HYCOM hydrodynamic model: funding for the development of HYCOM has been provided by the United States National Ocean Partnership Program and the Office of Naval Research. Data assimilative products using HYCOM are funded by the U.S. Navy. Computer time was made available by the DoD High Performance Computing Modernization Program, and the output is publicly available at http://hycom.org. All authors read and approved the final manuscript.

Competing interests

All authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The work described herein has been carried out (where appropriate) in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for animal experiments. All oysters were handled in accordance with James Cook University's animal ethics requirements and guidelines.

Author details

[1]Centre for Sustainable Tropical Fisheries and Aquaculture, and College of Science and Engineering, James Cook University, Townsville QLD 4811, QLD, Australia. [2]Australian Centre for Pacific Islands Research, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Maroochydore QLD 4558, QLD, Australia. [3]Geoscience Division, Secretariat of the Pacific Community, 241 Mead Road, Nabua, Suva, Fiji Islands.

Received: 8 September 2016 Accepted: 12 December 2016
Published online: 10 January 2017

References

1. Guo Q. Incorporating latitudinal and central-marginal trends in assessing genetic variation across species ranges. Mol Ecol. 2012;21(22):5396–403.
2. André C, Larsson LC, Laikre L, Bekkevold D, Brigham J, Carvalho GR, Dahlgren TG, Hutchinson WF, Mariani S, Mudde K, et al. Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. Heredity. 2011;106(2):270–80.
3. Waples RS, Punt AE, Cope JM. Integrating genetic data into management of marine resources: how can we do it better? Fish Fish. 2008;9(4):423–49.
4. Lal MM, Southgate PC, Jerry DR, Bosserelle C, Zenger KR. A parallel population genomic and hydrodynamic approach to fishery management of highly-dispersive marine invertebrates: the case of the Fijian black-lip pearl oyster *Pinctada margaritifera*. PloS ONE. 2016;11(8):e0161390.
5. Eckert CG, Samis KE, Lougheed SC. Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. Mol Ecol. 2008;17(5):1170–88.
6. Sexton JP, McIntyre PJ, Angert AL, Rice KJ. Evolution and Ecology of Species Range Limits. Annu Rev Ecol Evol Syst. 2009;40:415–36.
7. Brussard PF. Geographic Patterns and Environmental Gradients: The Central-Marginal Model in *Drosophila* Revisited. Annu Rev Ecol Syst. 1984;15(1):25–64.
8. Liggins L, Booth DJ, Figueira WF, Treml EA, Tonk L, Ridgway T, Harris DA, Riginos C. Latitude-wide genetic patterns reveal historical effects and contrasting patterns of turnover and nestedness at the range peripheries of a tropical marine fish. Ecography. 2015;38(12):1212–24.
9. Vucetich JA, Waite TA. Spatial patterns of demography and genetic processes across the species' range: Null hypotheses for landscape conservation genetics. Conserv Genet. 2003;4(5):639–45.
10. Hardie DC, Hutchings JA. Evolutionary ecology at the extremes of species' ranges. Environ Rev. 2010;18(NA):1–20.

11. Liggins L, Gleeson L, Riginos C. Evaluating edge-of-range genetic patterns for tropical echinoderms, *Acanthaster planci* and *Tripneustes gratilla*, of the Kermadec Islands, southwest Pacific. Bull Mar Sci. 2014;90(1):379–97.
12. Shanks AL. Pelagic Larval Duration and Dispersal Distance Revisited. Biol Bull. 2009;216(3):373–85.
13. Limborg MT, Helyar SJ, De Bruyn M, Taylor MI, Nielsen EE, Ogden ROB, Carvalho GR, Consortium FPT, Bekkevold D. Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). Mol Ecol. 2012;21(15):3686–703.
14. Hellberg ME, Burton RS, Neigel JE, Palumbi SR. Genetic assessment of connectivity among marine populations. Bull Mar Sci. 2002;70(1):273–90.
15. Broquet T, Viard F, Yearsley JM. Genetic drift and collective dispersal can result in chaotic genetic patchiness. Evolution. 2013;67(6):1660–75.
16. Strathmann MF, Strathmann RR. An Extraordinarily Long Larval Duration of 4.5 Years from Hatching to Metamorphosis for Teleplanic Veligers of *Fusitriton oregonensis*. Biol Bull. 2007;213(2):152–9.
17. Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM. Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. Mol Ecol. 2013;22(11):2953–70.
18. Thorpe JP, Solé-Cava AM, Watts PC. Exploited marine invertebrates: genetics and fisheries. Hydrobiologia. 2000;420(1):165–84.
19. Arnaud-Haond S, Vonau V, Rouxel C, Bonhomme F, Prou J, Goyard E, Boudry P. Genetic structure at different spatial scales in the pearl oyster (*Pinctada margaritifera cumingii*) in French Polynesian lagoons: beware of sampling strategy and genetic patchiness. Mar Biol. 2008;155(2):147–57.
20. Addison JA, Hart MW. Spawning, copulation and inbreeding coefficients in marine invertebrates. Biol Lett. 2005;1(4):450–3.
21. Gosling EM. Fisheries and management of natural populations. In: Gosling EM, editor. Marine Bivalve Molluscs. West Sussex, United Kingdom: Wiley; 2015. p. 270–324.
22. Gosling EM, Wilkins NP. Genetics of settling cohorts of *Mytilus edulis* (L.): Preliminary observations. Aquaculture. 1985;44(2):115–23.
23. Hare PM, Palumbi RS, Butman AC. Single-step species identification of bivalve larvae using multiplex polymerase chain reaction. Mar Biol. 2000;137(5):953–61.
24. Gaggiotti OE, Bekkevold D, Jørgensen HBH, Foll M, Carvalho GR, Andre C, Ruzzante DE. Disentangling the Effects of Evolutionary, Demographic, and Environmental Factors Influencing Genetic Structure of Natural Populations: Atlantic Herring as a Case Study. Evolution. 2009;63(11):2939–51.
25. Dao HT, Smith-Keune C, Wolanski E, Jones CM, Jerry DR. Oceanographic Currents and Local Ecological Knowledge Indicate, and Genetics Does Not Refute, a Contemporary Pattern of Larval Dispersal for The Ornate Spiny Lobster, *Panulirus ornatus* in the South-East Asian Archipelago. PLoS One. 2015;10(5):e0124568.
26. Pujolar JM, Bevacqua D, Andrello M, Capoccioni F, Ciccotti E, De Leo GA, Zane L. Genetic patchiness in European eel adults evidenced by molecular genetics and population dynamics modelling. Mol Phylogenet Evol. 2011;58(2):198–206.
27. Wood S, Paris CB, Ridgwell A, Hendy EJ. Modelling dispersal and connectivity of broadcast spawning corals at the global scale. Glob Ecol Biogeogr. 2014;23(1):1–11.
28. Thomas Y, Dumas F, Andréfouët S. Larval Dispersal Modeling of Pearl Oyster *Pinctada margaritifera* following Realistic Environmental and Biological Forcing in Ahe Atoll Lagoon. PLoS ONE. 2014;9(4):e95050.
29. Neo ML, Erftemeijer PLA, Beek KL, Maren DS, Teo SLM, Todd PA. Recruitment constraints in Singapore's fluted giant clam (*Tridacna squamosa*) population–A dispersal model approach. PLoS One. 2013;8(3):e58819.
30. Grosberg RK, Cunningham CW. Genetic structure in the sea: from populations to communities. In: Bertness MD, Gaines SD, Hay ME, editors. Marine Community Ecology. Sunderland, Massachussetts, USA: Sinauer Associates; 2001. p. 61–84.
31. Krück NC, Innes DI, Ovenden JR. New SNPs for population genetic analysis reveal possible cryptic speciation of eastern Australian sea mullet (*Mugil cephalus*). Mol Ecol Resour. 2013;13(4):715–25.
32. Nayfa MG, Zenger KR. Unravelling the effects of gene flow and selection in highly connected populations of the silver-lip pearl oyster (*Pinctada maxima*). Marine Genomics. 2016;28:99–106.
33. Nielsen EE, Hemmer-Hansen J, Poulsen NA, Loeschcke V, Moen T, Johansen T, Mittelholzer C, Taranger G-L, Ogden R, Carvalho GR. Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). BMC Evol Biol. 2009;9:276.
34. Southgate PC, Strack E, Hart A, Wada KT, Monteforte M, Cariño M, Langy S, Lo C, Acosta-Salmón H, Wang A. Exploitation and Culture of Major Commercial Species. In: Southgate PC, Lucas JS, editors. The Pearl Oyster. Amsterdam: Elsevier; 2008. p. 303–55.
35. Wada KT, Tëmkin I. Taxonomy and Phylogeny: Commercial Species. In: Southgate PC, Lucas JS, editors. The Pearl Oyster. Amsterdam: Elsevier; 2008. p. 54–7.
36. SPC. Profiles of high interest aquaculture commodities for Pacific Islands countries. Noumea: Secretariat of the Pacific Community; 2003. p. 71.
37. Arnaud-Haond S, Bonhomme F, Blanc F. Large discrepancies in differentiation of allozymes, nuclear and mitochondrial DNA loci in recently founded Pacific populations of the pearl oyster *Pinctada margaritifera*. J Evol Biol. 2003;16(3):388–98.
38. Arnaud-Haond S, Vonau V, Bonhomme F, Boudry P, Blanc F, Prou J, Seaman T, Goyard E. Spatio-temporal variation in the genetic composition of wild populations of pearl oyster (*Pinctada margaritifera cumingii*) in French Polynesia following 10 years of juvenile translocation. Mol Ecol. 2004;13(7):2001–7.
39. Lemer S, Planes S. Effects of habitat fragmentation on the genetic structure and connectivity of the black-lipped pearl oyster *Pinctada margaritifera* populations in French Polynesia. Mar Biol. 2014;161(9):2035–49.
40. Lemer S, Planes S. Translocation of wild populations: conservation implications for the genetic diversity of the black-lipped pearl oyster *Pinctada margaritifera*. Mol Ecol. 2012;21(12):2949–62.
41. Benzie JAH, Ballment E. Genetic differences among black-lipped pearl oyster (*Pinctada margaritifera*) populations in the western Pacific. Aquaculture. 1994;127(2–3):145–56.
42. Durand P, Blanc F. Genetic diversity in a tropical marine bivalve - *Pinctada margaritifera* (Linne, 1758). Bull Societe Zool France - Evolution Et Zoologie. 1988;113(3):293–304.
43. Durand P, Wada KT, Blanc F. Genetic variation in wild and hatchery stocks of the black pearl oyster, *Pinctada margaritifera*, from Japan. Aquaculture. 1993;110(1):27–40.
44. Gervis MH, Sims NA. The Biology and Culture of Pearl Oysters (Bivalvia: Pteriidae), vol. 21. Manila: ICLARM; 1992.
45. Cunha RL, Blanc F, Bonhomme F, Arnaud-Haond S. Evolutionary Patterns in Pearl Oysters of the Genus *Pinctada* (Bivalvia: Pteriidae). Mar Biotechnol. 2011;13(2):181–92.
46. Jameson LH. On the Identity and Distribution of the Mother-of-Pearl Oysters; with a Revision of the Subgenus *Margaritifera*. Proc Zool Soc London. 1901;70(2):372–94.
47. Wada KT, Jerry DR. Population Genetics and Stock Improvement. In: Southgate PC, Lucas JS, editors. The Pearl Oyster. Amsterdam: Elsevier; 2008. p. 437–71.
48. Lal MM, Southgate PC, Jerry DR, Zenger KR. Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. Mar Genomics. 2016;25:57–68.
49. Lind CE, Evans BS, Taylor JJU, Jerry DR. Population genetics of a marine bivalve, *Pinctada maxima*, throughout the Indo-Australian Archipelago shows differentiation and decreased diversity at range limits. Mol Ecol. 2007; 16(24):5193–203.
50. Tëmkin I. Molecular phylogeny of pearl oysters and their relatives (Mollusca, Bivalvia, Pterioidea). BMC Evol Biol. 2010;10(1):1–28.
51. Yu DH, Chu KH. Genetic variation in wild and cultured populations of the pearl oyster *Pinctada fucata* from southern China. Aquaculture. 2006;258(1–4):220–7.
52. Alagarswami K, Dharmaraj S, Chellam A, Velayudhan TS. Larval and juvenile rearing of black-lip pearl oyster, *Pinctada margaritifera* (Linnaeus). Aquaculture. 1989;76(1–2):43–56.
53. Doroudi MS, Southgate PC. Embryonic and larval development of *Pinctada margaritifera* (Linnaeus, 1758). Molluscan Res. 2003;23(2):101–7.
54. Pechenik JA. Larval Experience and Latent Effects: Metamorphosis Is Not a New Beginning. Integr Comp Biol. 2006;46(3):323–33.
55. Dawson MN, Raskoff KA, Jacobs DK. Field preservation of marine invertebrate tissue for DNA analyses. Mol Mar Biol Biotechnol. 1998;7(2):145–52.
56. Adamkewicz SL, Harasewych MG. Systematics and biogeography of the genus *Donax* (Bivalvia: Donacidae) in eastern North America. Am Malacol Bull. 1996;13(1–2):97–103.
57. GE. Illustra AutoSeq G-50 and AutoSeq 96 dye terminator removal. Data file 28-9175-28. In: Illustra AutoSeq G-50 documents. Buckinghamshire, United Kingdom: GE Healthcare UK Limited; 2007. p. 1–4.
58. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C, et al. Diversity arrays technology: A generic genome profiling technology on open platforms. Methods Mol Biol. 2012;888:67–89.
59. Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A. Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. BMC Proc. 2011;5 Suppl 7:54–P54.

60. Harrang E, Lapègue S, Morga B, Bierne N. A High Load of Non-neutral Amino-Acid Polymorphisms Explains High Protein Diversity Despite Moderate Effective Population Size in a Marine Bivalve With Sweepstakes Reproduction. Genes Genomes Genetics. 2013;3(2):333–41.
61. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE. 2011;6(5):e19379.
62. Ren R, Ray R, Li P, Xu J, Zhang M, Liu G, Yao X, Kilian A, Yang X. Construction of a high-density DArTseq SNP-based genetic map and identification of genomic regions with segregation distortion in a genetic population derived from a cross between feral and cultivated-type watermelon. Mol Genet Genomics. 2015;290(4):1457–70.
63. Mateos JM, Pérez JP. Image Processing with ImageJ. Olton, Birmingham: GBR: Packt Publishing Ltd; 2013.
64. Cruz VMV, Kilian A, Dierig DA. Development of DArT Marker Platforms and Genetic Diversity Assessment of the U.S. Collection of the New Oilseed Crop *Lesquerella* and Related Species. PLoS ONE. 2013;8(5):e64062.
65. Robasky K, Lewis NE, Church GM. The role of replicates for error mitigation in next-generation sequencing. Nat Rev Genet. 2014;15(1):56–62.
66. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evol Bioinformatics Online. 2005;1:47–50.
67. Rousset F. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour. 2008;8(1):103–6.
68. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. [http://www.genetix.univ-montp2.fr/genetix/intro.htm]. Accessed 12 Dec 2016.
69. Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol Ecol Resour. 2014;14(1):209–14.
70. Alho JS, Välimäki K, Merilä J. Rhh: an R extension for estimating multilocus heterozygosity and heterozygosity–heterozygosity correlation. Mol Ecol Resour. 2010;10(4):720–2.
71. Slate J, David P, Dodds KG, Veenvliet BA, Glass BC, Broad TE, McEwan JC. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Heredity. 2004;93:255–65.
72. Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods Ecol Evol. 2013;4(8):782–8.
73. Kalinowski ST. Counting Alleles with Rarefaction: Private Alleles and Hierarchical Sampling Designs. Conserv Genet. 2004;5(4):539–43.
74. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. Peer J. 2014;2:e281.
75. Peakall ROD, Smouse PE. GENEALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes. 2006;6(1):288–95.
76. Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011;27(21):3070–1
77. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24(11):1403–5.
78. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010;11(1):94.
79. Steinig EJ, Neuditschko M, Khatkar MS, Raadsma HW, Zenger KR. NetView P: A network visualization tool to unravel complex population structure using genome-wide SNPs. Mol Ecol Resour. 2016:1–12
80. Neuditschko M, Khatkar MS, Raadsma HW. NETVIEW: A High-Definition Network-Visualization Approach to Detect Fine-Scale Population Structures from Genome-Wide Patterns of Variation. PLoS ONE. 2012;7(10):e48375.
81. Sundqvist L, Zackrisson M, Kleinhans D. Directional genetic differentiation and asymmetric migration. arXiv pre-print: arXiv:13040118v2. 2013;14.
82. Foll M. BayeScan v2.1 User Manual. Ecology. 2012;20:1450–62.
83. Foll M, Gaggiotti O. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics. 2008;180(2):977–93.
84. Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G. LOSITAN: A workbench to detect molecular adaptation based on a F_{st}-outlier method. BMC Bioinformatics. 2008;9(1):323.
85. White TA, Stamford J, Rus Hoelzel A. Local selection and population structure in a deep-sea fish, the roundnose grenadier (*Coryphaenoides rupestris*). Mol Ecol. 2010;19(2):216–26.
86. Pujolar JM, Jacobsen MW, Als TD, Frydenberg J, Munch K, Jónsson B, Jian JB, Cheng L, Maes GE, Bernatchez L, et al. Genome-wide single-generation signatures of local selection in the panmictic European eel. Mol Ecol. 2014;23(10):2514–28.
87. Kovach RP, Gharrett AJ, Tallmon DA. Genetic change for earlier migration timing in a pink salmon population. Proc R Soc B Biol Sci. 2012;279(1743):3870–8.
88. Narum SR, Hess JE. Comparison of FST outlier tests for SNP loci under selection. Mol Ecol Resour. 2011;11:184–94.
89. Gogarten SM, Bhangale T, Conomos MP, Laurie CA, McHugh CP, Painter I, Zheng X, Crosslin DR, Levine D, Lumley T, et al. GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. Bioinformatics. 2012;28(24):3329–31.
90. Cummings JA. Operational multivariate ocean data assimilation. Q J R Meteorol Soc. 2005;131(613):3583–604.
91. Chassignet EP, Hurlburt HE, Smedstad OM, Halliwell GR, Hogan PJ, Wallcraft AJ, Baraille R, Bleck R. The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system. J Mar Syst. 2007;65(1–4):60–83.
92. Siegel DA, Mitarai S, Costello CJ, Gaines SD, Kendall BE, Warner RR, Winters KB. The stochastic nature of larval connectivity among nearshore marine populations. Proc Natl Acad Sci U S A. 2008;105(26):8974–9.
93. Siegel DA, Kinlan BP, Gaylord B, Gaines SD. Lagrangian descriptions of marine larval dispersion. Mar Ecol Prog Ser. 2003;260:83–96.
94. Viikmäe B, Torsvik T, Soomere T. Impact of horizontal eddy diffusivity on Lagrangian statistics for coastal pollution from a major marine fairway. Ocean Dyn. 2013;63(5):589–97.
95. Markey KL, Abdo DA, Evans SN, Bosserelle C. Keeping It Local: Dispersal Limitations of Coral Larvae to the High Latitude Coral Reefs of the Houtman Abrolhos Islands. PLoS ONE. 2016;11.
96. Halliwell GR. Evaluation of vertical coordinate and vertical mixing algorithms in the HYbrid-Coordinate Ocean Model (HYCOM). Ocean Model. 2004;7(3–4):285–322.
97. Saucedo PE, Southgate PC. Reproduction, Development and Growth. In: Southgate PC, Lucas JS, editors. The Pearl Oyster. Amsterdam: Elsevier; 2008. p. 133–86.
98. Song Z, Shu Q, Bao Y, Yin X, Qiao F. The prediction on the 2015/16 El Niño event from the perspective of FIO-ESM. Acta Oceanol Sin. 2015;34(12):67–71.
99. Varotsos CA, Tzanis CG, Sarlis NV. On the progress of the 2015–2016 El Niño event. Atmos Chem Phys. 2016;16(4):2007–11.
100. Wessel P, Smith WHF, Scharroo R, Luis J, Wobbe F. Generic Mapping Tools: Improved Version Released. Eos, Transactions American Geophysical Union. 2013;94(45):409–10.
101. Slatkin M. Isolation by Distance in Equilibrium and Non-Equilibrium Populations. Evolution. 1993;47(1):264–79.
102. Slatkin M. Gene Flow and the Geographic Structure of Natural Populations. Science. 1987;236(4803):787–92.
103. Waters JM, Fraser CI, Hewitt GM. Founder takes all: density-dependent processes structure biodiversity. Trends Ecol Evol. 2013;28(2):78–85.
104. Ganachaud A, Kessler W, Wijffels S, Ridgway K, Cai W, Holbrook N, Bowen M, Sutton P, Qiu B, Timmermann A, et al. Southwest Pacific Ocean Circulation and Climate Experiment (SPICE): Part I Scientific Background. In: CLIVAR Publication Series, NOAA OAR Special Report, vol. 111. Seattle: NOAA/OAR/PMEL; 2007. p. 1–46.
105. Shima JS, Swearer SE. The legacy of dispersal: larval experience shapes persistence later in the life of a reef fish. J Anim Ecol. 2010;79(6):1308–14.
106. Nosil P, Vines TH, Funk DJ. Perspective: Reproductive Isolation Caused by Natural Selection against Immigrants from Divergent Habitats. Evolution. 2005;59(4):705–19.
107. Simpson SD, Harrison HB, Claereboudt MR, Planes S. Long-Distance Dispersal via Ocean Currents Connects Omani Clownfish Populations throughout Entire Species Range. PLoS ONE. 2014;9(9):e107610.
108. Godhe A, Egardt J, Kleinhans D, Sundqvist L, Hordoir R, Jonsson PR. Seascape analysis reveals regional gene flow patterns among populations of a marine planktonic diatom. Proc R Soc Lond B Biol Sci. 2013;280(1773):20131599.
109. Hoffman JI, Peck LS, Linse K, Clarke A. Strong Population Genetic Structure in a Broadcast-Spawning Antarctic Marine Invertebrate. J Hered. 2011;102(1):55–66.
110. Kochzius M, Nuryanto A. 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. 2008;17(17):3775–87.

111. Sanciangco JC, Carpenter KE, Etnoyer PJ, Moretzsohn F. Habitat Availability and Heterogeneity and the Indo-Pacific Warm Pool as Predictors of Marine Species Richness in the Tropical Indo-Pacific. PLoS ONE. 2013;8(2):e56245.
112. Carpenter KE. An introduction to the oceanography, geology, biogeography, and fisheries of the tropical and subtropical western and central Pacific. In: Carpenter KE, Niem VH, editors. FAO Species Identification Guide for Fishery Purposes The Living Marine Resources of the Western Central Pacific, vol. 1. Rome: Food and Agriculture Organization of the United Nations (FAO), South Pacific Forum Fisheries Agency (FFA) and Norwegian Agency for International Development (NORAD); 1998. p. 1–19.
113. Barber PH, Palumbi SR, Erdmann MV, Moosa MK. Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. Mol Ecol. 2002;11(4):659–74.
114. Bruno JF, Selig ER. Regional Decline of Coral Cover in the Indo-Pacific: Timing, Extent, and Subregional Comparisons. PLoS One. 2007;2(8):e711.
115. Knutsen H, Olsen EM, Jorde PE, Espeland SH. Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. Mol Ecol. 2011;20(4):768.
116. Wang L, Liu S, Zhuang Z, Guo L, Meng Z, Lin H. Population Genetic Studies Revealed Local Adaptation in a High Gene-Flow Marine Fish, the Small Yellow Croaker (*Larimichthys polyactis*). PLoS ONE. 2013;8(12):e83493.
117. Bourret V, Kent MP, Primmer CR, Vasemaegi A, Karlsson S, Hindar K, McGinnity P, Verspoor E, Bernatchez L, Lien S. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). Mol Ecol. 2013;22(3):532–51.
118. Kvingedal R, Evans BS, Lind CE, Taylor JJU, Dupont-Nivet M, Jerry DR. Population and family growth response to different rearing location, heritability estimates and genotype × environment interaction in the silver-lip pearl oyster (*Pinctada maxima*). Aquaculture. 2010;304(1–4):1–6.
119. Nosil P, Funk DJ, Ortiz-Barrientos D. Divergent selection and heterogeneous genomic divergence. Mol Ecol. 2009;18:375–402.
120. Funk WC, McKay JK, Hohenlohe PA, Allendorf FW. Harnessing genomics for delineating conservation units. Trends Ecol Evol. 2012;27(9):489–96.
121. Miller MR, Brunelli JP, Wheeler PA, Liu S, Rexroad CE, Palti Y, Doe CQ, Thorgaard GH. A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. Mol Ecol. 2012;21(2):237–49.
122. Hemmer-Hansen J, Nielsen EE, Grønkjær P, Loeschcke V. Evolutionary mechanisms shaping the genetic population structure of marine fishes; lessons from the European flounder (*Platichthys flesus* L.). Mol Ecol. 2007;16(15):3104–18.
123. Nei M. Genetic Distance between Populations. Am Nat. 1972;106(949):283–92.
124. Nei M. Genetic distance and molecular phylogeny. In: Ryman N, Utter F, editors. Population Genetics and Fishery Management. Seattle: University of Washington Press; 1987. p. 193–223.
125. Kalinowski ST. Evolutionary and statistical properties of three genetic distances. Mol Ecol. 2002;11(8):1263–73.
126. Ranjbar MS, Zolgharnien H, Yavari V, Archangi B, Ali Salari M, Arnaud-Haond S, Cunha RL. Rising the Persian Gulf Black-Lip Pearl Oyster to the Species Level: Fragmented Habitat and Chaotic Genetic Patchiness in *Pinctada persica*. Evol Biol. 2016;43(1):131–43.
127. Williams ST, Jara J, Gomez E, Knowlton N. The Marine Indo-West Pacific Break: Contrasting the Resolving Power of Mitochondrial and Nuclear Genes. Integr Comp Biol. 2002;42(5):941–52.
128. Richards ZT, Berry O, van Oppen MJH. Cryptic genetic divergence within threatened species of *Acropora* coral from the Indian and Pacific Oceans. Conserv Genet. 2016;17(3):577–91.
129. Wörheide G, Epp LS, Macis L. Deep genetic divergences among Indo-Pacific populations of the coral reef sponge *Leucetta chagosensis* (Leucettidae): Founder effects, vicariance, or both? BMC Evol Biol. 2008;8(1):1–18.
130. Miller AD, Versace VL, Matthews TG, Montgomery S, Bowie KC. Ocean currents influence the genetic structure of an intertidal mollusc in southeastern Australia – implications for predicting the movement of passive dispersers across a marine biogeographic barrier. Ecol Evol. 2013;3(5):1248–61.
131. Liggins L, Treml EA, Possingham HP, Riginos C. Seascape features, rather than dispersal traits, predict spatial genetic patterns in co-distributed reef fishes. J Biogeogr. 2016;43(2):256–67.
132. Liggins L, Treml EA, Riginos C. Taking the Plunge: An Introduction to Undertaking Seascape Genetic Studies and using Biophysical Models. Geogr Compass. 2013;7(3):173–96.
133. Galstoff PS. Pearl and Hermes Reef, Hawaii: hydrographical and biological observations. Honolulu: Bernice P. Bishop Museum; 1933.
134. Ponia B. A review of aquaculture in the Pacific Islands 1998–2007: Tracking a decade of progress through official and provisional statistics. In: Aquaculture Technical Papers. 2nd ed. Noumea: Secretariat of the Pacific Community; 2010. p. 38.
135. Tranter DJ. Reproduction in Australian pearl oysters (Lamellibranchia). IV. *Pinctada margaritifera* (L.). Aust J Mar Freshwat Res. 1958;9:509–23.
136. Rose RA. A manual for the artificial propagation of the silverlip or goldlip pearl oyster, *Pinctada maxima* (Jameson) from Western Australia. 1990.
137. Rose RA, Baker SB. Larval and spat culture of the Western Australian silver- or goldlip pearl oyster, *Pinctada maxima* Jameson (Mollusca: Pteriidae). Aquaculture. 1994;126(1–2):35–50.
138. Arjarasirikoon U, Kruatrachue M, Sretarugsa P, Chitramvong Y, Jantataeme S. Gametogenic processes in the pearl oyster, *Pteria penguin* (Roding, 1798) (Bivalvia, Mollusca). J Shellfish Res. 2004;23:403–10.
139. Vilisoni MTJ. Recruitment patterns of molluscs in Savusavu Bay, Fiji with emphasis on the Blacklip Pearl Oyster, *Pinctada margaritifera* (Linnaeus, 1758). Suva, Fiji Islands: University of the South Pacific; 2012. MSc thesis.
140. Pouvreau S, Tiapari J, Gangnery A, Lagarde F, Garnier M, Teissier H, Haumani G, Buestel D, Bodoy A. Growth of the black-lip pearl oyster, *Pinctada margaritifera*, in suspended culture under hydrobiological conditions of Takapoto lagoon (French Polynesia). Aquaculture. 2000;184(1–2):133–54.