

# Population Structure, Genetic Connectivity, and Signatures of Local Adaptation of the Giant Black Tiger Shrimp (*Penaeus monodon*) throughout the Indo-Pacific Region

Nga T.T. Vu <sup>1,2,\*</sup>, Kyall R. Zenger<sup>1,2</sup>, Catarina N.S. Silva<sup>2</sup>, Jarrod L. Guppy<sup>1,2</sup>, and Dean R. Jerry<sup>1,2,3</sup>

<sup>1</sup>Australian Research Council Industrial Transformation Research Hub for Advanced Prawn Breeding, College of Science and Engineering, James Cook University, Townsville, Queensland, Australia

<sup>2</sup>Centre for Sustainable Tropical Fisheries and Aquaculture, College of Science and Engineering, James Cook University, Townsville, Queensland, Australia

<sup>3</sup>Tropical Futures Institute, James Cook University, Singapore

\*Corresponding author: E-mail: thithanhnga.vu1@my.jcu.edu.au.

Accepted: 11 September 2021

## Abstract

The giant black tiger shrimp (*Penaeus monodon*) is native to the Indo-Pacific and is the second most farmed penaeid shrimp species globally. Understanding genetic structure, connectivity, and local adaptation among Indo-Pacific black tiger shrimp populations is important for informing sustainable fisheries management and aquaculture breeding programs. Population genetic and outlier detection analyses were undertaken using 10,593 genome-wide single nucleotide polymorphisms (SNPs) from 16 geographically disparate Indo-Pacific *P. monodon* populations. Levels of genetic diversity were highest for Southeast Asian populations and were lowest for Western Indian Ocean (WIO) populations. Both neutral ( $n = 9,930$ ) and outlier ( $n = 663$ ) loci datasets revealed a pattern of strong genetic structure of *P. monodon* corresponding with broad geographical regions and clear genetic breaks among samples within regions. Neutral loci revealed seven genetic clusters and the separation of Fiji and WIO clusters from all other clusters, whereas outlier loci revealed six genetic clusters and high genetic differentiation among populations. The neutral loci dataset estimated five migration events that indicated migration to Southeast Asia from the WIO, with partial connectivity to populations in both oceans. We also identified 26 putatively adaptive SNPs that exhibited significant Pearson correlation ( $P < 0.05$ ) between minor allele frequency and maximum or minimum sea surface temperature. Matched transcriptome contig annotations suggest putatively adaptive SNPs involvement in cellular and metabolic processes, pigmentation, immune response, and currently unknown functions. This study provides novel genome-level insights that have direct implications for *P. monodon* aquaculture and fishery management practices.

**Key words:** SNP, prawn, aquaculture, population genetics, gene flow, DArTseq.

## Introduction

The evolutionary forces of gene flow, genetic drift, mutation, and natural selection, which are all influenced by organism-specific life history, collectively shape the genetic structure of wild populations (Matolweni et al. 2000; Hemmer-Hansen et al. 2007). Population genetics research provides valuable estimates regarding genetic diversity and local adaptation, which are equally important factors for the adaptive capacity of key species under climate change scenarios (Grant et al. 2017; Oleksiak 2019). Facilitation of appropriate fisheries management and aquaculture practices can also be enhanced

through population genetics research by providing novel insights into broodstock selection for breeding programs and discernment of hatchery stocks from wild populations (Williams and Hoffman 2009; Gjedrem 2012; Lorenzen et al. 2012; Bernatchez et al. 2017). In order to make accurate recommendations regarding aquaculture programs (e.g., selective breeding) or fishery management plans for commercially important species, it is important to comprehensively understand the population structure, genetic diversity, and local adaptation within and among broad-scale populations (Lal et al. 2017).

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

## Significance

Traditional genetic markers (e.g., microsatellite repeats and mitochondrial genomes) can have limited power to detect fine-scale genetic structure and subtle signatures of local adaptation both within and among populations of important aquaculture species. We determined the genetic population structure and potential signatures of local adaptation both within and among 16 geographically disparate Indo-Pacific populations of the black tiger shrimp *Penaeus monodon* by sequencing 131,929 unique single-nucleotide polymorphisms present within the genome of 532 individual wild-caught shrimp. This study provides novel genome-level insights about *P. monodon* genetic structure and signatures of local adaptation, which collectively assist in the advancement of global aquaculture breeding programs and fishery management practices.

The giant black tiger shrimp *Penaeus monodon* is native to and considered one of the most important aquaculture species throughout the Indo-Pacific region. The life history of *P. monodon* comprises an offshore planktonic larval phase (approximately 20 days) before migrating to nursery areas, such as in-shore areas and mangrove estuaries (Motoh 1985; Queiroga and Blanton 2004). For example, tagged juveniles and adult *P. monodon* from Trinity Inlet in north Queensland, Australia moved north between 5 and 100 km over a span of 10–111 days (Gribble et al. 2003). High gene flow has also been reported among *P. monodon* samples collected in Madagascar, Mozambique, and South Africa despite geographic distances up to 2,000 km (Forbes et al. 1999). Therefore, the capacity for moderate to high-levels of gene flow (i.e., genetic connectivity) exists among geographically disparate *P. monodon* populations.

To date, most studies that assessed population genetic structure and connectivity patterns of *P. monodon* throughout the Indo-Pacific region have used microsatellites, allozymes, and mitochondrial DNA (mtDNA) markers (Benzie et al. 2002; You et al. 2008; Waqairatu et al. 2012). These genetic markers can often have limited resolution power to detect fine-scale genetic structure and signatures of adaptation within and among populations (Nielsen et al. 2009; Allendorf et al. 2010). Despite limited resolution these previous studies demonstrated low-level of population differentiation (PD) throughout the Indo-Pacific Ocean, as well as lower genetic diversity in Indian Ocean than Pacific Ocean *P. monodon* populations; however, levels of predicted gene flow were not determined, nor were patterns of genetic connectivity among populations consistent across these studies (You et al. 2008; Waqairatu et al. 2012).

The recent development of high-throughput sequencing techniques permits determination of both genome-wide neutral and adaptive genetic variation for nonmodel species, which can provide unique insight into population structure and local adaptation. For example, analysis of neutral genetic loci has been utilized for population structure determination (Batista et al. 2016; Segovia et al. 2017; Woodings et al. 2018), whereas outlier detection methods permit the separation of putatively adaptive (i.e., outlier) and neutral loci into

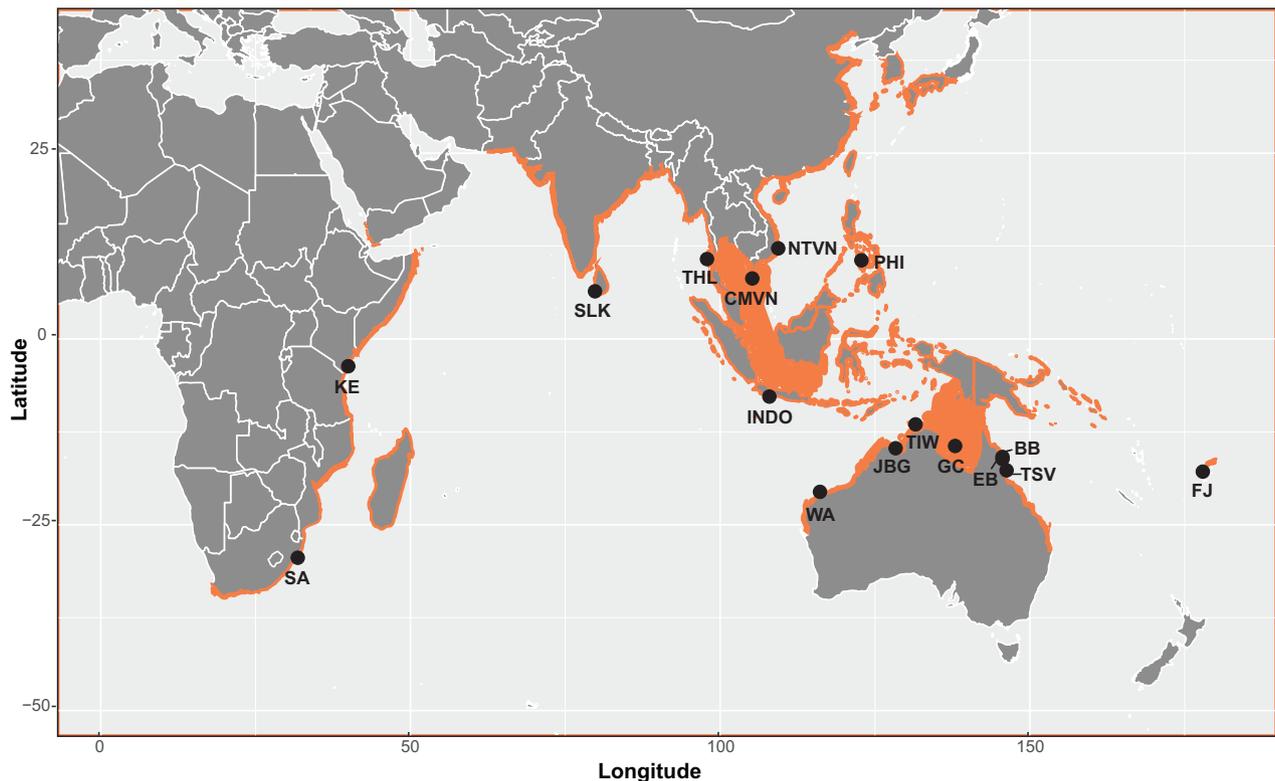
two discrete datasets. Recent studies have demonstrated that outlier loci provide unique and useful insights into the genetic structure and management units of aquatic species that neutral loci often do not show (Gagnaire et al. 2015; Woodings et al. 2018; Anderson et al. 2019; Nowland et al. 2019).

Collectively, the high-resolution power of genome-wide single nucleotide polymorphism (SNP) genotyping offers the potential for heightened understanding of *P. monodon* population genetic structure, connectivity, and local adaptation, which could translate into advancements in aquaculture and fishery practices. Accordingly, the main objective of this study was to utilize both neutral and outlier SNP datasets to determine the population structure, connectivity, and local adaptation of *P. monodon* populations from throughout the Indo-Pacific region (fig. 1 and [supplementary table S1, Supplementary Material online](#)). More specifically, this study aimed to: 1) evaluate genetic diversity and determine the spatial population structure among *P. monodon* populations using a combined (i.e., neutral and outlier loci) dataset, 2) determine the genetic connectivity and evolutionary relationship among populations, and 3) identify putatively adaptive outlier SNP (i.e., candidate genes).

## Results

### Genotyping and SNP Quality Control

A total of 131,929 unique SNPs were obtained for each of 532 *P. monodon* individuals (fig. 1) using DArTseq<sup>TM</sup>. Initial quality control assessment of DArTseq<sup>TM</sup> data using *dartQC* removed 119,514 SNPs, or 90.6% of loci ([supplementary table S2, Supplementary Material online](#)). The remaining 12,415 loci were then tested for linkage disequilibrium (LD) and Hardy–Weinberg equilibrium (HWE), which removed an additional 1,822 SNPs, or 1.4% of loci. Following genotype filtering, 12 individuals (nine from Kenya, one from Philippines, one from Sri Lanka, and one from Western Australia) were removed due to high rates of missing data (>26%). A final set comprised 10,593 SNPs (with missing data per SNP < 21% and per individual < 26%) for each retained *P. monodon* individual ( $n = 520$ ) was subjected to further analyses.



**FIG. 1.**—Sampling sites and the known distribution of *Penaeus monodon* sourced in this study. Orange areas denote the known distribution of *P. monodon* (modified from Kongkeo [2005–2020]). See Schott et al. (2009) for detailed Indo-Pacific summer and winter monsoon currents maps. FJ: Fiji; Eastern Australia (EA); BB: Bramston Beach, EB: Etty Bay; TSV: Townsville; Northern Australia (NT); GC: Gulf of Carpentaria, JBG: Joseph Bonaparte Gulf, TIW: Tiwi Island; Western Australia (WA); Vietnam (CMVN: Ca Mau, NTVN: Nha Trang); PHI: Philippines; INDO: Indonesia; THL: Thailand; SLK: Sri Lanka; KE: Kenya; SA: South Africa. Map generated using R version 4.0.3.

### Population Genetic Diversity

The dataset of 10,593 SNPs for each *P. monodon* individual ( $n = 520$ ) exhibited a mean allelic richness ( $A_R$ ) per location that ranged from 1.10 in SA to 1.62 in INDO (table 1). The mean private allelic richness ( $A_{PR}$ ) per population ranged from 0.002 in BB to 0.014 in SLK. The percentage of polymorphic loci (PPL) ranged from 18% to 76% for KE and TIW populations, respectively. Average observed heterozygosity ( $H_O$ ) across populations ranged from 0.02 to 0.15, whereas average expected heterozygosity ( $H_E$ ) ranged from 0.10 to 0.17. Of note is that  $H_O$  and  $H_E$  values across 14 of 16 populations were almost the same (0.01 difference), whereas two Western Indian Ocean (WIO) populations (KE and SA) exhibited a 5-fold difference between their  $H_O$  and  $H_E$  values (0.02 and 0.10), respectively (table 1). Samples from SA and KE populations exhibited the lowest values for  $H_E$  (0.10), average multilocus heterozygosity (Av. MLH = 0.023), and standardized multilocus heterozygosity (sMLH = 0.20), whereas their average internal relatedness (IR) was much higher than within other populations. Inbreeding coefficient ( $F_{IS}$ ) values among sampling locations varied from  $-0.06$  in WA to 0.80 in KE and 0.78 in SA. Among populations, the four Southeast Asia

populations (NTVN, CMVN, INDO, and THL) displayed higher values for observed  $H_E$  (0.16–0.17), average multilocus heterozygosity (Av. MLH: 0.15), and sMLH (1.19–1.25) than other populations (table 1).

### Detection of Outlier Loci

Outlier loci were determined from the dataset of 10,593 SNPs for each *P. monodon* individual ( $n = 520$ ) using two independent approaches: BAYESCAN and *PCAdapt* (see Materials and Methods). BAYESCAN and *PCAdapt* analyses (FDR = 0.01 for both) identified 965 SNPs and 1,252 SNPs as being under putative selection (associated with positive alpha values), respectively. Of these SNPs 663 were overlapping and, thus, deemed the outlier loci dataset, whereas the remaining SNPs ( $n = 9,930$ ) were deemed the neutral loci data set.

### Genetic Population Structure

Cross-validation by ADMIXTURE inferred the lowest cross-validation errors (ADMIXTURE approach) ranged from  $K = 6$  to  $K = 8$  for the genetic structure of neutral and outlier loci (supplementary fig. S1, Supplementary Material online). We

**Table 1**  
Genetic Diversity Indices among 16 Populations of *Penaeus monodon* Assessed Using All 10,593 High-Quality-Filtered SNP Loci

Population	<i>n</i>	$A_R$ ( $\pm$ SD)	$A_{PR}$ ( $\pm$ SD)	PPL (Av. MAF)	$H_0$ ( $\pm$ SD)	$H_E$ ( $\pm$ SD)	Av. MLH ( $\pm$ SD)	sMLH ( $\pm$ SD)	IR ( $\pm$ SD)	$F_{IS}$ ( $P < 0.05$ )
Fiji (FJ)	49	1.39 $\pm$ 0.45	0.008 $\pm$ 0.074	0.48	0.11 $\pm$ 0.16	0.12 $\pm$ 0.17	0.109 $\pm$ 0.005	0.88 $\pm$ 0.04	0.04 $\pm$ 0.02	0.06
Bramston Beach (BB)	51	1.49 $\pm$ 0.43	0.002 $\pm$ 0.019	0.64	0.12 $\pm$ 0.15	0.13 $\pm$ 0.16	0.117 $\pm$ 0.005	0.95 $\pm$ 0.04	0.003 $\pm$ 0.027	0.07
Ety Bay (EB)	31	1.48 $\pm$ 0.45	0.002 $\pm$ 0.027	0.58	0.12 $\pm$ 0.16	0.13 $\pm$ 0.16	0.118 $\pm$ 0.003	0.95 $\pm$ 0.03		
-0.003 $\pm$ 0.021	0.06									
Townsville (TSV)	22	1.49 $\pm$ 0.45	0.005 $\pm$ 0.049	0.56	0.12 $\pm$ 0.16	0.13 $\pm$ 0.16	0.117 $\pm$ 0.003	0.95 $\pm$ 0.03	0.004 $\pm$ 0.020	0.07
Gulf of Carpentaria (GC)	30	1.54 $\pm$ 0.44	0.003 $\pm$ 0.031	0.65	0.13 $\pm$ 0.16	0.14 $\pm$ 0.16	0.131 $\pm$ 0.008	1.06 $\pm$ 0.07	-0.03 $\pm$ 0.04	0.07
Joseph Bonaparte Gulf (JBG)	34	1.54 $\pm$ 0.43	0.003 $\pm$ 0.030	0.66	0.13 $\pm$ 0.16	0.14 $\pm$ 0.16	0.132 $\pm$ 0.004	1.07 $\pm$ 0.03	-0.04 $\pm$ 0.02	0.07
Tiwi Island (TIW)	50	1.55 $\pm$ 0.42	0.003 $\pm$ 0.024	0.71	0.13 $\pm$ 0.15	0.14 $\pm$ 0.16	0.131 $\pm$ 0.004	1.06 $\pm$ 0.03	-0.03 $\pm$ 0.02	0.08
Western Australia (WA)	22	1.46 $\pm$ 0.47	0.006 $\pm$ 0.060	0.5	0.14 $\pm$ 0.19	0.13 $\pm$ 0.17	0.138 $\pm$ 0.03	1.13 $\pm$ 0.26	-0.03 $\pm$ 0.06	-0.06
Philippines (PHI)	12	1.49 $\pm$ 0.50	0.005 $\pm$ 0.059	0.49	0.14 $\pm$ 0.19	0.14 $\pm$ 0.17	0.138 $\pm$ 0.004	1.12 $\pm$ 0.03	-0.06 $\pm$ 0.01	0.01
Nha Trang Vietnam (NTVN)	30	1.58 $\pm$ 0.44	0.007 $\pm$ 0.065	0.68	0.15 $\pm$ 0.17	0.16 $\pm$ 0.17	0.148 $\pm$ 0.004	1.20 $\pm$ 0.03	-0.11 $\pm$ 0.02	0.08
Ca Mau Vietnam (CMVN)	39	1.61 $\pm$ 0.40	0.004 $\pm$ 0.031	0.76	0.15 $\pm$ 0.16	0.16 $\pm$ 0.17	0.146 $\pm$ 0.005	1.19 $\pm$ 0.04	-0.10 $\pm$ 0.02	0.10
Java Indonesia (INDO)	38	1.62 $\pm$ 0.40	0.004 $\pm$ 0.035	0.76	0.15 $\pm$ 0.16	0.17 $\pm$ 0.17	0.150 $\pm$ 0.013	1.22 $\pm$ 0.11	-0.10 $\pm$ 0.03	0.09
Thailand (THL)	46	1.54 $\pm$ 0.46	0.010 $\pm$ 0.078	0.60	0.15 $\pm$ 0.18	0.16 $\pm$ 0.17	0.154 $\pm$ 0.005	1.25 $\pm$ 0.04	-0.13 $\pm$ 0.02	0.03
Sri Lanka (SLK)	19	1.49 $\pm$ 0.47	0.014 $\pm$ 0.090	0.53	0.11 $\pm$ 0.16	0.12 $\pm$ 0.16	0.113 $\pm$ 0.020	0.92 $\pm$ 0.16	0.01 $\pm$ 0.07	0.09
Kenya (KE)	16	1.11 $\pm$ 0.30	0.010 $\pm$ 0.076	0.18	0.02 $\pm$ 0.08	0.10 $\pm$ 0.28	0.023 $\pm$ 0.002	0.20 $\pm$ 0.02	0.65 $\pm$ 0.03	0.80
South Africa (SA)	31	1.10 $\pm$ 0.28	0.005 $\pm$ 0.055	0.19	0.02 $\pm$ 0.08	0.10 $\pm$ 0.27	0.024 $\pm$ 0.001	0.20 $\pm$ 0.01	0.65 $\pm$ 0.02	0.78

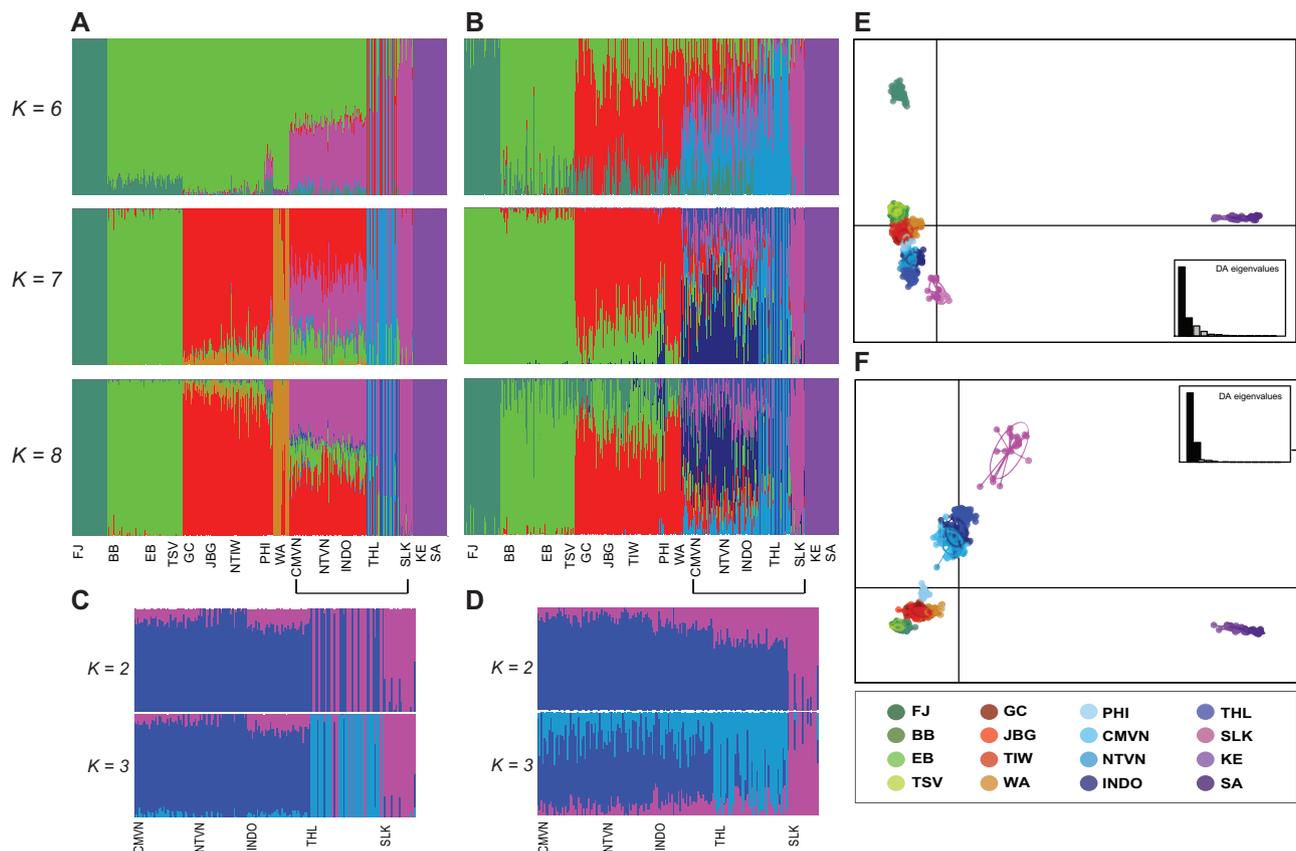
Note.—*n*, number of samples;  $A_R$ , mean allelic richness;  $A_{PR}$ , private allelic richness; PPL, percentage of polymorphic loci; Av. MAF, average minor allele frequency of polymorphic loci;  $H_0$ , mean observed heterozygosity;  $H_E$ , average expected heterozygosity; Av. MLH, average multilocus heterozygosity; sMLH, standardized multilocus heterozygosity; IR, internal relatedness;  $F_{IS}$ , significant ( $P < 0.05$ ) inbreeding coefficient; SD, standard deviation.

visualized clustering patterns for  $K$  values 6 through 8 to identify hierarchical clustering patterns as  $K$  increases (fig. 2A and B).

At  $K=6$ , for both neutral and outlier datasets, *P. monodon* from FJ and WIO clusters were separated from the rest of the sampled populations (fig. 2A and B). In addition, at  $K=6$ , based on the neutral loci, all seven Australian and Southeast Asia populations (PHI, CMVN, NTVN, and INDO) exhibited a high degree of admixture, whereas the THL population divided into three distinct groups (includes two single clusters). At  $K=7$ , for neutral loci, the main signals derived from ADMIXTURE approach was: 1) separation of FJ and WIO clusters from the rest of the sampled populations, 2) clustering of Northern Australia and PHI populations (NAP cluster; GC, JBG, TIW, and PHI), and 3) overlapping genetic clusters for five populations (CMVN, NTVN, INDO, THL, and SLK) (figs. 2A and 4A). At  $K=7$  and  $K=8$  for outlier loci, shrimp from five populations (CMVN, NTVN, INDO, PHI, and SLK) exhibited some degree of admixture with six genetic clusters (fig. 2B). The admixture between five populations (CMVN, NTVN, INDO, THL, and SLK) was further investigated in a substructure analysis for both neutral and outlier loci datasets (fig. 2C and D, see Materials and Methods), which revealed THL to be an admixture between two clusters: the SLK population and Southeast Asia group (CMVN, NTVN, and INDO).

The patterns identified by ADMIXTURE analysis were supported by the discriminant analysis of principal components (DAPC) analysis (fig. 2E and F). The Bayesian Information Criterion (BIC) statistic generated by DAPC indicated that the optimal number of clusters for neutral and outlier loci was  $K=7$  and  $K=6$ , respectively (supplementary fig. S1, Supplementary Material online). As with the ADMIXTURE results for neutral loci, DAPC confirmed the presence of seven clusters: 1) Fiji (FJ); 2) Eastern Australia cluster (EA; BB, EB, and TSV); 3) NAP cluster (GC, JBG, TIW, and PHI); 4) Western Australia (WA); 5) Southeast Asia cluster (CMVN, NTVN, INDO, and THL); 6) Sri Lanka (SLK); and 7) WIO cluster (KE and SA). The results of DAPC based on the outlier dataset revealed the following six clusters: 1) FJ; 2) EA; 3) NAP and WA cluster (GC, JBG, TIW, PHI, and WA); 4) Southeast Asia; 5) SLK; and 6) WIO (fig. 2).

The majority of population-pairwise  $F_{ST}$  values differed significantly from one another ( $P < 0.01$ ) for both neutral and outlier loci datasets with the following three exceptions: 1) KE and SA populations in both neutral and outlier loci datasets, 2) JBG and TIW populations in neutral loci dataset, and 3) EA group (BB, EB, and TSV) in outlier loci dataset (supplementary table S3, Supplementary Material online). The highest pairwise  $F_{ST}$  values occurred between the WIO (KE and SA) and all the Pacific Ocean populations, with  $F_{ST}$  values ranging from 0.33 to 0.56 and 0.51 to 0.96 for neutral and outlier loci datasets, respectively ( $P < 0.01$ ). Interestingly, in both neutral and outlier datasets, pairwise

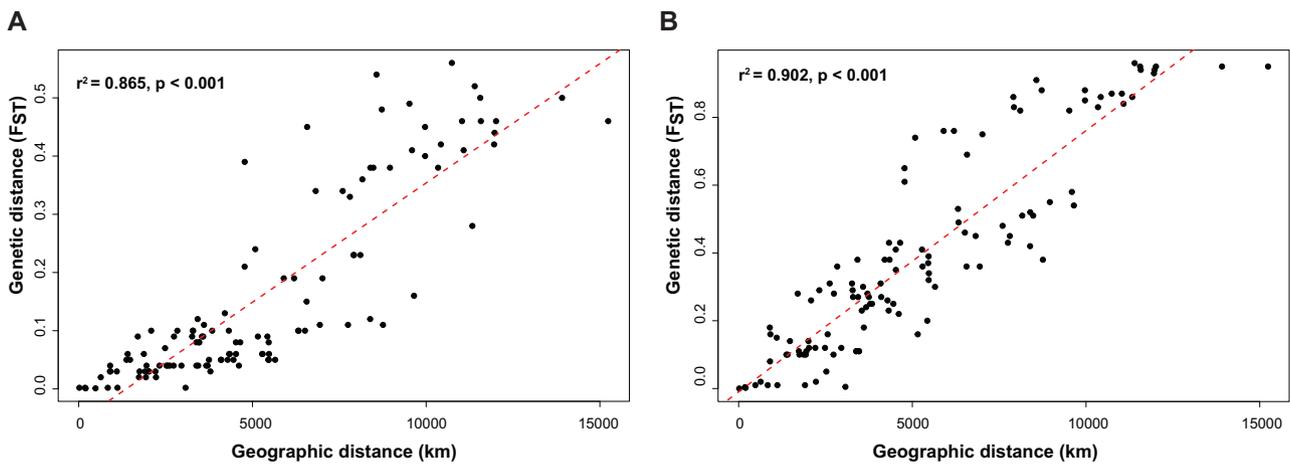


**Fig. 2.**—Population structure results for *Penaeus monodon* populations ( $n = 16$ ) using neutral loci (A, C, E) and outlier loci (B, D, F). Plots of individual admixture determined using ADMIXTURE with cross-validation error using three unique  $K$  values ( $K = 6$ ,  $K = 7$ , and  $K = 8$ ) for both neutral (A) and outlier (B) loci. Assignment of Southeast Asia group and SLK to cluster at  $K = 2$  and  $K = 3$  as inferred ADMIXTURE analysis using neutral loci (C) and outlier loci (D). DAPC analysis plots with BIC recommended  $K$  value of  $K = 7$  for neutral loci (E) and  $K = 6$  for outlier loci (F). Individuals of the same color belong to the same cluster. Population names are defined in figure 1.

$F_{ST}$  values between WIO and Fiji, as well as between WIO and Australian populations were higher than  $F_{ST}$  values observed among WIO and Southeast Asian populations (CMVN, NTVN, INDO, and THL). We also observed that geographically proximal populations exhibited the lowest  $F_{ST}$  values (e.g., between KE and SA and among EA populations; [supplementary table S3, Supplementary Material online](#)). Overall, average  $F_{ST}$  was higher for outlier loci ( $F_{ST} = 0.40 \pm 0.29$ ) than neutral loci ( $F_{ST} = 0.16 \pm 0.16$ ) ([supplementary table S3, Supplementary Material online](#)).

A series of Mantel tests were conducted for both neutral and outlier datasets in different subsets of sampling sites to assess the correlation between geographic and genetic distance. Significant isolation by distance (IBD) was found when all 16 locations were considered for both neutral and outlier datasets ( $r^2 = 0.865$  and  $r^2 = 0.902$ ,  $P < 0.001$ , respectively; [fig. 3](#)). Similarly, IBD relationships among the seven clusters were significant and slightly decreased for both neutral and outlier datasets ( $r^2 = 0.793$  and  $r^2 = 0.824$ ,  $P < 0.001$ , respectively; [supplementary fig. S2, Supplementary Material online](#)).

Analyses of molecular variance (AMOVAs) for the two oceans (Indian and Pacific) using neutral loci showed a weak, but significant population structure ([table 2A](#);  $F_{SC} = 0.075$ ,  $P < 0.01$ ). More specifically, 22.8% of molecular variation occurred between the Indian and Pacific Oceans, whereas 5.8% of molecular variation occurred among populations within groups ( $n = 7$ ). AMOVA of both neutral and outlier datasets, which was based on seven clustering populations (see Materials and Methods), revealed a significant genetic divergence within and among all sampled populations ( $P < 0.01$ ; [table 2B](#)). For neutral loci, the majority of genetic variability (83.01%) was explained within populations and 13.9% among groups; however, for outlier loci, the genetic variability explained within populations (51.9%) was only slightly higher than the genetic variability explained among groups (46.7%). The patterns observed for both neutral and outlier datasets were supported by pairwise  $F_{ST}$  comparisons, which were significant at the group level ([table 2](#)). AMOVA results also yielded significant  $F_{SC}$  values in both neutral and outlier analyses ( $F_{SC} = 0.01711$  and  $0.04685$ , respectively), indicating significant genetic division among these seven groups ( $P < 0.01$ ).



**FIG. 3.**—IBD plots with Mantel correlograms for the relationship between genetic ( $F_{ST}$ ) and geographic (km) distances among *Penaeus monodon* populations ( $n = 16$ ) using (A) 9,930 neutral SNPs and (B) 663 outlier SNPs.

**TreeMix Analysis**

The generated maximum likelihood (ML) phylogenetic tree, which considered entire neutral loci dataset ( $n = 9,930$  SNPs), was rooted to Fiji (FJ) given the geographic separation of this population from all other sampled populations. Five migration events ( $m = 5$ ) provided the best explanation of sample covariance (fig. 4B) out of all migration events tested ( $m = 0–10$ ; supplementary fig. S3, Supplementary Material online). After adding five migration events (log-likelihood = 1,001.24), the likelihood was not further increased by increasing the number of migration events up to 10 (supplementary fig. S4, Supplementary Material online). Consistent with DAPC and ADMIXTURE analyses (with or without migration events included), TreeMix analysis indicated migration to Southeast

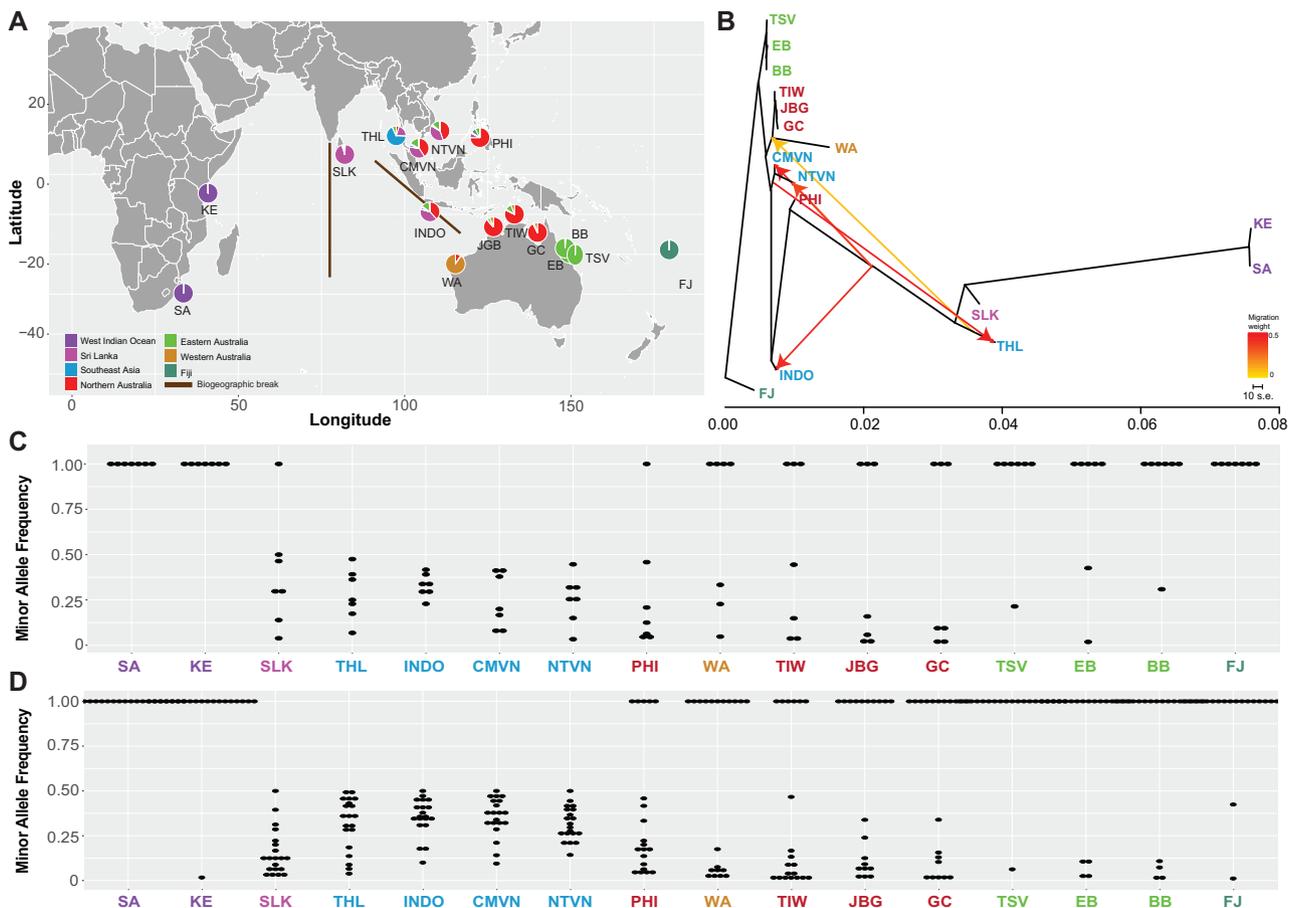
Asia from the WIO sites and a major phylogeographic split between WIO and Pacific Ocean populations (fig. 4B and supplementary fig. S3, Supplementary Material online).

There were three additional migration events included in the ML tree from the WIO to Southeast Asia with strong migration weights ( $>0.35$ ; fig. 4B). Within these regions, THL shared a branch with Indian Ocean populations (SLK, KE, and SA), which demonstrated the significant relationship between these populations and provided evidence for gene flow from the WIO to SLK and from THL to each of the Southeast Asia populations (NTVN, CMVN, and INDO). The fourth migration event suggested a slightly weighted migration from THL to the Northern and Western Australian populations (TIW, JBG, GC, and WA). The last migration event estimated a strongly

**Table 2**

AMOVA among *Penaeus monodon* Populations ( $n = 16$ ), among Regional Groups of Populations Identified by ADMIXTURE and DAPC Analyses, and among Individuals within Populations Using Neutral and Outlier SNP Datasets

Variance Partition		Degrees of Freedom	Sum of Squares	Variance Component	Variation (%)	Fixation Indexes	P Value
<b>(A) Between the Indian and Pacific Oceans</b>							
Neutral dataset (9,930 SNPs)	Among groups	1	17,168.99	69.54	22.75	FSC: 0.07529	<0.01
	Among populations within groups	14	19,378.48	17.78	5.82	FCT: 0.22748	<0.01
	Among individuals within populations	504	112,215.79	4.27	1.4		
	Within individuals	520	111,340.50	214.12	70.04	FIT: 0.29960	<0.01
<b>(B) Among seven regional groups</b>							
Neutral dataset (9,930 SNPs)	Among groups	6	32,347.27	35.74	13.86	FSC: 0.01711	<0.01
	Among populations within groups	9	4,200.20	3.80	1.47	FCT: 0.13857	<0.01
	Among individuals within populations	504	112,215.79	4.27	1.65		
	Within populations	520	111,340.50	214.12	83.01	FIT: 0.16986	<0.01
Outlier dataset (663 SNPs)	Among groups	6	15,306.66	17.93	46.68	FSC: 0.04685	<0.01
	Among populations within groups	9	727.41	0.96	2.50	FCT: 0.46681	<0.01
	Among individuals within populations	504	9,691.87	-0.29	-0.76		
	Within individuals	520	10,304.5	19.82	51.58	FIT: 0.48416	<0.01



**Fig. 4.**—(A) Map of biogeographic break and the proportion of each of the 16 *Penaeus monodon* populations assigned to seven clusters identified in the program ADMIXTURE using 9,963 neutral loci. The proportion of individuals assigned to each cluster per site depicted in colored pie charts. (B) Population relationships and migration edges of *Penaeus monodon* inferred by TreeMix for neutral SNP dataset at five migration events ( $m=5$ ). Fiji (FJ) was used as outgroup to root the tree. The migration events are colored according to their weight. (C) and (D) show the minor allele frequency distribution of seven and 20 temperature correlated putatively adaptive SNPs for each Indo-Pacific population, respectively. Population names are the same as defined in figure 1.

weighted migration of NTVN and CMVN to THL (fig. 4B). These results indicated either migration to THL from both the Indian Ocean (SA, KE, and SLK) and the Southeast Asia (CMVN and NTVN) populations or bidirectional gene flow within the THL population.

To confirm gene flow among *P. monodon* populations  $f_3$ -statistics were also calculated, which resulted in 212 significant  $f_3$  scores with Z scores ranging from  $-32.35$  to  $-0.07$  (supplementary table S4, Supplementary Material online). These significant  $f_3$ -statistics provided further support for the five migration events demonstrated by TreeMix analysis in that eight populations (BB, GC, JBG, TIW, CMVN, NTVN, INDO, and THL) each received gene flow from two other populations (supplementary table S4, Supplementary Material online). Sixteen  $f_3$ -statistics were significantly positive (Z score  $< -20$ ) when CMVN and INDO were admixed by SLK, PHI, and seven Oceania populations (GC, JBG, TIW, BB, EB, TSV, and FJ). Interestingly, all 16 significantly positive  $f_3$ -statistics included SLK as one of the two source populations.

### Identification of Putatively Adaptive SNPs

The transcriptome BLAST approach (see Materials and Methods) revealed that 20.5% of outlier loci ( $n=136$ ) matched coding contigs ( $n=1,408$  total) with an E-value  $< 10^{-5}$  (supplementary table S5, Supplementary Material online). Of these only 19.5% ( $n=27$ ) exhibited 100% pairwise identity to 62 transcriptome contigs (supplementary table S6, Supplementary Material online). Putative functional annotations were classified into 17 functional groups with eight (12.9%) annotated as “Predicted: uncharacterized protein”, 13 (20.9%) annotated as “N/A”, and 41 (66.2%) annotated as cellular and metabolic processes, pigmentation, and immune response genes (supplementary table S6, Supplementary Material online). Three of these putatively adaptive SNPs were missing from WIO populations and thus removed from subsequent correlation analysis. Pearson correlations revealed that the minor allele frequency (MAF) for four and three of these 24 putatively adaptive SNPs was

**Table 3**

Pearson Correlations for Putatively Adaptive SNPs that Exhibited a Significant Relationship between Minor Allele Frequency and Sea Surface Temperature Maximum or Minimum

SNP ID	SST_max		SST_min	
	<i>R</i>	<i>P</i> value	<i>R</i>	<i>P</i> value
(A) Candidate SNPs identified by the first analysis				
2019PM_4058	<b>-0.666</b>	<b>0.005</b>	0.005	0.986
2019PM_4727	<b>-0.667</b>	<b>0.005</b>	-0.162	0.550
2019PM_6523	<b>-0.629</b>	<b>0.009</b>	-0.248	0.354
2019PM_7307	<b>-0.541</b>	<b>0.030</b>	-0.325	0.220
2019PM_3144	-0.209	0.437	<b>-0.504</b>	<b>0.047</b>
2019PM_5152 <sup>a</sup>	-0.305	0.252	<b>-0.650</b>	<b>0.006</b>
2019PM_6229	-0.249	0.352	<b>-0.763</b>	<b>0.001</b>
(B) Candidate SNPs identified by the second analysis				
2019PM_665	<b>-0.580</b>	<b>0.020</b>	-0.080	0.770
2019PM_4130	<b>-0.600</b>	<b>0.010</b>	-0.060	0.840
2019PM_1552	<b>-0.639</b>	<b>0.008</b>	-0.237	0.378
2019PM_2811	<b>-0.655</b>	<b>0.006</b>	-0.251	0.349
2019PM_3848	<b>-0.670</b>	<b>0.005</b>	-0.222	0.409
2019PM_6807	<b>-0.596</b>	<b>0.015</b>	-0.296	0.266
2019PM_7282	<b>-0.588</b>	<b>0.017</b>	-0.177	0.513
2019PM_10077	<b>-0.633</b>	<b>0.008</b>	-0.098	0.719
2019PM_10517	<b>-0.564</b>	<b>0.023</b>	-0.143	0.596
2019PM_1233	-0.313	0.238	<b>-0.708</b>	<b>0.002</b>
2019PM_1287	-0.244	0.362	<b>-0.760</b>	<b>0.001</b>
2019PM_2814	-0.220	0.413	<b>-0.561</b>	<b>0.024</b>
2019PM_3748	-0.471	0.065	<b>-0.635</b>	<b>0.008</b>
2019PM_5152 <sup>a</sup>	-0.304	0.252	<b>-0.650</b>	<b>0.006</b>
2019PM_6872	-0.209	0.437	<b>-0.685</b>	<b>0.003</b>
2019PM_6979	-0.208	0.440	<b>-0.686</b>	<b>0.003</b>
2019PM_8233	-0.263	0.324	<b>-0.500</b>	<b>0.048</b>
2019PM_9353	-0.289	0.278	<b>-0.653</b>	<b>0.006</b>
2019PM_9861	-0.254	0.343	<b>0.540</b>	<b>0.031</b>
2019PM_10081	-0.315	0.234	<b>-0.632</b>	<b>0.009</b>

NOTE.—SST\_max, sea surface temperature maximum; SST\_min, sea surface temperature minimum. Significant values ( $P < 0.05$ ) are bolded.

<sup>a</sup>Overlapping SNP across both utilized approaches for putatively adaptive SNPs identification.

significantly correlated ( $P < 0.05$ ) with sea surface temperature maximum (SST\_max) and minimum (SST\_min), respectively (table 3A and supplementary fig. S5, Supplementary Material online). Of these, SNPs 2019PM\_4058 (C > G) and 2019PM\_6523 (T > G) matched transcriptome contigs annotated as “AChain Apocrustacyanin C1 Crystals” and “C-type lectin,” respectively (supplemental table S6, Supplementary Material online), whereas the other five temperature correlated putatively adaptive SNPs matched contigs that were annotated as “Predicted: uncharacterized protein”, “nucleic acid binding”, or “N/A” (supplemental table S6, Supplementary Material online).

The combined environmental association (EnA) and PD approaches (see Materials and Methods) revealed full RDA

model support for the role of sea surface temperature in *P. monodon* SNP genotype distribution ( $P < 0.001$ ;  $R^2 = 0.058$ ; adjusted  $R^2 = 0.055$ ). Based on these two significantly constrained axes the RDA model identified 1,256 putatively adaptive SNPs. LFMM analysis (FDR = 0.05) identified 424 putatively adaptive SNPs that were significantly associated with SST\_max or SST\_min of which 66 overlapped with RDA identified putatively adaptive SNPs. Of these, 35 overlapped with the PD determined outlier SNP dataset (see Materials and Methods; supplementary fig. S6, Supplementary Material online). Three of these were missing from WIO populations and thus removed from subsequent correlation analysis. Pearson correlations revealed that the MAF for 11 and 9 of these 35 putatively adaptive SNPs was significantly correlated ( $P < 0.05$ ) with SST\_max and SST\_min, respectively (table 3B and supplementary fig. S7, Supplementary Material online). Of these only 2019PM\_5152 exhibited 100% pairwise identity with two transcriptome contigs annotated as “transcriptional regulatory-like” and “N/A” (supplementary table S6, Supplementary Material online).

MAF distribution for the seven temperature correlated putatively adaptive SNPs identified by the transcriptome BLAST approach varied between 0.02 and 0.5 for all Southeast Asia populations (CMVN, NTVN, INDO, and THL) and between 0.2 and 1 for all other Indo-Pacific populations except for South Africa (SA), Kenya (KE), and Fiji (FJ), which all exhibited a MAF of 1 (fig. 4C). Similarly, MAF distribution for the 20 temperature correlated putatively adaptive SNPs identified by the combined EnA and PD approach varied between 0.02 and 0.5 for Southeast Asia (CMVN, NTVN, INDO, and THL) and Sri Lanka (SLK) populations and between 0.2 and 1 for all other Indo-Pacific populations except for South Africa (SA), which exhibited a MAF of 1 (fig. 4D).

## Discussion

The giant black tiger shrimp is an important aquaculture and fishery species in Indo-Pacific countries and comprehensive understanding of population genetic structure, genetic connectivity, and local adaptation is of value to assist with identifying: 1) source broodstocks, 2) wild stocks with different genetic backgrounds, and 3) appropriate fishery management strategies. Using whole-genome SNP sequencing we provide the most comprehensive analysis of *P. monodon* population genetic structure within the Indo-Pacific region to date, as well as new insights regarding gene flow from the Indian Ocean to the Pacific Ocean. The results from this study revealed a strong pattern of genetic structure within and among *P. monodon* Indo-Pacific populations that corresponded with broad geographical regions and clear genetic breaks among samples within regions, which in several cases is associated with geographical barriers, ocean surface currents, and environmental variables. Analyses based on neutral

SNPs ( $n = 9,930$ ) revealed: 1) five migration events that indicated gene flow to Southeast Asia from the WIO sites, 2) gene flow among *P. monodon* populations in the Pacific region, and 3) partial connectivity among populations native to both oceans. Identification of 26 putatively adaptive SNPs revealed that sea surface temperature may be important factors driving local adaptation of *P. monodon* across the Indo-Pacific. This study highlights the importance of conducting analyses on both neutral and outlier loci given the different requirements for *P. monodon* aquaculture and fishery management.

## Population Genetic Structure and Gene Flow

### Major Genetic Break within the Indo-Pacific

Using both neutral and outlier loci, the results from this study revealed a pattern of strong genetic structure for Indo-Pacific *P. monodon* that corresponded with broad geographical regions and clear genetic breaks among samples within geographically discrete regions. Neutral loci, which exhibited seven discernable genetic clusters, indicated a weak genetic structure across Southeast Asia and the Northwest Indian Ocean, as well as the separation of Fiji and WIO clusters from all other Indo-Pacific populations. Outlier loci, which exhibited six discernable genetic clusters, indicated a stronger genetic structure among Indo-Pacific populations as well as revealed a high degree of admixture (i.e., shared ancestry) among populations within each genetic cluster.

Genetic differentiation among Indo-Pacific *P. monodon* populations included in the study exhibited notable variation (supplementary table S3, Supplementary Material online). Outlier loci revealed an increase in genetic divergence across sample sites, which could be due to directional selection because identified genetic clusters correspond to locally adapted populations (see below). In addition to genetic differences, Mantel tests based on both neutral and outlier loci also showed a significant positive correlation between genetic and geographic distances, which indicates that geographic distance plays an important role in limiting gene flow among *P. monodon* populations. AMOVA results indicated a weak, but significant ( $P < 0.01$ ), population structure between the Indian and Pacific Oceans, as well as among the seven identified genetic clusters, which suggests that substantial gene flow may occur between the Indian and Pacific Oceans (table 2).

Our results indicated significant genetic differences among *P. monodon* populations. This finding is supported by similar previous observations for *P. monodon* (Benzie et al. 2002; You et al. 2008; Waqairatu et al. 2012), giant red shrimp *Aristaeomorpha foliacea* (Fernández et al. 2013), skunk clownfish *Amphiprion akallopisos* (Huyghe and Kochzius 2018), and holothurian *Stichopus chloronotus* (Pirog et al. 2019). Particularly, the three previous *P. monodon* studies demonstrated that the greatest genetic differentiation among populations existed between Southeast Asia and Australia

and between Southeast Asia and southeast Indian Ocean (Benzie et al. 2002; You et al. 2008), or between Indian and Pacific Ocean populations (Waqairatu et al. 2012).

There are many factors that contribute to the genetic differentiation of Indo-Pacific *P. monodon*, including environmental and ecological factors, biogeographic barriers between the Indian and Pacific Oceans, and life-history traits (Vogler et al. 2012; Hui et al. 2016; Amaral et al. 2017; Huyghe and Kochzius 2017). It is possible that biogeographic barriers between the Indian and Pacific Oceans could restrict gene flow and obstruct the physical movement of many tropical marine species between these ocean basins. One such biogeographic barrier is the Indonesian Sunda Shelf, which has been considered to play a role in (if not entirely responsible for) the genetic differences observed among Indo-Pacific *P. monodon* (Benzie et al. 2002; You et al. 2008; Waqairatu et al. 2012), as well as other Indo-Pacific marine species (Horne et al. 2008; Gaither et al. 2009; Simmonds et al. 2018; Wainwright et al. 2018). Of note is that the influence of this biogeographic barrier on gene flow remains during times of sea level rise following last glacial maximum (Randall 1998; Rocha et al. 2007; Crandall et al. 2019).

### Regional Population Structure within the Pacific Ocean

Significant population genetic structure was evident within Pacific Ocean *P. monodon*; neutral loci indicated five distinct genetic clusters are present, whereas outlier SNPs found evidence for four clusters (fig. 2). The main difference in genetic structure was that the Western Australian (WA) population was a discrete cluster (neutral SNPs) versus part of the Northern Australia and Philippines (NAP) cluster (outlier SNPs). Recently, Vu et al. (2020) demonstrated that *P. monodon* from a Western Australian population was genetically divergent from multiple Eastern and Northern Australia populations (i.e., discrete cluster) based on outlier loci determined using EnA analyses ( $n = 89$  SNPs). We believe that the difference between the present study and Vu et al. (2020) is most likely due to outlier determination being based on PD versus EnA analyses ( $n = 663$  vs. 89), respectively. Other non-SNP studies also reported *P. monodon* from Western Australia as being genetically distinct from other Australian *P. monodon* populations (Brooker et al. 2000; Benzie et al. 2002). Taken together, we suggest that the Western Australia *P. monodon* population should be managed as a unique genetic stock.

Within the Pacific Ocean, the Southeast Asian populations displayed a mixed pattern of differentiation and gene flow with those from Northern Australia (fig. 2). For example, ADMIXTURE analysis of both neutral and outlier loci revealed that populations in the Southeast Asia group shared ancestry with other clusters in the Pacific Ocean; Philippines was grouped with the Northern Australian group. TreeMix analysis also revealed low-level gene flow from Thailand to Northern-Western Australian populations (GC, JBG, TIW, and WA).

Additionally, many Southeast Asian countries have long histories of trading in live *P. monodon* broodstock for aquaculture which could lead to the high admixture between this region and hard to determine genetic relationships of the shrimp examined from this region.

Based on a whole SNP dataset ( $n = 10,593$ ) all Pacific Ocean populations, except Fiji, exhibited a high level of within-population genetic diversity, whereas Southeast Asia populations (e.g., Indonesia, Vietnam, and Thailand) exhibited the highest among-population genetic diversity (table 1). This elevated genetic diversity among Southeast Asia populations could be driven by the migration of genetically divergent Indian Ocean shrimp into Southeast Asia (fig. 4A and B) and/or environmental variables within Southeast Asia that promote genetic diversification through local adaptation to niche habitats. For example, the expansive mangroves found in Southeast Asia account for almost half of the total global above-ground biomass (Hutchison et al. 2014), which could lead to *P. monodon* genetic diversification through population expansion given that mangroves are known to provide ideal nursery habitat for *P. monodon* larvae (i.e., food and protection; Lee 2004; Sheaves et al. 2012). In addition genetic diversification being driven by migration events and/or potential local adaptation, Benzie et al. (2002) demonstrated the maintenance of large *P. monodon* populations within Southeast Asia as well as the evolution of genetic variants in peripheral populations, which were concluded to be the main factors driving the high level of observed genetic diversity. The observed heterogeneity and reduced genetic diversity within Fiji are most likely driven by geographic isolation, which is consistent with the significant differentiation of Fiji compared with all other Pacific Ocean populations. Waqairatu et al. (2012) demonstrated that Fijian *P. monodon* are genetically distinct from other Pacific Ocean populations and that this population was derived from a relatively small founder population that has remained isolated due to its geographic remoteness. The reduction in genetic diversity observed by Waqairatu et al. (2012) and this study could also be due to the introduction of *P. monodon* from the Philippines (Choy 1982) and Eastern Australian (Waqairatu and Pickering 2010) for aquaculture purposes because this could have led to the release or escape of less diverse aquaculture progeny into the wild and interbreeding with the small and diversity-limited founder population. However, our SNP dataset is unable to conclusively determine whether this lack of genetic diversity is due to: 1) a recent population bottleneck, 2) genetic drift from a small number of founder individuals, 3) influence of escaped or released aquaculture individuals, or 4) the various combinations of population bottleneck, genetic drift, and/or aquaculture influence. Future studies aimed at improving Fiji-based *P. monodon* aquaculture should focus on determining demographic processes and introgression patterns to further understand the root cause(s) and identify the most effective means of counteracting this low genetic diversity.

### Regional Population Structure within the Indian Ocean

At a regional scale, *P. monodon* collected at three locations in the Northwest and West Indian Ocean were clustered into two distinct groups, with Sri Lanka forming one cluster and WIO (Kenya and South Africa) comprising the other cluster (pairwise  $F_{ST} \geq 0.39$ ). Genetic divergence of Sri Lanka from other WIO populations has also been documented among marine populations of other species and explained by life history characteristics, environmental transitions, large geographical distances, geographical barriers, and/or ocean surface currents (Vogler et al. 2012; Hui et al. 2016; Amaral et al. 2017; Huyghe and Kochzius 2017). For example, Vogler et al. (2012) concluded that geographical barriers and ocean surface currents were drivers of the observed differences in the population genetic structure of Indian Ocean crown-of-thorns starfish *Acanthaster planci*, which is also a benthic species with a pelagic larvae stage lasting 3–4 weeks. In the present study, TreeMix analysis revealed that there was a very low level of gene flow between WIO populations and Sri Lanka (fig. 4B), as well as the lack of any significant negative  $f_3$ -statistics for three-population tests for Sri Lanka, South Africa, and Kenya (supplementary table S4, Supplementary Material online). Genetic divergence of *P. monodon* in the east, north, and west Indian ocean could have been driven by the loss of habitat and geographic fragmentation experienced during the last glacial maximum when the sea level dropped by approximately 120 m given the reported recolonization (by shelf areas or populations further north/east) of lost habitat following sea level rise following last glacial maximum (Forbes et al. 1999; Benzie et al. 2002; Silva et al. 2009). Moreover, the Indian Ocean has a complex current system with the South Equatorial Current flowing westward across the Indian Ocean and splitting into the Southeast and Northeast Madagascar Current, which can facilitate long-distance larval dispersal as well as generate migration barriers (Schott et al. 2009; Otwoma and Kochzius 2016; Huyghe and Kochzius 2018). Considering the offshore planktonic larval phase (approximately 20 days) of *P. monodon* (Motoh 1985), a stepping-stone model could explain larval dispersal throughout the entire Indian Ocean over several generations because the period of seasonal monsoons (northeast winds in December/January and southwest winds in July/August) corresponds with the peak *P. monodon* spawning seasons in the Indian Ocean (Sk et al. 2015; Kaka et al. 2019). Although we do not know the exact divergence time between the ancestral (WIO populations) and the Indo-Polynesian lineage (Sri Lanka, Southeast Asia, Australia, and Fiji), it could have occurred during the late Pleistocene given the lineage divergence timing of other species in this region (Vogler et al. 2012; Hui et al. 2016; Farhadi et al. 2017; Huyghe and Kochzius 2017). Future studies should further elucidate the degree of and potential reasons for genetic divergence among populations within the Indian Ocean by undertaking

extensive sampling within this region of *P. monodon*'s distribution.

At a local scale, genetic population structure analysis using both outlier and neutral loci also failed to determine the presence of genetic structure among WIO *P. monodon* populations. Strong gene flow between Kenya and South Africa could be due to the southward Mozambique Current, which is a branch that splits off from the Northeast Madagascar Current at the northern tip of Madagascar and creates a number of eddies on its way through the Mozambique Channel to the southward WIO (Schott et al. 2009; Huyghe and Kochzius 2018). Population genetic structure and gene flow analyses (DAPC, ADMIXTURE, and TreeMix) revealed that the WIO cluster was genetically divergent from all Indo-Polynesian populations (Sri Lanka, Southeast Asia, Australia, and Fiji). More specifically, significantly elevated and positive  $F_{IS}$  values were observed in WIO populations ( $F_{IS} = 0.80$  and  $0.78$  for Kenya and South Africa, respectively), which suggests heterozygosity deficiency (table 1). The significantly elevated  $F_{IS}$  values and reduced heterozygosity observed for these WIO populations could, therefore, be due to a genuine reduction in genome-wide heterozygosity (e.g., Wahlund effect; De Meeüs 2018) and/or the presence of null alleles (i.e., technical artifacts of the RAD-seq genotyping approaches; Davey et al. 2013) because selection pressure is not likely the driving factor (i.e., neutral loci exhibited elevated  $F_{IS}$  values; supplementary table S7, Supplementary Material online). A reanalysis of data that reduced the entire SNP dataset from 10,593 to 2,155 loci for all individuals ( $n = 520$ ) by retaining only the most informative SNPs ( $> 95\%$  concurrence in technical replicates with missing data per SNP  $< 2\%$ ) reduced the  $F_{IS}$  values of Kenya and South Africa populations to  $0.03$  and  $0.02$  (from  $0.80$  and  $0.78$ ; table 1), respectively, whereas  $H_O$  and  $H_E$  only differed by  $0.01$ – $0.02$  across all 16 populations (supplementary table S8, Supplementary Material online). Interestingly, this subset of SNPs had no effect on the diversity metrics of any other populations (e.g., significant Pearson correlations across all 16 populations were identified between  $H_E$ ,  $H_O$ , and  $A_R$  based on the whole 10,593 SNPs dataset as well as the 2,155 SNPs subset;  $R = 0.9$ – $1$ ,  $P < 0.00001$ ; data not shown), which confirms that, the observed elevation in  $F_{IS}$  values were was specific to WIO populations. Taken together, it is unclear whether the elevated  $F_{IS}$  values observed in WIO populations are the result of null alleles (i.e., artifacts) or/and the Wahlund effect (i.e., real); however, given that elevated  $F_{IS}$  values were exclusively observed for WIO populations, null alleles are more likely the driving factor than the Wahlund effect (Waples 2018). Regardless, both factors have been suggested as the two most common causes for the elevated  $F_{IS}$  values observed in Pacific whiteleg shrimp *Litopenaeus vannamei* (Valles-Jimenez et al. 2004), and kuruma shrimp *Penaeus japonicus* (Tsoi et al. 2007).

### Signatures of Local Adaptation

Vu et al. (2020) recently demonstrated the role of environmental factors (e.g., sea surface temperature) in shaping the genetic structure and local adaptation of Australian *P. monodon* populations. Based on the results of Vu et al. (2020) we hypothesized that *P. monodon* may also exhibit signatures of local adaptation to sea surface temperature on a larger scale (e.g., across entire Indo-Pacific region). Using two discrete analytical approaches, we identified 26 temperature correlated putatively adaptive SNPs that exhibited a significant correlation with SST\_max or SST\_min, which suggests that sea surface temperature could be contributing to genetic variation and/or local adaptation within and among Indo-Pacific *P. monodon* populations. Of note is that more or different putatively adaptive SNPs might have been identified by the transcriptome BLAST approach if different criteria were used to determine the outlier loci dataset; however, the use of both transcriptome BLAST (outlier SNPs only) and combined EnA and PD (all SNPs) approaches ensures that all putatively adaptive SNPs associated with temperature were identified. Our findings are consistent with the observed role of sea surface temperature on hierarchical genetic structure and genetic differences within and among populations of American lobster *Homarus americanus*, giant California sea cucumber *Parastichopus californicus*, sea scallop *Placopecten magellanicus*, and European green crab *Carcinus maenas* (Benestan et al. 2016; Jeffery et al. 2018; Van Wyngaarden et al. 2018; Xuereb et al. 2018).

One of the identified temperatures correlated putatively adaptive SNPs was 2019PM\_4058, which matched transcriptome contigs annotated as "AChain Apocrustacyanin C1 Crystals" (table 3 and supplementary table S6, Supplementary Material online). In *P. monodon* crustacyanin protein is known to be involved in temperature sensitive hypodermis pigmentation (Wade et al. 2014, 2015; Budd et al. 2017; Fan et al. 2021) and substrate color adaptation (Wade et al. 2012), which have important implications for *P. monodon*, *P. japonicus*, and *L. vannamei* aquaculture (Bernal Rodríguez et al. 2017). Another identified temperature correlated putatively adaptive SNP was 2019PM\_6523, which matched transcriptome contigs annotated as "C-type lectin" (table 3 and supplementary fig. S5 and table S6, Supplementary Material online). C-type lectins are known to be involved in the immune response to pathogen infection via the hemocyte prophenoloxidase system and might also enhance immune system recognition of certain bacteria and viruses in *L. vannamei* (Wei et al. 2012). The putatively adaptive SNPs that did not exhibit a significant correlation with sea surface temperature might still be useful candidate genes that correlate with other untested environmental factors (Milano et al. 2014). For example, 2019PM\_3043 matched one transcriptome contig annotated as one of the *P. monodon* chitinase genes (*PmChi-5*). *PmChi-5* may be involved in

molting, larval metamorphosis, immune defense against pathogen infections, and ammonia nitrate stress response in *P. monodon* (Zhou et al. 2018) as well as regulation of both humoral and cellular immune responses in *L. vannamei* (Niu et al. 2018). Although these putatively adaptive SNP annotations suggest potential roles in *P. monodon*, local adaptation no firm conclusions be made given the likelihood that all identified putatively adaptive SNPs are involved in a diversity of functions and/or pathways (Pavlidis et al. 2012).

### Implications for Aquaculture and Fishery Management

Within the Indo-Pacific, *P. monodon* is considered as one of the most important aquaculture species (Waqairatu et al. 2012). Genome-wide SNP genotyping (divided into neutral and outlier loci datasets) provides an essential tool for clarifying genetic population structure, which is of paramount importance for effective aquaculture and fishery management practices (Benestan 2019). The results of the present genome-wide SNP study may help inform *P. monodon* Indo-Pacific aquaculture and fishery management in several ways. Firstly, the presence of seven genetically distinct *P. monodon* stocks (Fiji, Eastern Australia, Northern Australia and Philippines, Western Australia, Southeast Asia, Sri Lanka, and the West Indian Ocean) could be considered as a single management unit at the regional level. However, more refined management of *P. monodon* wild stocks should be considered in Southeast Asia and Sri Lanka because these populations exhibited high admixture in both neutral and outlier loci analyses. In the case of Southeast Asia, where wild stocks encompass ocean territories of several countries, we recommend a multijurisdictional approach to management of these wild stocks as opposed to independent management by each country. Lastly, future research aimed at the development of selective breeding programs based on thermal adaptation, establishment of pathogen resistant, and/or disease-free aquaculture stocks should focus on outlier loci for candidate gene discovery (Lebeuf-Taylor et al. 2019).

## Materials and Methods

### Sample Collection and DNA Extraction

The giant black tiger shrimp *P. monodon* is widely distributed throughout the Indo-Pacific region, spanning from Fiji in the central Pacific, throughout Australia, Southeast Asia, and India, to the south western coast of the African continent. In order to study the genetic relationships across the entire Indo-Pacific range we collected wild adult of *P. monodon* ( $n = 532$ ) from 16 discrete Indo-Pacific locations between 2015 and 2019 (fig. 1 and supplementary table S1, Supplementary Material online). Pleopod tissue samples were excised from each wild adult *P. monodon* sample with scissors, preserved in ethanol and stored at  $-20^{\circ}\text{C}$  until DNA extraction.

Genomic DNA (gDNA) was extracted from each pleopod sample following a modified cetyl-trimethyl ammonium bromide protocol (Adamkewicz and Harasewych 1996; Vu et al. 2020). All gDNA extractions were subsequently cleaned using Sephadex G-50 (GE HLS 2007) and gDNA integrity was confirmed using 0.8% agarose gel containing GelGreen (ThermoFisher Scientific Pty Ltd, Australia). All gDNA samples were normalized to 50 ng/ $\mu\text{l}$  (20  $\mu\text{l}$  final volume) before submission to Diversity Arrays Technology (DArTseq; Canberra Australia) for library preparation and DArTseq<sup>TM</sup> sequencing.

### SNP Genotyping and Quality Control

All shrimp samples were sequenced and genotyped using DArTseq<sup>TM</sup> technology, as previously described (Kilian et al. 2012; Anderson et al. 2019; Vu et al. 2020). Briefly, DArTseq<sup>TM</sup> technology relies on a complexity reduction method in order to obtain genome sequences using next-generation sequencing technology. DArTseq<sup>TM</sup> was optimized for *P. monodon* by selecting the most appropriate complexity reduction method, which used a combination of *PstI* and *SphI* methylation-sensitive restriction enzymes. Only "mixed fragments" (*PstI-SphI*) were effectively amplified in 30 rounds of PCR using the following reaction conditions: 94  $^{\circ}\text{C}$  for 1 min, then 30 cycles of 94  $^{\circ}\text{C}$  for 20 s, 58  $^{\circ}\text{C}$  for 30 s, 72  $^{\circ}\text{C}$  for 45 s, and 72  $^{\circ}\text{C}$  for 7 min. After PCR amplification, equimolar amplicons from each sample were pooled and subjected to c-Bot (Illumina) bridge PCR before 77 cycles of single read sequencing on Illumina HiSeq2500.

Raw SNP sequence data were obtained from DArT after initial SNP calling, but without filtering by proprietary analysis pipelines (DArTsoft). SNP data were then initially filtered using custom *dartQC* python scripts (available from <https://github.com/esteinig/dartqc>). More specifically, SNPs matching the following stringent criteria were excluded from the dataset: 1) low read depth ( $-\text{read\_counts} < 7$ ), 2) average repeatability  $< 90\%$ , 3) call rate  $< 80\%$ , 4) duplicated SNPs base on similar sequence clusters ( $-\text{cluster} > 0.95$ ), and 5) MAF  $< 0.02$ . Furthermore, all remaining loci and individuals were tested for LD ( $r^2 > 0.2$ ) using *PLINK 1.9* (Purcell et al. 2007). For SNP pairs in LD, the SNP with the lowest call rate value was removed. Each SNP was also assessed for deviation from HWE at a population level using Arlequin version 3.5 (Excoffier et al. 2005) to run an exact test with 10,000 steps in the Markov Chain and 100,000 dememorizations. If a SNP significantly deviated from HWE ( $P < 0.0009$ ) in all populations, it was removed from further analyses. This filtering pipeline left  $n = 10,593$  high-quality SNP remaining for population genetic analyses.

### Population Genetic Diversity

Population genetic diversity statistics were calculated on all available SNPs ( $n = 10,593$ ). Observed heterozygosity ( $H_O$ ), expected heterozygosity ( $H_E$ ), and inbreeding coefficient ( $F_{IS}$ )

were computed using the *divBasic* function of R package *divRsity* version 1.9.90 (Keenan et al. 2013), with 1,000 bootstrap replicates. HP-RARE version 1.1 (Kalinowski 2005) was used to calculate allelic richness ( $A_R$ ), private allelic richness ( $A_{RA}$ ), and PPL using rarefaction to avoid sampling size bias. Average multilocus heterozygosity (Av. MLH) and sMLH were calculated for all individuals using R package *inbreedR* version 0.3.2 (Stoffel et al. 2016), whereas IR was calculated for each population using the R package *Rhh* version 1.0.1 (Alho et al. 2010).

### Determination of Outlier Loci

Loci under putative selection were identified using two independent PD approaches: BayeScan version 2.1 (Foll and Gaggiotti 2008; Foll 2012) and R package *PCAdapt* (Luu et al. 2017). The two approaches were based on underlying hypotheses that take into account different genome-wide signatures of local adaptation as both showed promise for accurately identifying outliers under a variety of complex demographic scenarios (Lotterhos and Whitlock 2015; Luu et al. 2017). BayeScan outlier detection was based on a logistic regression model that decomposed locus-population  $F_{ST}$  coefficients into population-specific and locus-specific components. More specifically, BayeScan analysis was conducted with 20 pilot runs of 5,000 iterations each followed by 100,000 iterations and 50,000 additional burn-in iterations followed by outlier identification using Bayescan 2.01 function *plot\_R.r* (Benjamini and Hochberg 1995). To decrease false positives due to multiple testing, we applied the false discovery rate (FDR) correction of  $q$  values at 0.01. Outlier detection using R package *PCAdapt* was based on principal components analysis (PCA) and identified loci that excessively related to population structure as being under putative selection. More specifically, *PCAdapt* was run using two principal components ( $K=2$ ) as they explained genetic structure of populations and a  $\text{min.maf} = 0.05$ ; candidate loci then were determined at FDR of 0.01 using R package *qvalue* (Storey et al. 2019).

### Population Genetic Structure: Neutral versus Outlier Divergence

Population differentiation and genetic structure were calculated separately for neutral ( $n = 9,930$ ) and outlier ( $n = 663$ ) SNP datasets.

Genetic differences among populations and individuals were estimated using R package *StAMPP* version 1.5.1 (Pembleton et al. 2013). Pairwise genetic differentiation values ( $F_{ST}$ ) and their significance between populations, were calculated according to Wright (1949) and Weir and Cockerham (1984).

Population genetic structure was determined using two approaches: 1) DAPC implemented in the R package *adegenet* version 2.1.1 (Jombart 2008), and 2) ADMIXTURE version

1.3.0 (Alexander et al. 2009). First, DAPC was used on individual genotypes in a multivariate analysis to determine the best number of genetic clusters ( $K$ ) by running ten iterations of the “*find.clusters*” function and the visualize BIC values for all ten iterations along with estimated SD for each  $K$  value ( $K$  values of 1–15) before subsequently identification the optimal number of retained principal components (PC) using the *optim.a.score* function. Second, ADMIXTURE estimated the ancestry and genetic structure of each individual based on ML using  $K$  values that ranged from 1 to 15 (as ancestral modes) to determine the most appropriate number of distinct genetic clusters. A substructure analysis for both neutral and outlier loci datasets to determine the admixture between five populations (CMVN, NTVN, INDO, THL, and SLK) was investigated using ADMIXTURE.

AMOVA was used to test genetic diversity structure between populations and geographic regions. AMOVA was run with ARLEQUIN version 3.5.2.2 (Excoffier and Lischer 2010) and statistical significance of each variance component was assessed using 10,000 permutations for each of the following hierarchical comparisons: 1) among groups, 2) among populations within groups, 3) among individuals within populations, and 4) within populations. Two independent AMOVAs were carried out based on oceans and regions. Firstly, in order to better understand the genetic structure of *P. monodon* between the Indian and Pacific Oceans, one hierarchical AMOVA was conducted for neutral loci based on two groups: 1) the Indian Ocean (SLK, KE, and SA) and 2) the Pacific Ocean (BB, EB, TSV, JBG, GC, TIW, WA, FJ, PHI, NTVN, CMVN, and THL). Secondly, we conducted independent analysis for both neutral and outlier loci datasets, all *P. monodon* individuals ( $n = 520$ ) were pooled into seven groups based on clustering of individuals by ADMIXTURE and DAPC analysis: 1) BB, EB, TSV; 2) GC, JBG, TIW, PHI; 3) WA; 4) FJ; 5) CMVN, NTVN, INDO, THL; 6) SLK; 7) KE, SA.

Isolation by distance was determined by discrete Mantel tests for each dataset (neutral and outlier) using the *mantel.randtest* function with 999 permutations in R package *ade4* (Dray and Dufour 2007). Four IBD analyses were estimated separately using both neutral and outlier loci for: 1) all 16 sampled sites and 2) seven sampled groups as described in AMOVA. Contemporary least-cost oceanographic distances between each pair of sampling sites were estimated using R package *marmap* (Pante and Simon-Bouhet 2013).

### TreeMix Analysis

To quantify historical connectivity across the Indo-Pacific region, a ML tree of the entire neutral dataset ( $n = 9,930$  SNPs) was constructed using TreeMix version 1.12 (Pickrell and Pritchard 2012). An ML tree without migration ( $m = 0$ ) was initially constructed and then migration events (i.e., edges) were added ( $m = 1–10$ ) to improve model fitness. Turn off sample size correction (*-noss*), bootstrap replicates (-

*bootstrap*), and *-root* flag (Fiji) were added to evaluate the confidence in the ML tree and the weight of migration events. The most likely number of migration events was selected based on the log-likelihood of the event and the plotted residuals, with each migration event analysis repeated ten times to evaluate consistency. Of note is that two root populations were assessed (Fiji and South Africa) with log-likelihood being more consistent when rooted to Fiji than South Africa. All trees and migration events were plotted in R using *plotting\_funcs.R* script within TreeMix version 1.12.

Lastly, to confirm the admixture of *P. monodon* among all populations, the *f*<sub>3</sub>-statistics were calculated using the three-pop approach (Reich et al. 2009) within TreeMix version 1.12. This approach calculated *f*<sub>3</sub>-statistics for all three possible populations in the form of (X; A, B) following a bifurcating tree. A significantly negative *f*<sub>3</sub>-statistic value means that X is the product of admixture between A and B (Reich et al. 2009).

### Identification of Putatively Adaptive SNPs

Each outlier SNP under putative selection was assessed against the assembled *P. monodon* transcriptome (Huerlimann et al. 2018) using Geneious Prime version 2019.1.3 (Kearse et al. 2012) with a maximum E-value of 10<sup>-5</sup>. Contigs within the assembled *P. monodon* transcriptome that exhibited 100% pairwise identity with each queried SNP sequence (69 bp for each allele) were compiled and reported along with contig annotations provided by Huerlimann et al. (2018).

To ensure comprehensiveness we undertook two discrete analytical approaches to identify putatively adaptive SNPs present within all 16 Indo-Pacific *P. monodon* populations: 1) transcriptome BLAST of entire outlier dataset and 2) combined EnA and PD analyses as per Vu et al. (2020). Putatively adaptive SNPs identified by both approaches were then independently assessed for correlation between MAF and location-specific sea surface temperature maximum (SST\_max) and minimum (SST\_min) using a linear regression (Pearson correlation method) following Dalongeville et al. (2018).

For the transcriptome BLAST approach each outlier SNP was assessed against the assembled *P. monodon* transcriptome (Huerlimann et al. 2018) using Geneious Prime version 2019.1.3 (Kearse et al. 2012) with a maximum E-value of 10<sup>-5</sup>. Contigs within the assembled *P. monodon* transcriptome that exhibited 100% pairwise identity with each queried SNP sequence (69 bp for each allele) were compiled and reported along with contig annotations provided by Huerlimann et al. (2018). Each identified putatively adaptive SNP was then assessed for correlation between MAF and location-specific sea surface temperature maximum (SST\_max) and minimum (SST\_min) using R package *ggpubr* (Kassambara A and Kassambara MA 2020).

For the Vu et al. (2020) approach the whole 10,593 SNPs dataset was subjected to two discrete EnA analyses to identify putatively adaptive loci associated with SST\_max or SST\_min: 1) redundancy analysis (RDA) in the R package *vegan* (Oksanen et al. 2013, 2018) and 2) latent factor mixed models (LFMM) in the R package *LEA* (Frichot and François 2015). LFMM analysis used seven latent factors because this matched the number of population clusters identified in DAPC and ADMIXTURE (fig. 2). Each identified putatively adaptive SNP was then: 1) assessed against the assembled *P. monodon* transcriptome (Huerlimann et al. 2018) using Geneious Prime version 2019.1.3 (Kearse et al. 2012) with 100% pairwise identity and E-value < 10<sup>-5</sup> cut-offs and 2) assessed for correlation between MAF and SST\_max and SST\_min using R package *ggpubr* (Kassambara A and Kassambara MA 2020).

SST\_max and SST\_min data (supplementary table S9, Supplementary Material online) for each Indo-Pacific sampling location were obtained from the Bio-ORACLE database (<http://www.bio-oracle.org/downloads-to-email.php>) (Tyberghein et al. 2012; Assis et al. 2018) as described in Vu et al. (2020).

### Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

### Acknowledgments

This work was supported by the Australian Research Council Industrial Transformation Research Hub for Advanced Prawn Breeding (IH130200013). N.T.T.V. was supported by Vietnam International Cooperation Department via Biotechnology of Aquaculture scholarship program and James Cook University School/Faculty Scholarship. For either providing or assisting with collection of shrimp tissue samples, we would like to thank to Monal M. Lal, Jerome G. Genodepa, Molatelo J. Madibana, Sriyan Chandrakeerthi, Saman Athauda, Sila Sukhavachana, Prapansak Srisapoom, Admad N.M. Muchtar, Purnomo Hadi, Samsul Bahrawi, Miera O. David, Melony J. Sellars, Dao Tan Hoc, Le Thi Kieu Thu, Le Van Truc, Nguyen Hong Lan, Tran Nhu Han, Nguyen Quoc Dat, Nguyen Viet Vuong, Dong Thanh Ha, Pham The Viet, Tran Hoang Ai Ly, Tran Van Huan, Le Vu Ni Na, and Nguyen Van Manh. We also would like to thank Research Institute for Aquaculture No.1, Dinhbang, Tuson, Bacninh, Vietnam for supporting N.T.T.V., David Jones for advice on data analyses, and Richard C. Edmunds for assistance with manuscript preparation. All authors declared no conflict of interest.

### Author Contributions

N.T.T.V. carried out tissue collections, laboratory work, participated in the investigation design and conceptualization,

performed all data analyses, and prepared the manuscript; J.L.G. and C.N.S.S. participated in laboratory work, data analyses, and manuscript editing; K.R.Z. and D.R.J. supervised project design and analyses and assisted with manuscript editing. All authors read and approved the final manuscript.

## Data Availability

All data analyzed during this study are included in this published article and its **Supplementary Data**. Raw datasets used for the current study are available from the corresponding author upon reasonable request. The *P. monodon* transcriptome used for gene ontology in this study is available from the National Center for Biotechnology Information database under the following accession numbers (<https://www.ncbi.nlm.nih.gov/nucleotide/GGLH000000000.1>): BioProject: PRJNA421400, BioSamples: SAMN08741487–SAMN08741521, SRA: SRP127068 (RR6868116–SRR6868172), and TRA: GGLH000000000 (Huerlimann et al. 2018).

## Literature Cited

- Adamkewicz SL, Harasewych M. 1996. Systematics and biogeography of the genus *Donax* (Bivalvia: Donacidae) in eastern North America. *Am Malacol Bull.* 13:97–104.
- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–1664.
- Alho JS, Välimäki K, MERILÄ J. 2010. Rhh: an R extension for estimating multilocus heterozygosity and heterozygosity–heterozygosity correlation. *Mol Ecol Resour.* 10(4):720–722.
- Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation genetics. *Nat Rev Genet.* 11(10):697–709.
- Amaral AR, Smith BD, Mansur RM, Brownell RL, Rosenbaum HC. 2017. Oceanographic drivers of population differentiation in Indo-Pacific bottlenose (*Tursiops aduncus*) and humpback (*Sousa* spp.) dolphins of the northern Bay of Bengal. *Conserv Genet.* 18(2):371–381.
- Anderson G, Hampton J, Smith N, Rico C. 2019. Indications of strong adaptive population genetic structure in albacore tuna (*Thunnus alalunga*) in the southwest and central Pacific Ocean. *Ecol Evol.* 9(18):10354–10364.
- Assis J, et al. 2018. Bio-ORACLE v2. 0: extending marine data layers for bioclimatic modelling. *Global Ecol Biogeogr.* 27(3):277–284.
- Batista PD, Janes JK, Boone CK, Murray BW, Sperling FA. 2016. Adaptive and neutral markers both show continent-wide population structure of mountain pine beetle (*Dendroctonus ponderosae*). *Ecol Evol.* 6(17):6292–6300.
- Benestan L. 2019. Population genomics applied to fishery management and conservation. In: Oleksiak MF, Rajora OP, editors. *Population genomics: marine organisms*. Cham: Springer International Publishing. p. 399–421.
- Benestan L, et al. 2016. Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Mol Ecol.* 25(20):5073–5092.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol.* 57(1):289–300.
- Benzie J, et al. 2002. Mitochondrial DNA variation in Indo-Pacific populations of the giant tiger prawn, *Penaeus monodon*. *Mol Ecol.* 11(12):2553–2569.
- Bernal Rodríguez C, García A, Ponce-Palafox J, Spanopoulos-Hernández M, Puga-López D. 2017. The color of marine shrimps and its role in the aquaculture. *Int J Aquac Fish Sci.* 3:62–65.
- Bernatchez L, et al. 2017. Harnessing the power of genomics to secure the future of seafood. *Trends Ecol Evol.* 32(9):665–680.
- Brooker A, Benzie J, Blair D, Versini J-J. 2000. Population structure of the giant tiger prawn *Penaeus monodon* in Australian waters, determined using microsatellite markers. *Mar Biol.* 136(1):149–157.
- Budd AM, Hinton TM, Tonks M, Cheers S, Wade NM. 2017. Rapid expansion of pigmentation genes in penaeid shrimp with absolute preservation of function. *J Exp Biol.* 220(Pt 22):4109–4118.
- Choy S. 1982. The biology of littoral penaeid prawns in Fiji waters with particular reference to *Penaeus* (*Melicertus*) *canaliculatus* Pérez-Farfante. Suva (Fiji): The University of the South Pacific. p. 161.
- Crandall ED, et al. 2019. The molecular biogeography of the Indo-Pacific: testing hypotheses with multispecies genetic patterns. *Global Ecol Biogeogr.* 28(7):943–960.
- Dalongeville A, Benestan L, Mouillot D, Lobreaux S, Manel S. 2018. Combining six genome scan methods to detect candidate genes to salinity in the Mediterranean striped red mullet (*Mullus surmuletus*). *BMC Genomics* 19(1):217.
- Davey JW, et al. 2013. Special features of RAD sequencing data: implications for genotyping. *Mol Ecol.* 22(11):3151–3164.
- De Meeüs T. 2018. Revisiting F IS, F ST, Wahlund effects, and null alleles. *J Hered.* 109(4):446–456.
- Dray S, Dufour A-B. 2007. The ade4 package: implementing the duality diagram for ecologists. *J Stat Soft.* 22(4):1–20.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online.* 1:117693430500100.
- Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 10(3):564–567.
- Fan S, et al. 2021. Molecular characterization and functional analysis of scavenger receptor class B from black tiger shrimp (*Penaeus monodon*). *Electron J Biotechnol.* 51:40–49.
- Farhadi A, et al. 2017. Mechanisms of peripheral phylogeographic divergence in the Indo-Pacific: lessons from the spiny lobster *Panulirus homarus*. *BMC Evol Biol.* 17(1):195.
- Fernández M, Heras S, Maltagliati F, Roldán M. 2013. Deep genetic divergence in giant red shrimp *Aristaeomorpha foliacea* (Risso, 1827) across a wide distributional range. *J Sea Res.* 76:146–153.
- 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(2):977–993.
- Forbes A, Demetriades N, Benzie J, Ballment E. 1999. Allozyme frequencies indicate little geographic variation among stocks of giant tiger prawn *Penaeus monodon* in the south-west Indian Ocean. *Afr J Mar Sci.* 21(1):271–277.
- Frichot E, François O. 2015. LEA: an R package for landscape and ecological association studies. *Methods Ecol Evol.* 6(8):925–929.
- Gagnaire PA, et al. 2015. Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evol Appl.* 8(8):769–786.
- Gaither MR, Toonen RJ, Robertson DR, Planes S, Bowen BW. 2009. Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira* and *Lutjanus fulvus*). *J Biogeogr.* 37(1):133–147.
- GE HLS. 2007. Illustra AutoSeq G-50 and AutoSeq 96 dye terminator removal. In: Data files 28–9175-28. Illustra AutoSeq G-50 documents. p.1–4. Available from: [https://gels.yilimart.com/Assets/Images/doc/file/17085501\\_DATAFILE\\_01.PDF](https://gels.yilimart.com/Assets/Images/doc/file/17085501_DATAFILE_01.PDF)

- Gjedrem T. 2012. Genetic improvement for the development of efficient global aquaculture: a personal opinion review. *Aquaculture* 344–349:12–22.
- Grant WS, Jasper J, Bekkevold D, Adkison M. 2017. Responsible genetic approach to stock restoration, sea ranching and stock enhancement of marine fishes and invertebrates. *Rev Fish Biol Fisheries*. 27(3):615–649.
- Gribble N, Atfield J, Dredge M, White D, Kistler S. 2003. Sustainable *Penaeus monodon* (Black Tiger prawn) populations for broodstock supply the department of primary industries. Brisbane (QLD): Queensland Department of Primary Industries.
- 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.). *Mol Ecol*. 16(15):3104–3118.
- Horne JB, van Herwerden L, Choat JH, Robertson DR. 2008. High population connectivity across the Indo-Pacific: congruent lack of phylogeographic structure in three reef fish congeners. *Mol Phylogenet Evol*. 49(2):629–638.
- Huerlimann R, et al. 2018. De novo assembly, characterization, functional annotation and expression patterns of the black tiger shrimp (*Penaeus monodon*) transcriptome. *Sci Rep*. 8(1):13553–13514.
- Hui M, et al. 2016. Comparative genetic population structure of three endangered giant clams (Cardiidae: *Tridacna* species) throughout the Indo-West Pacific: implications for divergence, connectivity and conservation. *J Mollus Stud*. 82(3):403–414.
- Hutchison J, Manica A, Swetnam R, Balmford A, Spalding M. 2014. Predicting global patterns in mangrove forest biomass. *Conserv Lett*. 7(3):233–240.
- Huyghe F, Kochzius M. 2017. Highly restricted gene flow between disjunct populations of the skunk clownfish (*Amphiprion akallopisos*) in the Indian Ocean. *Mar Ecol*. 38(1):e12357.
- Huyghe F, Kochzius M. 2018. Sea surface currents and geographic isolation shape the genetic population structure of a coral reef fish in the Indian Ocean. *PLoS One* 13(3):e0193825.
- Jeffery NW, et al. 2018. Genomewide evidence of environmentally mediated secondary contact of European green crab (*Carcinus maenas*) lineages in eastern North America. *Evol Appl*. 11(6):869–882.
- Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24(11):1403–1405.
- Kaka RM, Jung'a JO, Badamana M, Ruwa RK, Karisa HC. 2019. Morphometric length-weight relationships of wild penaeid shrimps in Malindi-Ungwana Bay: implications to aquaculture development in Kenya. *Egypt J Aquat Res*. 45(2):167–173.
- Kalinowski ST. 2005. hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes*. 5(1):187–189.
- Kassambara A, Kassambara MA. 2020. ggpubr: 'ggplot2' based publication ready plots. R package version 0.4.0. Available from: <https://CRAN.R-project.org/package=ggpubr>. Accessed April 2021.
- Kearse M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649.
- 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 Ecol Evol*. 4(8):782–788.
- Kilian A, et al. 2012. Diversity arrays technology: A generic genome profiling technology on open platforms. In: Pompanon F, Bonin A, editors. *Data production and analysis in population genomics. Methods in molecular biology (methods and protocols)*. Totowa (NJ): Humana Press. p. 67–89
- Kongkeo H. 2005–2020. Cultured Aquatic Species Information Programme: *Penaeus monodon*. FAO Fisheries Division [online]. Rome [Cited 12 September 2020].
- Lal MM, Southgate PC, Jerry DR, Bosserelle C, Zenger KR. 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(1):66.
- Lebeuf-Taylor E, McCloskey N, Bailey SF, Hinz A, Kassen R. 2019. The distribution of fitness effects among synonymous mutations in a gene under directional selection. *eLife* 8:e45952.
- Lee SY. 2004. Relationship between mangrove abundance and tropical prawn production: a re-evaluation. *Mar Biol*. 145(5):943–949.
- Lorenzen K, Beveridge M, Mangel M. 2012. Cultured fish: integrative biology and management of domestication and interactions with wild fish. *Biol Rev Camb Philos Soc*. 87(3):639–660.
- Lotterhos KE, Whitlock MC. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol Ecol*. 24(5):1031–1046.
- Luu K, Bazin E, Blum MG. 2017. pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Mol Ecol Resour*. 17(1):67–77.
- Matolweni LO, Balkwill K, McLellan T. 2000. Genetic diversity and gene flow in the morphologically variable, rare endemics *Begonia dregei* and *Begonia homonyma* (Begoniaceae). *Am J Bot*. 87(3):431–439.
- Milano I, et al. 2014. Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Mol Ecol*. 23(1):118–135.
- Motoh H. 1985. Biology and ecology of *Penaeus monodon*. In: Paper presented at the Proceedings of the First International Conference on the Culture of Penaeid Prawns/Shrimps; 1984 Dec 4–7; Iloilo City, Philippines: Aquaculture Department, Southeast Asian Fisheries Development Center. p. 27–36.
- Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D. 2009. Population genomics of marine fishes: identifying adaptive variation in space and time. *Mol Ecol*. 18(15):3128–3150.
- Niu S, et al. 2018. A chitinase from Pacific white shrimp *Litopenaeus vannamei* involved in immune regulation. *Dev Comp Immunol*. 85:161–169.
- Nowland SJ, Silva CN, Southgate PC, Strugnell JM. 2019. Mitochondrial and nuclear genetic analyses of the tropical black-lip rock oyster (*Saccostrea echinata*) reveals population subdivision and informs sustainable aquaculture development. *BMC Genomics* 20(1):711.
- Oksanen J, et al. 2013. Package 'vegan', Community Ecology Package. R package version 2.0-10. Available from: <https://CRAN.R-project.org/package=vegan>. Accessed February 2021.
- Oksanen J, et al. 2018. Package 'vegan': Community Ecology Package. R package version 2.5-2. Available from: <https://cran.r-project.org/web/packages/vegan/index.html>. Accessed February 2021.
- Oleksiak MF. 2019. Adaptation without boundaries: population genomics in marine systems. In: Rajora OP, editor. *Population genomics: concepts, approaches and applications*. Cham (Switzerland): Springer International Publishing. p. 587–612.
- Otwoma LM, Kochzius M. 2016. Genetic population structure of the coral reef sea star *Linckia laevigata* in the Western Indian Ocean and Indo-West Pacific. *PLoS One* 11(10):e0165552.
- Pante E, Simon-Bouhet B. 2013. marmap: a package for importing, plotting and analyzing bathymetric and topographic data in R. *PLoS One* 8(9):e73051.
- Pavlidis P, Jensen JD, Stephan W, Stamatakis A. 2012. A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. *Mol Biol Evol*. 29(10):3237–3248.

- Pembleton LW, Cogan NO, Forster JW. 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol Ecol Resour.* 13(5):946–952.
- Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8(11):e1002967.
- Pirog A, et al. 2019. High clonal propagation and low population connectivity in the holothurian *Stichopus chloronotus* from the Indo-Pacific. *Mar Biol.* 166(5):63.
- Purcell S, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81(3):559–575.
- Queiroga H, Blanton J. 2004. Interactions between behaviour and physical forcing in the control of horizontal transport of decapod crustacean larvae. *Adv Mar Biol.* 47:107–214.
- Randall JE. 1998. Zoogeography of shore fishes of the Indo-Pacific region. *Zool Stud.* 37(4):227–268.
- Reich D, Thangaraj K, Patterson N, Price AL, Singh L. 2009. Reconstructing Indian population history. *Nature* 461(7263):489–494.
- Rocha L, Craig M, Bowen B. 2007. Phylogeography and the conservation of coral reef fishes. *Coral Reefs.* 26(3):501–512.
- Schott FA, Xie SP, McCreary JP Jr. 2009. Indian Ocean circulation and climate variability. *Rev Geophys.* 47(1):RG1002.
- Segovia NI, Gallardo-Escárate C, Poulin E, Haye PA. 2017. Lineage divergence, local adaptation across a biogeographic break, and artificial transport, shape the genetic structure in the ascidian *Pyura chilensis*. *Sci Rep.* 7:44559.
- Sheaves M, Johnston R, Connolly RM, Baker R. 2012. Importance of estuarine mangroves to juvenile banana prawns. *Estuarine Coastal Shelf Sci.* 114:208–219.
- Silva IC, Mesquita N, Paula J. 2009. Genetic and morphological differentiation of the mangrove crab *Perisesarma guttatum* (Brachyura: Sesamididae) along an East African latitudinal gradient. *Biol J Linn Soc.* 99(1):28–46.
- Simmonds SE, et al. 2018. Evidence of host-associated divergence from coral-eating snails (genus *Coralliophila*) in the Coral Triangle. *Coral Reefs* 37(2):355–371.
- Sk NU, Ghosh S, Maity J. 2015. Reproductive biology, maturation size and sex ratio of black tiger shrimp (*Penaeus monodon* Fabricius, 1798) from fishing grounds of Digha coast, West Bengal, India. *Int J Aquat Biol.* 3(6):372.
- Stoffel MA, et al. 2016. inbreedR: an R package for the analysis of inbreeding based on genetic markers. *Methods Ecol Evol.* 7(11):1331–1339.
- Storey JD, Bass AJ, Dabney A, Robinson D. 2019. qvalue: Q-value estimation for false discovery rate control. Available from: <http://github.com/jdstorey/qvalue>. Accessed November 2019.
- Tsoi K, Chan T, Chu K. 2007. Molecular population structure of the kuruma shrimp *Penaeus japonicus* species complex in western Pacific. *Mar Biol.* 150(6):1345–1364.
- Tyberghein L, et al. 2012. Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Glob Ecol Biogeogr.* 21(2):272–281.
- Valles-Jimenez R, Cruz P, Perez-Enriquez R. 2004. Population genetic structure of Pacific white shrimp (*Litopenaeus vannamei*) from Mexico to Panama: microsatellite DNA variation. *Mar Biotechnol.* 6(5):475–484.
- Van Wyngaarden M, et al. 2018. Oceanographic variation influences spatial genomic structure in the sea scallop, *Placopecten magellanicus*. *Ecol Evol.* 8(5):2824–2841.
- Vogler C, et al. 2012. Phylogeography of the crown-of-thorns starfish in the Indian Ocean. *PLoS One* 7(8):e43499.
- Vu NTT, et al. 2020. Fine-scale population structure and evidence for local adaptation in Australian giant black tiger shrimp (*Penaeus monodon*) using SNP analysis. *BMC Genomics* 21(1):669.
- Wade NM, Budd A, Irvin S, Glencross BD. 2015. The combined effects of diet, environment and genetics on pigmentation in the giant tiger prawn, *Penaeus monodon*. *Aquaculture* 449:78–86.
- Wade NM, et al. 2012. Mechanisms of colour adaptation in the prawn *Penaeus monodon*. *J Exp Biol.* 215(2):343–350.
- Wade NM, et al. 2014. Quantitative methods to measure pigmentation variation in farmed Giant Tiger Prawns, *Penaeus monodon*, and the effects of different harvest methods on cooked colour. *Aquaculture* 433:513–519.
- Wainwright BJ, Arlyza IS, Karl SA. 2018. Population genetics of the collector urchin, *Tripneustes gratilla*, in the Indonesian archipelago. *Mar Ecol.* 39(6):e12530.
- Waples RS. 2018. Null Alleles and F IS × F ST Correlations. *J Hered.* 109(4):457–461.
- Waqairatu S, Pickering T. 2010. Confirmatory testing of the viral status of *Penaeus monodon* (Black Tiger shrimp) populations in the Fiji Islands. Retrieved from The University of the South Pacific.
- Waqairatu SS, et al. 2012. Genetic analysis of Black Tiger shrimp (*Penaeus monodon*) across its natural distribution range reveals more recent colonization of Fiji and other South Pacific islands. *Ecol Evol.* 2(8):2057–2071.
- Wei X, et al. 2012. Two C-type lectins from shrimp *Litopenaeus vannamei* that might be involved in immune response against bacteria and virus. *Fish Shellfish Immunol.* 32(1):132–140.
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358–1370.
- Williams SE, Hoffman EA. 2009. Minimizing genetic adaptation in captive breeding programs: a review. *Biol Conserv.* 142(11):2388–2400.
- Woodings LN, et al. 2018. Outlier SNPs detect weak regional structure against a background of genetic homogeneity in the Eastern Rock Lobster, *Sagmariasus verreauxi*. *Mar Biol.* 165(12):185.
- Wright S. 1949. The genetical structure of populations. *Ann Eugenics.* 15(1):323–354.
- Xuereb A, Kimber CM, Curtis JM, Bernatchez L, Fortin MJ. 2018. Putatively adaptive genetic variation in the giant California sea cucumber (*Parastichopus californicus*) as revealed by environmental association analysis of restriction-site associated DNA sequencing data. *Mol Ecol.* 27(24):5035–5048.
- You EM, et al. 2008. Microsatellite and mitochondrial haplotype diversity reveals population differentiation in the tiger shrimp (*Penaeus monodon*) in the Indo-Pacific region. *Anim Genet.* 39(3):267–277.
- Zhou F, et al. 2018. Characterization and expression analysis of a chitinase gene (*PmChi-5*) from black tiger shrimp (*Penaeus monodon*) under pathogens infection and ambient ammonia-N stress. *Fish Shellfish Immunol.* 72:117–123.

Associate editor: Gwenaél Piganeau