Population genomics of the giant black tiger shrimp *Penaeus monodon* to understand wild fishery and aquaculture production

Thesis submitted by

Nga Thi Thanh Vu (MSc)

10 November 2021

In fulfilment of the requirements for Doctor of Philosophy

In Agriculture, Environmental and Related Studies

At College of Science and Engineering

James Cook University

Townsville, QLD Australia

Declaration

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 Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is provided.

_____

Nga T. T. Vu

10 November 2021

Acknowledgements

I never imagined that undertaking a PhD would be part of my life journey! But, alas, the unexpected happened and I am so very grateful for everyone who helped guide, support, fund, motivate, and challenge me throughout this life-changing experience.

Foremost, I would like to thank my primary supervisor Professor Dean Jerry for believing in me and encouraging me to pursue a PhD when I finished my Master of Science at James Cook University (JCU) in 2012. Your immense knowledge, patience, and motivation inspired and guided me through the various challenges that arose during my PhD journey. Thank you so much for everything!

I would also like to express my sincere gratitude to my second supervisor, Professor Kyall Zenger, for his support, guidance, and mentorship throughout my candidature. When I arrived at JCU to commence my candidature I did not have any experience in genetic analyses, so I am very grateful to you, Kyall, for all your patience, guidance, and knowledge. I could not have done it without you!

My special thanks to Monal Lal, Jerome Genodepa, Molatelo Madibana, Sriyan Chandrakeerthi, Saman Athauda, Sila Sukhavachana, Prapansak Srisapoome, Admad N.M. Muchtar, Purnomo Hadi, Samsul Bahrawi, Miera O. David, Dao Tan Hoc, Le Van Truc, Nguyen Hong Lan, Tran Nhu Han, Nguyen Quoc Dat, Tran Hoang Ai Ly, Tran Van Huan, Le Vu Ni Na, and Nguyen Van Manh for either providing or assisting with collection of shrimp tissue samples throughout the Indo-Pacific regions. You all contributed to making my PhD possible, for which I am forever grateful. I wish you all the best and look forward to future collaborations.

I would like to thank all manuscript co-authors (Jarrod Guppy, Catarina Silva, Melony Sellars, Shannon Kjeldsen, Dave Jones, and Richard Edmunds) for their time, support, and knowledge. I would also like to thank Julie Goldsbury and Carolyn Smith-Keune for mentorship in the Molecular Ecology and Evolution Laboratory.

I have also received amazing support from my friends and colleagues in the Aquaculture Genetics Group, Molecular Ecology and Evolution Laboratory, and Vietnamese students at JCU. I would like to thank Bill Chen, Sally Lau, Caroline Candebet, Samuel Payet,

Maria Nayfa, Eike Steinig, Brooke Whitelaw, Roger Huerlimann, Sandra Infante-Villamil, Cecile Massault, Sandra Hughes, Truong–Yen, Kien–Quyen, Thu–Huan, Hung–Phuong, Chung–Diem, The Anh–Dao, Luan, Hang, Thanh, Kieu, and many others for your advice, support, and friendship.

I would like to thank Professor Phan Thi Van and all of my colleagues at the Research Institute for Aquaculture No 1 in Vietnam for encouraging and supporting me. I really appreciate for all your help!

I would like to thank Associated Professor Pham Quoc Hung for introducing me to science during my undergraduate studies at Nha Trang University and being my mentor ever since. Thank you for your endless support and motivation!

I would like to thank my dear friends in Vietnam and Australia who provided unquestioning support as my "mental health team" (The Anh, Khoi, Phuong, Thuy, Thoa, Lan, Ha, Ni Na, Chinh). I really appreciate for all you did for me and am forever grateful to you all!

I would like to thank all my international and domestic Townsville friends. If wasn't for you Marina, Caiua, Jason, Scott, Fei, Lori, Susan, Anna, and Steve my PhD experience and life in Townsville would have not attained the magnitude of madness that I did. I love and appreciate all the Ship's Cats!

I would like to thank my family in Vietnam (Dinh, Van, Nhat, Thuong, Duc, Dung, Ngoc, An, Dat, and Diep), and my partner's family in USA (Daryn, Barbara, Phil, and Lynda Edmunds) for their love and endless support throughout my PhD candidature. You all played a central role in helping me to achieve this life goal. For this I am forever grateful!

I would like to thank my dog, Monkey, for laying by my side all the time that I worked from home and sharing his energy and love of life with me. You are truly one of a kind and I feel so lucky that we found you (at the last minute)!

Last but not least I would like to thank my life partner, Richard, for always believing in me more than I believed in myself and helping, supporting, and encouraging me with his unconditional love and knowledge about science and mental health. With all my love I thank you so much for everything as I honestly do not feel that I could have completed this journey without you!

"THIS IS THE WAY"

- The Mandalorian Creed

Statement of the contribution of others

My PhD candidate was supported by Vietnam International Cooperation Department via the Biotechnology of Aquaculture scholarship program which provided funding for living costs in Australia as well as a half of the enrolment and tuition fees at JCU Townsville. Further living expenses were supported by ARC-Topup (JCU) with a stipend of $5,273AUD per year for 2.5 years. All primary research costs were supported by the Australian Research Council Industrial Transformation Research Hub for Advanced Prawn Breeding (IH130200013), which was led by Professor Dean Jerry at JCU Townsville. Supplementary research costs were supported by a grant award of $500AUD from the College of Science and Engineering (CSE) Noncompetitive Research Training Grant round in 2017 as well as $3,000AUD grant awards from the CSE Competitive Research Training Grant rounds in 2018 and 2020.

I was involved in the conceptualisation and experimental design of all work presented in this thesis, including tissue sample collections and genomic DNA extractions. I was also primarily responsible for data management, integrity, analyses, and interpretation. This included SNP/PAV genotyping and filtering (Chapters 2 – 5), population genomic analyses (Chapter 2 – 3), phylogenomic analyses (Chapter 4), and commercially-viable low-density SNP panel identification (Chapter 5). I was the sole author for all written work presented in the six chapters that comprise this thesis and was lead author on all peer-reviewed journal article manuscripts that were derived from this body of work, both published and in preparation. Specific co-author contributions for this thesis are outlined by Chapter in the Table below.

Thesis chapter	Publication output on which the chapter is based	Nature and extend of contribution of co-contributors
2	Vu, N. T. T., Zenger, K. R., Guppy, J. L., Sellars, M. J., Silva, C. N. S., Kjeldsen, S. R., & Jerry, D. R. (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.	**Vu, N. T. T.**[a,b]: participated in study design and conceptualisation, performed all data analyses, and prepared the manuscript . **Guppy, J. L.**[a,b], **Sellars, M. J.**[a,c,*], **Silva, C. N. S.**[b], and **Kjeldsen, S. R.**[b]: facilitated the collection of wild samples, participated in laboratory work, data analyses and manuscript editing. **Zenger, K. R.**[a,b] and **Jerry, D. R.**[a,b,d]: supervised project design and analyses as well as edited manuscript.

3	Vu, N. T. T., Zenger, K. R., Silva, C. N. S., Guppy, J. L., & Jerry, D. R. (2021). Population structure, genetic connectivity, and signatures of local adaptation of the giant black tiger shrimp (*Penaeus monodon*) throughout the Indo–Pacific Region. ***Genome biology and evolution***. doi:10.1093/gbe/evab214	**Vu, N. T. T.**[a,b]: carried out tissue collections, laboratory work, participated in study design and conceptualisation, performed all data analyses, and prepared the manuscript. **Silva, C. N. S**.[b] and **Guppy, J. L**.[a,b]: participated in laboratory work, data analyses, and manuscript editing. **Zenger, K. R**.[a,b] and **Jerry, D. R**.[a,b,d]: supervised project design and analyses and assisted with manuscript editing.
4	Prepared as thesis chapter only	**Vu, N. T. T.**[a,b]: participated in study design and conceptualisation, performed all data analyses, and prepared the manuscript. **Zenger, K. R**.[a,b,] **Jerry, D. R**.[a,b,d]: supervised project design and analyses and assisted with manuscript editing.
5	Vu, N. T. T., Zenger, K. R., Jones, B. J. [a,b], Edmunds, R. C., & Jerry, D. R. (*in final prep*). Identification of suitable SNP markers for rapid and accurate diversity, provenance, and pedigree analyses of Indo-Pacific giant black tiger shrimp (*Penaeus monodon*). In preparation for ***Aquaculture***.	**Vu, N. T. T.**[a,b]: participated in study design and conceptualisation, performed all data analyses, and prepared the manuscript. **Jones, B. J.**[a,b] and **Edmunds, R. C**.[e**]: participated in data analyses and manuscript editing. **Zenger, K. R**.[a,b] and **Jerry, D. R**.[a,b,d]: supervised project design and analyses and assisted with manuscript editing.

Author affiliations

[a] ARC Industrial Transformation Research Hub for Advanced Prawn Breeding, College of Science and Engineering, James Cook University, Townsville, 4811, Australia.

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

[c] CSIRO Agriculture & Food, Integrated Sustainable Aquaculture Production Program, Queensland Bioscience Precinct, St Lucia, 4067, Australia.

[d] Tropical Futures Institute, James Cook University, Singapore, Singapore.

[e] Animal Genetics and Nutrition, Veterinary Sciences Discipline, College of Public Health, Medical and Veterinary Sciences, Division of Tropical Health and Medicine, James Cook University, Townsville, Queensland 4811, Australia

* Current address: Genics Pty Ltd, Level 5, Gehrmann Building. 60 Research Road, St Lucia QLD Australia 4067.

** Current address: Australian Institute of Marine Science, PMB 3, Townsville QLD 4810 Australia

Thesis abstract

The giant black tiger shrimp (*Penaeus monodon*) is native to the Indo-Pacific and is the second most farmed penaeid shrimp species globally and is one of the major contributors to global food security. Despite the commercial importance of *P. monodon,* the population structure and genetic connectivity within and among Indo-Pacific populations remains unclear (i.e., inconsistencies across previous studies) while information about local adaptation among Indo-Pacific populations remains limited (i.e., few informative studies to date). Consequently, comprehensive population genomics information will help to improve the sustainable development of commercial aquaculture practices (e.g., selective breeding) and wild fishery management strategies (e.g., persistence of healthy wild populations). To address important knowledge gaps, I undertook three interdependent investigations (Chapters 2 – 4) that collectively aimed to advance the current understanding about Australian (fine-scale) and Indo-Pacific (broad-scale) *P. monodon* population structure, genetic connectivity, local adaptation, and evolutionary relationships. To address current commercial needs, I undertook a fourth interdependent investigation (Chapter 5) that aimed to identify accurate and cost-effective population genomics analysis tools for diversity, provenance, and pedigree determination.

The first investigation (Chapter 2) addressed the knowledge gap about Australian *P. monodon* population genomics by generating a high-quality genome-wide single nucleotide polymorphism (SNP) dataset (n = 10,624 total) for individuals (n = 278 total) collected from geographically disparate populations (n = 7). Then, for the first time using a high-quality genome-wide SNP dataset, I determined the fine-scale population structure, genetic diversity, and putative local adaptation within and among Australian *P. monodon* populations. This approach revealed significant population structure and differentiation among all Australian populations (mean F_{ST} = 0.001 – 0.107; $p < 0.05$). Clear population structure with similar spatial patterns were observed in both neutral and outlier marker datasets with three genetically distinct groups identified (Eastern Australia, Northern Australia, and Western Australia). Eighty-nine putative outlier SNPs were identified as being potentially associated with environmental variables using both population differentiation (BayeScan and *PCAdapt*) and environmental association (redundancy analysis and latent factor mixed model) analysis approaches. Redundancy, partial redundancy, and multiple regression on distance matrices analyses using these 89 outlier SNPs revealed that both geographical distance and

environmental factors interact to generate the structure observed across Australian *P. monodon* populations. Interestingly, only 11 outlier SNPs matched *P. monodon* transcriptome contigs (Huerlimann *et al*., 2018) with subsequent protein translation of eight outlier SNPs demonstrating more non-synonymous changes than synonymous changes across all six reading frames. This study provides the most comprehensive and first SNP-based analysis of fine-scale Australian *P. monodon* population genomics to date, which can help aquaculture practices and fishery management strategies through improved stock identification, genetic diversity estimations, local adaptation determination, and selective breeding programs.

The second investigation (Chapter 3) addressed the knowledge gap about Indo-Pacific *P. monodon* population genomics by generating a high-quality genome-wide SNP dataset (n = 10,593 total) for each individual (n = 532 total) collected from geographically discrete populations (n = 16). Then, for the first time using a high-quality genome-wide SNP dataset, I determined the broad-scale population structure, genetic connectivity, genetic diversity, and local adaptation signatures within and among Indo-Pacific *P. monodon* populations. This approach revealed that genetic diversity levels were highest for Southeast Asian (SEA) populations and 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 among Indo-Pacific *P. monodon* populations that corresponded with broad geographical regions and clear genetic breaks among samples within regions. Neutral loci revealed seven genetic clusters and the separation of Fiji (FJ) 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 SEA from WIO with partial connectivity to populations in both oceans. I also identified 26 putatively adaptive SNPs from the outlier dataset that exhibited significant Pearson correlation ($p < 0.05$) between minor allele frequency and maximum or minimum sea surface temperature. Matched transcriptome contig annotations suggest that these putatively adaptive SNPs could be located within genes that are involvement in cellular and metabolic processes, pigmentation, immune response, and currently unknown functions. This study provides the most comprehensive and first SNP-based analysis of broad-scale Indo-Pacific *P. monodon* population genomics to date, which can help aquaculture practices and fishery management strategies through improved stock identification, genetic connectivity establishment, genetic diversity estimation, local adaptation determination, and selective breeding programs.

The third investigation (Chapter 4) addressed the knowledge gap about Indo-Pacific *P. monodon* phylogeography and phylogenomic relationships with sister taxa by identifying genome-wide SNP and presence-absence variant (PAV) marker subsets (n = 4,496 and n = 7,054), respectively. Then, using genome-wide SNP and PAV subsets in *P. monodon*, phylogenomic reconstructions were generated using neighbour-joining, maximum-likelihood, and Bayesian inference approaches, which all indicated a monophyletic structure for Indo-Pacific *P. monodon* and confirmed the evolutionary relationships between Indo-Pacific *P. monodon* and six sister taxa that were previously reported by non-SNP studies. More specifically, phylogeographic structures revealed that Indo-Pacific *P. monodon* appears to be divided into two lineages: 1) Indo-Polynesian (containing Sri Lanka, SEA, Australia, and FJ) and 2) WIO (containing Kenya and South Africa). The phylogenomic reconstructions collectively support the broad-scale population genetic structure (Chapter 3) conclusion that the WIO lineage is ancestral to the Indo-Polynesian lineage and that Sri Lanka (SLK) serves as a transition zone between these geographically distinct regions. This study provides comprehensive SNP- and PAV-based phylogenomic analyses of Indo-Pacific *P. monodon* and sister taxa, which can help aquaculture practices and fishery management strategies through improved stock identification, species determination, and genetic connectivity monitoring.

The fourth and final investigation (Chapter 5) addressed the commercial need for accurate and cost-effective population genomics analysis tools by identifying five low-density SNP panels from the broad-scale Indo-Pacific SNP dataset: 1) global diversity SNP panel (n = 2,155), 2) global provenance SNP panel (n = 1,200), 3) Pacific Ocean (PO) pedigree SNP panel (n = 217), 4) SLK pedigree SNP panel (n = 226), and 5) WIO pedigree SNP panel (n = 220). Multiple genetic diversity analyses demonstrated that the diversity SNP panel was able to estimate the genetic diversity within and among all Indo-Pacific *P. monodon* populations as accurately as the entire broad-scale SNP dataset (n = 10,593). Both DAPC and *NetView* analyses demonstrated that the provenance SNP panel was able to trace *P. monodon* collected from the same region back to their original basin and genetic cluster while Monte-Carlo tests demonstrated that individuals could be accurately assigned back to their original basin, region, and basin-specific sampling location. *In silico* analyses demonstrated that the PO, SLK, and WIO pedigree SNP panels were able to accurately assign theoretical progenies back to their parent-pair with 100% accuracy using 30 - 40 SNP markers. This study provides a globally effective and accurate low-density SNP panels for the Indo-Pacific *P. monodon* industry, which can help aquaculture practices and fishery management strategies through improved diversity,

provenance, and pedigree determination as well as facilitation of global data consistency, global database development, and global analysis standards establishment.

Using high-quality genome-wide SNP data, these four studies provide the most comprehensive evaluation of Australian (fine-scale) and Indo-Pacific (broad-scale) *P. monodon* population genomics to date as well as identify a suite of accurate and cost-effective genetic tools for global industry use. Taken together, these four interdependent studies provide new population genomics insights and resources that contribute towards closing the current knowledge gaps about key aspects of *P. monodon* aquaculture and fishery management practices.

Table of Contents

Frontispiece i

Declaration ii

Acknowledgements iii

Statement of the contribution of others vii

Thesis abstract viiii

Table of Contents xiii

List of Figures xviii

List of Tables xxii

List of Appendices xxiiii

Abbreviations xxv

Chapter 1. General introduction 1

1.1 Food security and aquaculture 1

1.2 Gene flow and local adaptation in marine species 2

1.2.1 Gene flow in marine species 3

1.2.2 Local adaptation in marine species 4

1.3 Application of genetic markers and techniques to aquaculture and fishery management 5

1.4 Population genetics and knowledge gaps of Black tiger shrimp aquaculture and fishery management 8

1.5 Thesis overview 10

1.6 Publication arising from this thesis 10

Chapter 2. Fine-scale population structure and evidence for local adaptation in Australia giant black tiger shrimp (*Penaeus monodon*) using SNP analysis 12

2.1 Introduction 12

2.2 Methods 14

2.2.1 SNP collection and DNA extraction 14

2.2.2 Genotyping and quality control 16

2.2.3 Genetic diversity, population differentiation and structure 17
2.2.4 Environmental data collection and processing 18
2.2.5 Detection of loci under selection 19
2.2.6 Environmental factors association with genetic variation 20
2.2.7 Gene ontology 21
2.3 Results 21
2.3.1 Genotyping and SNP quality control 21
2.3.2 Population genetic diversity 22
2.3.3 Population differentiation and genetic structure 23
2.3.4 Detection of outlier loci 27
2.3.5 Genetic structure based on outlier and neutral markers 28
2.3.6 Environmental factors association with genetic variation 29
2.3.7 Gene ontology 31
2.4 Discussion 32
2.4.1 Population genetic diversity 33
2.4.2 Population differentiation and genetic structure 34
2.4.3 Evidence for local adaptation 35
2.4.4 Environmental variables contributing to genetic structure 36
2.4.5 Highlights from Gene Ontology 37
2.4.6 Implications for aquaculture management and future research directions 38
2.5 Conclusion 39
Chapter 3. Population structure, genetic connectivity, and signatures of local adaptation of the giant black tiger shrimp (*Penaeus monodon*) throughout the Indo–Pacific region 40
3.1 Introduction 40
3.2 Material and Methods 42
3.2.1 Sample collection and DNA extraction 42
3.2.2 SNP genotyping and quality control 44
3.2.3 Population genetic diversity 44

3.2.4 Determination of outlier loci 45
3.2.5 Population genetic structure: neutral vs. outlier divergence 45
3.2.6 TreeMix analysis 47
3.2.7 Identification of putatively adaptive SNPs 47
3.3 Results 49
3.3.1 Genotyping and SNP quality control 49
Table 3.2 Filtering steps and SNPs counts retained after each step 49
3.3.2 Population genetic diversity 50
3.3.2 Determination of outlier loci 50
3.3.3 Genetic population structure 52
3.3.4 TreeMix analysis 57
3.3.5 Identification of putatively adaptive SNPs 58
3.4 Discussion 61
3.4.1 Population genetic structure and gene flow 62
3.4.1.1 Major genetic break within the Indo-Pacific 62
3.4.1.2 Regional population structure within the Pacific Ocean 63
3.4.1.3 Regional population structure within the Indian Ocean 65
3.4.2 Signatures of local adaptation 68
3.4.3 Implications for aquaculture and fishery management 69
3.5 Conclusion 70
Chapter 4. Genome-wide SNP markers reveal biogeographic impacts and confirm the sister taxa relationships in giant black tiger shrimp (*Penaeus monodon*) across the Indo–Pacific region 71
4.1 Introduction 71
4.2 Methods 73
4.2.1 Sample collection and DNA extraction 73
4.2.2 Sequence quality control and marker filtering 76
4.2.3 Phylogenetic analysis 77
4.3 Results 78

4.3.1 Genotyping and SNP/PAV filtering 78
4.3.2 Phylogenetic relationships among seven Penaeidae species 79
4.3.2.1 SNP-based ML and NJ trees 79
4.3.2.2 PAV-based ML and BI trees 81
4.3.3 Phylogeographic relationships among Indo–Pacific *P. monodon* samples 82
4.4 Discussion 86
4.4.1 Phylogeny of seven Penaeidae species 86
4.4.2 Phylogeographic relationships of Indo–Pacific *P. monodon* samples 87
4.5 Conclusion 89
Chapter 5. Development of a global SNP resource for diversity, provenance, and parentage analyses in the Indo-Pacific giant black tiger shrimp (*Penaeus monodon*) 90
5.1 Introduction 90
5.2 Methods 93
5.2.1 Tissue sample collection and genomic DNA extraction 93
5.2.2 SNP genotyping and quality control 94
5.2.3 Diversity SNP panel 95
5.2.4 Provenance SNP panel 97
5.2.5 Pedigree SNP panels 99
5.2.6 Genomic distribution of SNP panels 102
5.3 Results 102
5.3.1 Diversity SNP panel 102
5.3.2 Provenance SNP panel 108
5.3.3 Pedigree SNP panels 112
5.3.4 Genomic distribution of SNP panels 117
5.4 Discussion 118
5.5 Conclusion 122
Chapter 6. General Discussion 123
6.1 Significance of research and contribution to the field 123

6.2 Future research 128

6.2.1 Improved origin and demographic history reconstruction for Indo-Pacific black tiger shrimp *P. monodon* 128

6.2.2 Advancement of *P. monodon* selective breeding programs 129

6.2.3 *In situ* validation of identified low-density SNP panels 130

References 132

Appendices 156

List of Figures

Figure 2.1 Map showing the seven localities where 283 wild Australian *Penaeus monodon* samples were collected (BB: Bramston Beach, EB: Etty Bay, Townsville: TSV, Gulf of Carpentaria: GC, Joseph Bonaparte Gulf: JBG, Tiwi Island: TIW, and Nickol Bay: NKB). 15

Figure 2.2 SNP data analysis workflow from raw (*n* = 125,511) to final neutral (*n* = 10,535) and outlier (*n* = 89) loci datasets. 17

Figure 2.3 Population structure of 278 individuals of *Penaeus monodon* samples using 10,624 SNPs. (**A**) Discriminant Analysis of Principal Components (DAPC) scatterplot and (**B**) an individual density plot on the first discriminant function created through the R package '*adegenet* '; and (**C**) and (**D**) population networks constructed using the *NetView* P v.1.0 pipeline at kNN = 30 and 60. Dots represent individuals, whereas colored ellipses correspond to sampling origin. Population names are defined in Figure 2.1..... 25

Figure 2.4 Plots of isolation by distance (IBD) with Mantel correlograms for the relationship between genetic distance (F_{ST}) and geographic distance (km) among *Penaeus monodon* populations (*n* = 7) using 10,624 SNPs. 26

Figure 2.5 Venn diagram showing the total number of putative SNP significantly associated with at least one environmental variable. The total number of SNP is reported in each panel. The 89 outlier SNP on the intersections (shaded area) were retained for further analysis. .. 28

Figure 2.6 Principal components analysis based allele frequencies for (**A**) 10,535 neutral SNPs loci and (**B**) 89 outlier SNPs loci. Population names are defined in Figure 2.1 28

Figure 2.7 Redundancy analysis on (A) 10,535 neutral loci and (B) 89 outlier SNPs allele frequencies in seven populations of *Penaeus monodon* in Australia. Explanatory variables (arrows) were RDA axes retained as important variable selection accounting for genetic variation (Env_2: surface temperature maximum, Env_3: surface temperature minimum, Env_5: surface phytoplankton mean, Env_7: benthic current velocity mean, and Env_1: surface salinity mean). 30

Figure 3.1 Sampling sites and the known distribution of *Penaeus monodon* (modified from (Kongkeo, 2005–2020)) sourced in this study. Orange areas denote the know distribution of *P. monodon*. 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..... 42

Figure 3.2 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. Individuals of the same color belong to the same cluster. Population names are defined in Figure 3.1. 53

Figure 3.3 Plots of isolation by distance (IBD) with Mantel correlograms for the relationship between genetic (F_{ST}) and geographic (km) distances among *Penaeus monodon* populations (*n* = 16; **A**, **C**) and among clusters (*n* = 7; **B**, D) using 9,930 neutral SNPs (**A**, **B**) and 663 outlier SNPs (**C**, **D**)..... 54

Figure 3.4 (**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 use 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 3.1. 58

Figure 3.5 Venn diagram showing the overlapped putative adaptive SNPs identified by both environmental association and population differentiation approaches..... 60

Figure 4.1 Sampling locations for seven Penaeidae species (*n* = 340 individuals) used in this study. *Penaeu monodon* samples: Eastern Australia (PmoBB, Bramston Beach; PmoEB, Etty Bay; PmoTSV, Townsville); Northern Australia (PmoGC, Gulf of Carpentaria; PmoJBG, Joseph Bonaparte Gulf; PmoTIW, Tiwi Island); Western Australia (PmoWA); Vietnam (PmoCMVN, Ca Mau; PmoNTVN, Nha Trang); Fiji (PmoFJ); Indonesia (PmoINDO); Thailand (PmoTHL); Philippines (PmoPHI); Sri Lanka (PmoSLK); Kenya (PmoKE); South Africa (PmoSA). Non- *P. monodon* samples: *Fenneropenaeus merguiensis*: Fmerg; *Litopenaeus vannamei*: Lvan; *Penaeus esculentus*: Pescu; *Penaeus semisulcatus*: Psemi; *Penaeus longistylus*: Plong; and *Penaeus latisulcatus*: Plati. Map generated using R version 4.0.3..... **74**

Figure 4.2 Phylogenetic relationships of seven species from the Family Penaeidae using 4,496 SNP loci (**A**) and 7,054 PAV loci (**B**). (**A**) The maximum likelihood reconstruction was conducted using IQ-TREE v2.1.0 under HKY+F model with 10,000 bootstrap replicates and log likelihood scores of ln L = -20,273.726. (B) neighbour-joining (NJ) based on Nei's (1972) genetic distances in MEGA-X with 1000 bootstrap replicates. *Penaeus longistylus* was used as the outgroup taxon. Numbers at nodes represent bootstrap values in percentage..... 80

Figure 4.3 Phylogenetic relationships of seven Penaeidae species using 7,054 PAV loci. **(A)** Bayesian reconstruction generated in MrBayes v3.2.7 from 5,953 of the most credible set of trees. **(B)** The maximum likelihood reconstruction was conducted using IQ-TREE v2.1.0 under the GTR2+FO+R2 model with 10,000 bootstrap replicates. *Penaeus longistylus* was

used as the outgroup. Numbers at nodes represent posterior probabilities and bootstrap values in percentage..... 82

Figure 4.4 Phylogeographic analyses of *Penaeus monodon* using SNP and PAV loci datasets. Tree topology corresponds to the best ML tree. (A) ML tree generated in IQ-TREE v2.1.0 using 4,496 SNP loci under KY+F+R2 model. (B) BI tree generated in MrBayes using 7,054 PAV loci from 11,971 of the most credible set of trees. Numbers at nodes represent posterior probabilities and bootstrap values in percentage. Out-group Pescu: *Penaeus esculentus*; and Psemi: *Penaeus semisulcatus*. Legend for locality codes is the same as defined in Figure 4.1..... 83

Figure 4.5 Phylogeographic analyses of *P. monodon* using SNP and PAV loci datasets. (A) NJ tree generated in MEGA-X using 4,496 SNP loci based on Nei's (1972) genetic distances. (B) ML tree generated in IQ-TREE v2.1.0 using 7,054 PAV loci under GTR2+FO+R4 model. Numbers at nodes represent posterior probabilities and bootstrap values in percentage. Both *P. esculentus* (Pescu) and *P. semisulcatus* (Psemi) were used as outgroup. All abbreviations are the same as defined in Figure 4.1. 85

Figure 5.1 Sampling sites and the known distribution of Indo-Pacific *Penaeus monodon* (modified from Kongkeo (2005–2020)) sourced in this study. FJ: Fiji; Eastern Australia (EA; BB: Bramston Beach, EB: Etty Bay; TSV: Townsville); Northern Australia (NA; GC: Gulf of Carpentaria, JBG: Joseph Bonaparte Gulf, TIW: Tiwi Island); Western Australia (WA); Vietnam (VN; 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. 94

Figure 5.2 SNP filtering workflow from raw (n = 131,929) to SNP panels for determining genetic diversity, provenance, and pedigree analyses of *Penaeus monodon*. MAF: minor allele frequency, A_R: mean allelic richness, H_O: mean observed heterozygosity, H_E: mean expected heterozygosity, F_{ST}: pairwise population difference value..... 95

Figure 5.3. (A) Comparison of MAF fraction between the entire dataset (n = 10,593 SNPs), the full diversity SNP panel (n = 2,155), and the three smaller genetic diversity subsets (n = 1,500, 1,000, and 500) across sampled Indo-Pacific *Penaeus monodon* individuals (n = 520). (B-E) Pearson correlations between the MAF of fraction determined for each Indo-Pacific population using entire SNP dataset (n = 10,593; Chapter 4) and the full diversity SNP panel (B), 1,500 SNP marker subset (C), 1,000 SNP marker subset (D), and 500 SNP marker subset (E). 103

Figure 5.4 Pearson correlations between the entire SNP dataset (n = 10,593; Vu, et al., 2021) and the full diversity SNP panel (n = 2,155) and the 500 SNP marker subset for mean allelic richness (A_R; panels A and B), mean observed heterozygosity (H_O; panels C and D), and mean expected heterozygosity (H_E; panels E and F), respectively. 105

Figure 5.5. Mantel correlograms demonstrating the relationship between pairwise genetic differentiation (F_{ST}) based on the entire dataset dataset (n = 10,593; Chapter 3) and the full diversity SNP panel (n = 2,155; panel **A**) as well as the three smaller genetic diversity subsets consisting of 1,500 SNPs (panel **B**), 1,000 SNPs (panel **C**), and 500 SNPs (panel **D**)..... 108

Figure 5.6 Genomic clustering of Indo-Pacific *Penaeus monodon* sample (*n* = 520) based on the provenance SNP panel (*n* = 1,200). Discriminant Analysis of Principal Components (DAPC) scatterplots showing PC1 and PC2 (**A**) as well as PC2 and PC3 (**B**). Percentage of variance explained by each PC is indicated in parenthesis on each axis. Population networks constructed using *NetView* at kNN = 15 (**C**), kNN = 30 (**D**), and kNN = 50 (**E**). Dots represent individuals and colours correspond to sampling origin. FJ: Fiji; Eastern Australia (EA; BB: Bramston Beach, EB: Etty Bay; TSV: Townsville); Northern Australia (NA; GC: Gulf of Carpentaria, JBG: Joseph Bonaparte Gulf, TIW: Tiwi Island); Western Australia (WA); Vietnam (VN; CMVN: Ca Mau, NTVN: Nha Trang); PHI: Philippines; INDO: Indonesia; THL: Thailand; SLK: Sri Lanka; KE: Kenya; SA: South Africa..... 109

Figure 5.7 Monte Carlo tests based on the provenance SNP panel (*n* = 1,200) demonstrating the accuracy of tracing Indo-Pacific *Penaeus monodon* individuals (*n* = 520) back to their original basin (**A**), original region (**B**), original sampling location within Pacific Ocean (**C**), and original sampling location within Indian Ocean (**D**). Three proportions of individuals (50%, 70% and 90%) and F_{ST} ranked SNPs (top 25%, top 50%, top 75%, and 100%) were considered for all tests. FJ: Fiji; Eastern Australia (EA; BB: Bramston Beach, EB: Etty Bay; TSV: Townsville); Northern Australia (NA; GC: Gulf of Carpentaria, JBG: Joseph Bonaparte Gulf, TIW: Tiwi Island); Western Australia (WA); Vietnam (VN; CMVN: Ca Mau, NTVN: Nha Trang); PHI: Philippines; INDO: Indonesia; THL: Thailand; SLK: Sri Lanka; KE: Kenya; SA: South Africa; SEA: Southeast Asia. 111

Figure 5.8 Relationship between the number of pedigree SNP markers and exclusion probability (PE) estimates for correctly assigning parent-pair to an individual within each of the nine genetic units (Table 5.3). PE calculations were performed in CERVUS using multiple SNP subsets (*n* = 14) that consisted of various SNP marker quantities (*n* = 5 – 100) across 10 replicates for three discrete test scenarios (PE1, PE2, and PE3). PE1: incremental SNP marker subsets were selected high to low MAF-ranking from the PO, SLK, and WIO pedigree SNP panels; PE2: incremental SNP marker subsets were selected randomly from the PO, SLK, and WIO pedigree SNP panels; PE3: incremental SNP marker subsets were selected randomly from the filtered polymorphic datasets (MAF ≤ 0.01 and missing data per SNP ≥ 5%)..... 115

Figure 5.9 Predicted accuracy of Pacific Ocean (PO), Sri Lanka (SLK), and West Indian Ocean (WIO) pedigree SNP panels to correctly assign progeny to parents with 95% confidence using different sized marker subsets. Subsets were randomly selected from the corresponding PO, SLK, or WIO pedigree SNP panels (Table 5.3). 116

List of Tables

Table 2.1 Sampling codes and site locations with number of individuals (N). 15

Table 2.3 Genetic diversity indices for the *Penaeus monodon* populations sampled. 24

Table 2.4 Population differentiation estimates for *Penaeus monodon* populations sampled. 26

Table 2.5 Analysis of molecular variance (AMOVA) of *Penaeus monodon* from seven Australian locations (groups were assigned from DAPC analysis (see Figure 1). 27

Table 2.6 Population differentiation estimates for *Penaeus monodon* populations sampled using neutral (n = 963) and outlier loci (n = 89)..... 29

Table 2.7 Results of multiple regression of distance matrices (MRM) tests of geographic distance, genetic differentiation (F_{ST}) of 89 outlier loci, and significant environmental variables between *Penaeus monodon* populations in Australia. 31

Table 2.8 Characterization of high-quality BLAST matches obtained in comparison with the transcriptome *Penaeus monodon* database..... 32

Table 3.1 Fishing area, sampling location, population ID, code, number of individuals (N), and coordinates for each population of *Penaeus monodon* sampled..... 43

Table 3.2 Filtering steps and SNPs counts retained after each step..... 49

Table 3.3 Genetic diversity indices among 16 population of *Penaeus monodon* assessed using all 10,593 high quality filtered SNP loci..... 51

Table 3.4 Population differentiation estimates for *Penaeus monodon* populations sampled using both neutral (n = 9,930) and outlier (n = 663) loci. The only non-significant F_{ST} value (p > 0.05) is in bold type..... 55

Table 3.5 Analysis of molecular variance (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. 56

Table 3.6 Pearson correlations for putatively adaptive SNPs that exhibited a significant relationship between minor allele frequency and maximum or minimum sea surface temperature. 60

Table 3.7 Genetic diversity indices among 16 populations of *Penaeus monodon* assessed using 2,155 SNP loci (> 99% concurrence in technical replicates with missing data per SNP < 2%) 67

Table 4.1 Species code, species, sampling location, and total number of individuals (*N1*) obtained for each of the seven shrimp species included in this study. Number of individuals (*N2* and *N3*) used for a) phylogeny of non-*Penaeus monodon* species and b) phylogeography of *P. monodon* analyses, respectively. 75

Table 4.2 Filtering steps and SNPs counts retained after each step 79

Table 5.1 Genetic diversity indices (mean ± SE) for the sixteen Indo-Pacific *Penaeus monodon* populations used assessed all five SNP marker panels/subsets. Fiji (FJ), Eastern Australia (BB, EB, and TSV), Northern Australia (GC, JBG, and TIW), Western Australia (WA), (5) Vietnam (CMVN and NTVN), Indonesia (INDO), Thailand (THL), Sri Lanka (SLK), Kenya (KE), and South Africa (SA). 106

Table 5.2 The number of SNPs retained for pedigree analysis after sequential filtering applications. The proportion of SNPs retained at each threshold are indicated in parenthesis. 113

Table 5.3 Summary statistics (mean ± SD) for the Pacific Ocean (PO), Sri Lanka (SLK), and West Indian Ocean (WIO) pedigree SNP panels. 114

Table 5.4 Pedigree determination accuracy (95% confidence) across 10 simulated generations of Eastern Australia progenies using four SNP marker subsets (n = 30, 35, 40, and 50) that were randomly selected from the Pacific Ocean (PO) pedigree SNP panel (n = 217). 117

List of Appendices

Appendix 2.1 Marine environmental variables (23) based on latitude and longitude of seven sampled sites. 156

Appendix 2.2 Characterization of high-quality BLAST matches obtained in comparison with *P. monod*on sequencing SNP against *P. monodo*n transcriptome..... 157

Appendix 2.3 Translation of eight outlier SNP with transcriptome matches to protein sequences using all three reading frames..... 166

Appendix 3.1 Sea surface temperature maximum (SST_max) and minimum (SST_min) relating to each population sample site were selected from the Bio-ORACLE database using monthly averages of climate data from 2000 – 2014. 168

Appendix 3.2 Graph of the number of clusters formed by DAPC and ADMIXTURE. The top plot shows the cross-validation error across multiple runs picking K values of 7 and 6 as the optimum number of clusters for neutral (**A**) and outlier loci (**C**) datasets. Inference of the number of clusters in the DAPC performed on neutral (**B**) and outlier loci (**D**) datasets; K values of 7 and 6 (the lowest BIC value) represented the best summary of neutral and outlier loci, respectively..... 169

Appendix 3.3 Population relationship for tree models with ten migration events (m = 0 – 10) inferred by Treemix for neutral SNP dataset. The drift parameter represents the degree to which genetic drift has occurred between each population. BB: Bramston Beach; EB: Etty Bay; TSV: Townsville; GC: Gulf of Carpentaria, JBG: Joseph Bonaparte Gulf; TIW: Tiwi Island, WA: Western Australia; CMVN: Ca Mau, Vietnam; NTVN: Nha Trang, Vietnam; INDO: Indonesia; THL: Thailand; PHI: Philippines; SLK: Sri Lanka; KE: Kenya; SA: South Africa. 170

Appendix 3.4 The fractions of variance in relatedness between populations explained by phylogenetic models with 0 to 10 migration events..... 171

Appendix 3.5 List of all population triples (X; A, B) with negative values of f3 (Z; A, B) in three-way admixture test among 16 *P. monodon* populations. BB: Bramston Beach; EB: Etty Bay; TSV: Townsville; GC: Gulf of Carpentaria; JBG: Joseph Bonaparte Gulf; TIW: Tiwi Island; NKB: Nickol Bay; CMVN: Ca Mau, Vietnam; NTVN: Nha Trang, Vietnam; FJ: Fiji; INDO: Java, Indonesia; THL: Thailand; PHI: Philippines; SLK: Sri Lanka; KE: Kenya; SA: South Africa. 172

Appendix 3.6 Characterization of high-quality BLAST matches of 663 outlier SNPs to *Penaeus monodon* transcriptome contigs 178

Appendix 3.7 Putative functional annotations for 62 contigs that exhibited 100% pairwise identity to 27 outlier SNPs..... 220

Appendix 3.8 Correlation between minor allele frequency (Y axis) and sea surface temprature maximum (SST_max) and minumum (SST_min) for each of the seven transcriptome BLAST

identified outlier SNPs. Each plot includes loess smoothing function and 95% confidence intervals (grey area). 223

Appendix 3.9 Pearson correlations between minor allele frequency and sea surface temperature (maximum and minimum) for the 20 temperatures correlated putatively adaptive SNPs identified by combined EA and PD approach. Each plot includes loess smoothing function and 95% confidence intervals (grey area). 224

Appendix 3.10 Genetic diversity indices among 16 populations of *Penaeus monodon* assessed using 9,930 neutral loci..... 225

Appendix 3.11 Genetic diversity indices among 16 populations of *Penaeus monodon* assessed using 2,155 SNP loci (> 99% concurrence in technical replicates with missing data per SNP < 2%). 226

Appendix 4.1 Phylogeographic analyses of 317 *Penaeus monodon* individuals using SNP loci (n = 4,496) datasets. Tree topology corresponds to the best ML tree. ML tree generated in IQ-TREE v2.1.0 using 4,496 SNP loci under K3Pu+F+R6 model. Legend for locality codes is the same as defined in Figure 4.1. 227

Appendix 5.1 Custom R scripts to simulate progeny genotypes in R 228

Appendix 5.2 Genomic distribution of five SNP panels exhibited the chromosome-scale genome assembly of *P. monodon* (Uengwetwanit et al., 2021)..... 235

Appendix 6.1 Published journal article reprints 269

Abbreviations

AMOVA	Analysis of molecular variance
A_R	Allelic richness
A_{RA}	Private allelic richness
Av. MLH	Average multi-locus heterozygosity
BIC	BIC: Bayesian information criterion
BI	Bayesian inference
BLAST	Basic local alignment and search tool
CLCA	Calcium-activated chloride channel regulator
CTAB	Cetyl-trimethyl ammonium bromide protocol
DAPC	Discriminant analysis of principal components
DArTcap	Target capture genotype by sequencing
DArTseq	Diversity Arrays Technology (Canberra Australia)
DNA	Deoxyribonucleic acid
EA	Environmental association
ESU	Evolutionary Significant Unit
FAO	Food and Agriculture Organisation
FDR	False discovery rate
F_{IS}	Inbreeding coefficient
F_{ST}	Pairwise genetic differentiation values
H_E	Expected heterozygosity
H_O	Observed heterozygosity
HWE	Hardy–Weinberg equilibrium
IBD	Isolation by distance
Illumina RNA-seq	Illumina iSelect genotyping array
IR	Internal relatedness
kNN	Multiple k-nearest neighbor
LD	Linkage disequilibrium
LFMM	Latent factor mixed models
MAF	Minor allele frequency
ML	Maximum likelihood
MRM	Multiple regression on distance matrices
mtDNA	Mitochondrial DNA
N_{eLD}	Effective population size
NJ	Neighbour-joining
NGS	Next-generation sequencing
PAV	Presence-absence variant
PC	Principal component
PCA	Principal component analysis

PD	Population differentiation
PPL	Percentage of polymorphic loci
pRDA	Partial redundancy analysis
R^2	Coefficient of determination
RDA	Redundancy analysis
SD	Standard deviation
SE	Standard error
SNP	Single nucleotide polymorphism
sMLH	Standardized multilocus heterozygosity

Chapter 1. General introduction

1.1 Food security and aquaculture

The global population is expected to increase from 7.7 billion in 2019 to 9.7 billion in 2050 and 10.9 billion in 2100 (DESA, 2019). According to FAO et al. (2021), in 2019 around 750 million people worldwide (i.e., approximate one in ten) were exposed to severe levels of food insecurity. The aquaculture sector is among the most significant and fastest-growing sectors of the global agrifood system and is considered as a primary protein source to feed the global population (Sheikha & Xu, 2017). There are currently approximate 400 species farmed using aquaculture methods (Naylor et al., 2021), which collectively produced 82.1 million tonnes in 2018 that accounted for 52% of human fish consumption (FAO, 2020). Due to the decrease in wild harvesting, aquaculture production must rapidly expand to meet the growing demand. The current limitations in aquaculture production could be due to the lack of genetically improved stocks, inconsistent seed stock quality, and/or disease outbreaks (Gjedrem & Robinson, 2014; Yue & Wang, 2017). Thus, sustainable aquaculture development and fisheries management require in-depth understanding about the biology (e.g., population genomics) of key species (e.g., crustaceans).

Crustaceans have important roles in many ecosystems and represent an important aquatic food source for humankind. Between 2010 and 2018, crustacean aquaculture production increased from 5.5 to 9.4 million tonnes (FAO, 2020). Among edible crustaceans in aquaculture, two penaeid shrimp species, whiteleg shrimp (*Litopenaeus vannamei*) and giant black tiger shrimp (*Penaeus monodon*), are predominately cultured (Thornber et al., 2020) and have an annual production exceeding 3.2 million tonnes since 2010. Of this, *P. monodon* farm production increased from 562,900 to 750,000 tonnes; however, its overall proportion of global crustacean aquaculture has decreased over this time period from 10% to 8% (FAO, 2020). Compared to *P. monodon*, the production of *L. vannamei* increased more than double from 2,648,500 tonnes in 2010 to 4,966,200 tonnes in 2018 (FAO, 2020). The rapid growth of the global *L. vannamei* industry has been attributed, in large part, to the success achieved in the domestication, selective breeding, and translocation (e.g., introduced to Asia) of this species (Briggs, 2005; Moss et al., 2012). In contrast, the limited production of *P. monodon* could be due to the lack of industry-ready genetically improved stocks, inconsistent seedstock quality,

and/or numerous devastating disease outbreaks (Guppy et al., 2020; Nguyen et al., 2020). Accordingly, there are ongoing genetic improvement programs for *P. monodon* in Australia, Thailand, Hawaii, Madagascar, and Vietnam (Foote et al., 2019; Nguyen et al., 2020; Norman-López et al., 2016). However, these projects still have limitations due to a number of challenges, including low survival during the grow-out phase and a rapid decline in the reproduction capacity of broodstock in captivity (Arnold et al., 2013; Nguyen et al., 2020). Consequently, research priorities within the *P. monodon* industry are now focusing on the development of breeding programs to address these issues and increase productions. Access to effective genetic marker suites is an integral resource for breeding programs and fishery management since they allow the: 1) assessment of wild and farmed population genetic diversity, 2) determination of the origin population for individuals (i.e., provenance), and 3) accurate parentage assignment (i.e., pedigree) throughout the hatchery and production systems.

1.2 Gene flow and local adaptation in marine species

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 (Hemmer-Hansen *et al*., 2007; Matolweni *et al*., 2000). Population genomics research provides valuable estimates regarding genetic diversity, population structure, genetic connectivity, and local adaptation, which are equally important factors for the adaptive capacity of key species under climate change scenarios (Grant et al., 2017; Lal et al., 2017; Oleksiak, 2019). Population genomics research also enhances the facilitation of appropriate aquaculture and fisheries management practices by providing novel insights into broodstock selection for breeding programs and discernment of hatchery stocks from wild populations (Bernatchez *et al*., 2017; Gjedrem, 2012; Lorenzen *et al*., 2012; Williams & Hoffman, 2009). In order to make accurate recommendations regarding aquaculture programs (e.g., seletive breeding) or fishery management plans for commercially important species (e.g., *P. monodon*), it is important to comprehensively understand the fine- and broad-scale genetic structure, genetic diversity, and local adaptation within and among populations (Lal et al., 2017; Woodings et al., 2018; Silliman, 2019).

1.2.1 Gene flow in marine species

Crustaceans are a major constituent of global aquatic ecosystems with over 66,000 known species (LeBlanc, 2007). The life history of many marine crustacean species comprises a planktonic larval phase that lasts for weeks-to-months-to-years depending on the species (Montgomery et al., 2006; Ptacek et al., 2001). For example, the larval development of *P. monodon* consists of four stages that span a total of approximate 20 – 26 days (Motoh, 1985) while the planktonic larval duration of the spiny lobsters *Panulirus ornatus* and *P. cygnus* are 4 – 8 months and 7 – 14 months, respectively (Phillips et al., 2013; Pitcher et al., 2005). Additionally, environmental factors such as oceanic currents and wind drift have the potential to disperse crustacean larvae over vast geographical distances (Hänfling et al., 2011; Teske et al., 2008). The interactions between physical and biophysical processes that predominantly influence the larval dispersal of crustaceans have been reviewed by Cowen and Sponaugle (2009). In summary, the physical dispersal process (e.g., waves, currents, tidal, and sediment movement) can mix larvae between populations or expand spatial distribution while larval behaviours (e.g., fecundities, mobility, and mortality) also influence dispersal patterns. Moreover, duration of the planktonic larval phase has been shown to be the dominant dispersal characteristic that determines genetic structuring and gene flow within and among localities (Pascual et al., 2017; Selkoe & Toonen, 2011).

Gene flow describes the exchange of genetic information between populations through migration whereas dispersal is defined as the movement of individuals from one genetic population to another (Allendorf & Luikart, 2007; Cowen et al., 2006). The particular type of connectivity varies depending on the species-specific duration of the pelagic larval phase, the velocity and spatiotemporal variability in ocean currents, and spatial configuration of localities (Treml et al., 2008). Many different approaches have been developed to estimate the connectivity between distal populations including geographic patterns, biophysical models, geochemical models, and genetic markers (reviewed by Cowen and Sponaugle (2009), Kool et al. (2013), and (Sexton et al., 2014)). For example, Vogler et al. (2012) presented the following three hypotheses as potential explanations for the observed differences in the genetic structure of crown-of-thorns starfish *Acanthaster planci* populations throughout the Indian Ocean, which are believed to have diverged during the late-Pliocene to early-Pleistocene (1.86 – 2.89 million years ago): 1) landmass distribution differences, 2) geographical barriers (e.g., lack of coral reefs between regions), and 3) ocean surface currents. Within the Indo-Pacific region, blacktip

shark (*Carcharhinus limbatus*) from the Western Indian Ocean (WIO) was genetically related to Australia and Indo-Pacific populations (Merwe et al., 2019) and deep-water snapper (*Pristipomoides filamentosus*) was genetically related between Guam, Australia (Ashmore Reef), New Caledonia, and Tongan populations (Gaither et al., 2011). Genetic structure of wild populations also appears to be strongly influenced by gene flow as a result of effective migration and/or dispersal (Hartl et al., 1997; Vogler et al., 2012). Interestingly, gene flow can constrain adaptation of populations inhabiting different ecological environments by homogenizing allele frequencies, which limits the effects of local selection pressures by reducing the number of interactions between genes and environmental variables (Hendry & Taylor, 2004).

1.2.2 Local adaptation in marine species

Local adaptation can occur when geographically discrete environmental conditions impose selection pressure on resident populations of widespread species. Within heterogeneous populations, local adaptation can arise when strong environmental selection pressure outweighs the influence of genetic drift and gene flow towards homogeneity (Gandon & Michalakis, 2002). The marine environment provides sufficient opportunities for homogenizing gene flow through passive dispersal of eggs and larvae with ocean currents and adult mobility (Waples, 1998). Resident populations of discrete marine environments can initially differ genetically at a few sites in their genomes (e.g., allele frequency differences; (Beaumont & Balding, 2004)) or rapidly adapt to new environments due to the presence of intense selection pressure(s) (Gandon & Michalakis, 2002; Morin et al., 2004). Thus, understanding how environmental features shape the genetic structure of populations (i.e., extent and scale of local adaptation) is crucial because it helps determine how populations evolve (Hecht et al., 2015; Jump et al., 2006; Rellstab et al., 2015). Recently, local adaptation of marine organisms to environmental conditions has been investigated in commercially important marine invertebrates, including greenlip abalone (*Haliotis laevigata*) (Sandoval-Castillo et al., 2018), eastern oyster (*Crassostrea virginica*) (Bernatchez et al., 2019), American lobster (*Homarus americanus*) (Benestan et al., 2016), and New Zealand scallop (*Pecten novaezelandiae*) (Silva & Gardner, 2015).

1.3 Application of genetic markers and techniques to aquaculture and fishery management

Molecular genetic (i.e., DNA) markers are powerful tools for the accurate determination of genetic uniqueness or divergence within and among individuals, populations, and species (Askari et al., 2013; Z. Liu & Cordes, 2004; Park & Moran, 1995). There are a number of well-established DNA markers used to assess genetic diversity and population structure, including allozymes, mitochondrial DNA (mtDNA), Restriction Fragment Length Polymorphism (RFLP), Randomly Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphisms (AFLP), microsatellites, and Single-Nucleotide Polymorphisms (SNPs) (Benzie, 2000; Liu & Cordes, 2004; Liu, 2011; Guppy et al., 2020). The application of these DNA markers in aquaculture and fishery management has been reviewed by Liu and Cordes (2004), Askari et al. (2013), Tanya and Kumar (2010), and Robledo et al. (2018). Of these DNA markers, three (mtDNA, microsatellites, and SNPs) have been demonstrated to be the most useful for ecology, conservation, population genomics, and local adaptation studies (Hauser & Seeb, 2008; Schoville et al., 2012; Guppy et al., 2020).

Despite being highly variable in natural populations, effectively haploid, and only inherited maternally mtDNA is easy to isolate and manipulate (e.g., amplify and sequence), which makes it a particularly useful DNA marker for tracking divergence among lineages, variance, and population size fluctuations (Galtier et al., 2009; McCusker & Bentzen, 2010). Datasets based on mtDNA markers have been recently reported for many shrimp species (Alam et al., 2016; Benzie et al., 2002; Tsoi et al., 2007; Zitari-Chatti et al., 2009); however, the non-Mendelian inheritance of mtDNA may undermine its application as a robust indicator of gender-biased migration, species relationships, or demographic history given that it does not provide insights regarding the Mendelian inherited nuclear genome (Ballard & Whitlock, 2004). Microsatellites are co-dominant, simple nucleotide sequence repeat DNA markers that are abundant, distributed throughout the genome, exhibit high levels of allele polymorphism, and occur in both coding and noncoding DNA regions (Abdul-Muneer, 2014; Chistiakov et al., 2006). Microsatellites have been used to assess population genetic structure, genetic differentiation, and gene flow in various marine invertebrates, including: 1) zhikong scallop, *Chlamys farreri* (Zhan et al., 2009), 2) western rock lobster, *Panulirus cygnus* (Kennington et al., 2013), 3) swamp ghost crab, *Ucides Cordatus* (Oliveira-Neto et al., 2014), and 4) ornate spiny lobster, *Panulirus ornatus* (Dao et al., 2015). Microsatellites have also been used for

pedigree determination in a variety of aquaculture species (Jerry et al., 2006; Liu & Cordes, 2004; Vandeputte & Haffray, 2014; Vela-Avitúa et al., 2013). However, microsatellites do have some disadvantages such as risk of homoplasy for allele size, the presence of null alleles (Putman & Carbone, 2014), and the requirement for large sample sizes to generate accurate allele frequency and diversity estimates (Hale et al., 2012).

SNP markers are rapidly replacing mtDNA and microsatellite markers in seascape genomics, population genomics, phylogenomic, genetic diversity, provenance, and pedigree studies given their demonstrated advantages, which include: 1) genome-wide coverage, 2) reliable population structure inferences, 3) potential for genome-wide trait association studies, and 4) improved genomic breeding value estimates (Arbon et al., 2021; Bruneaux et al., 2013; Nguyen et al., 2014). SNPs have been shown to be the most prevalent type of polymorphism within investigated genomes (Morin et al., 2004) as well as useful in the assessment of local adaptation within and among populations through the identification of putatively adaptive loci (Excoffier & Lischer, 2010; Foll & Gaggiotti, 2008; Frichot & François, 2015; Luu et al., 2017; Oksanen et al., 2015; Whitlock & Lotterhos, 2015). More specifically, SNP-based identification of putatively adaptive loci has been divided into two main approaches: 1) population differentiation (PD) and 2) environmental association (EA). The PD approach addresses traditional population differentiation by estimating the inherent variance in pairwise genetic differentiation values (F_{ST}) of each locus across the entire genome without assuming the effect of environmental variables on outlier SNP generation (Jensen et al., 2016; Luu et al., 2017). Alternatively, the EA approach addresses the environmental influence on outlier SNP generation by identifying candidate adaptive loci that covary with environmental factors based on allele frequency (Dalongeville et al., 2018; De Mita et al., 2013). Given the advantages and disadvantages of each approach, unique combinations of PD and EA may help to reduce the number of false negatives and promote uncovering potential genomic footprints of selection (Dalongeville et al., 2018). 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 (Anderson et al., 2019; Nowland et al., 2019; Woodings et al., 2018). 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) while alternative analysis methods permit the separation of putatively adaptive (i.e., outlier) and neutral loci into two discrete datasets (Excoffier & Lischer, 2010; Luu et al., 2017).

Phylogenomics play an important role in describing evolutionary processes and identifying historical events such as geographic and ecological drivers that shaped the current population genetic structure within species (Avise, 2000; Bowen et al., 2016; Van Cise et al., 2019). Waqairatu et al. (2012) and You et al. (2008) generated phylogeographic trees for Indo-Pacific *P. monodon* based on mtDNA and microsatellites datasets, which revealed the existence of two divergent lineages (Pacific and Indian Oceans); however, these studies did not include a comprehensive sampling of sites throughout the Indian Ocean so some phylogenomic resolution may have been lost (i.e., existence of additional lineages). Additionally, mtDNA datasets exhibit non-Mendelian inheritance and thus provide only a single (maternal) perspective, which may not reliably reflect species relationships or demographic history (Ballard & Whitlock, 2004) whereas microsatellite datasets tend to lack critical components necessary for robust phylogenomic analyses (Edwards et al., 2015); therefore, mtDNA and microsatellite datasets have limited power to resolve genetic structure and phylogeographic patterns (Jeffries et al., 2016; Zarraonaindia et al., 2012). In light of these limitations, SNPs are becoming the marker of choice in phylogeographic and phylogenomic studies (Hodel et al., 2017; Jeffries et al., 2016; Morin et al., 2004) because this approach provides more (i.e., cost-effective) data for resolving phylogenetic relationships at deep evolutionary timescales and alleviates concerns regarding intragenic recombination (Harvey et al., 2016; Leaché & Oaks, 2017).

The application of population genomics to fisheries and aquaculture management has been throughly reviewed by Benestan (2019); Valenzuela-Quiñonez (2016); Casey et al. (2016); Dudgeon et al. (2012); and Ovenden et al. (2015). These studies presented that genetic techniques have developed rapidly with the increasing affordability of new analyses, which bring a new paradigm for fisheries and aquaculture management, including 1) genomic stock identification, 2) traceability of fisheries resources, 3) population demography, 4) genetic diversity, population abundance and resilience, 5) and detection of pathogens and invasive species. The information of population genomics helps stakeholders anticipate the potential effects on the resilience of fisheries stocks and improve management of marine organisms. Additionally, the detection of adaptive loci (outlier loci) provides unprecedented opportunity to understand bio-complexity of the stock and potential to understand the genetic basis of micro-evolutionary effects of fisheries-induced evolution and climate change (Ovenden et al., 2015; Valenzuela-Quiñonez, 2016). For example, outlier loci revealed additional barriers to gene flow in Atlantic cod *Gadus morhua* populations throughout the North Atlantic, which

were not apparent at neutral loci (Bradbury et al., 2013). Thus, fisheries management requires an accurate spatial determination of local adapted populations (Valenzuela-Quiñonez, 2016).

1.4 Population genetics and knowledge gaps of Black tiger shrimp aquaculture and fishery management

The giant black tiger shrimp *P. 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 (approximate 20 days) before migrating to nursery areas, such as in-shore areas and mangrove estuaries (Motoh, 1985; Queiroga & Blanton, 2004). Larvae and post-larvae can display daily vertical migration behaviours in synchrony with tidal current (Queiroga & Blanton, 2004; Rothlisberg, 1982). 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 to 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 despite the existence of geographic barriers between the Indian and Pacific Oceans (e.g., Sunda and Sahul shelves) and between the East and West Indian Ocean (Waqairatu et al., 2012; You et al., 2008).

To date, most studies that assessed population genetic structure and connectivity patterns of *P. monodon* throughout the Indo-Pacific region have used mtDNA, microsatellites, or allozyme markers (Benzie et al., 2002; Waqairatu et al., 2012; You et al., 2008). These markers can often have limited resolution power to detect fine-scale genetic structure and signatures of adaptation within and among populations (Allendorf et al., 2010; Nielsen et al., 2009). Previous research on the fine-scale population structure of Australian *P. monodon* that used mtDNA, microsatellite, and allozyme markers suggested that there was no genetic differentiation between Northern and Eastern populations, but that *P. monodon* from Western Australia was a separate genetic stock with reduced allelic variation due to colonisation by east coast *P. monodon* individuals (Benzie et al., 2002; Benzie et al., 1993; Benzie et al., 1992; Brooker et al., 2000). Another microsatellite-based study on Indo-Pacific *P. monodon* reported that *P. monodon* from Northern Australia separated into a discrete cluster while *P. monodon* from Queensland and Western Australia grouped with Thailand, Palau, PNG, Taiwan,

Philippines, and Vietnam (Bac Lieu and Can Tho) into one cluster (Waqairatu et al., 2012). More broadly, previous studies using mtDNA and/or microsatellites demonstrated low-level population differentiation throughout the Indo-Pacific region as well as lower genetic diversity within Indian Ocean (IO) than Pacific Ocean (PO) *P. monodon* populations; however, levels of predicted gene flow were not determined nor were patterns of genetic connectivity among populations consistent across these studies (Waqairatu et al., 2012; You et al., 2008). Despite limited resolution provided by mtDNA and microsatellite markers, the population structure and genetic connectivity of *P. monodon* within and among Indo-Pacific populations remains unclear (i.e., inconsistencies across previous studies) while information about local adaptation among Indo-Pacific populations remains limited (i.e., few informative studies to date). Attainment of such population genomics information (e.g., using SNPs) could help improve the sustainable development of commercial aquaculture practices (e.g., selective breeding) and wild fishery management strategies (e.g., persistence of healthy wild populations).

Various next-generation sequencing technologies have been used for *P. monodon* SNP genotyping, including Illumina iSelect genotyping array (Baranski et al., 2014)) and target capture genotype by sequencing (DArTcap; (Guppy et al., 2020)). However, the subsequently identified SNP marker subsets (i.e., panels) for application-specific genotyping have not been broadly implemented to date because SNP genotyping with high-density (n > 10,000) panels can be cost-prohibitive for routine and/or high-volume applications (Gebrehiwot et al., 2021; Herry et al., 2018; Song & Hu, 2021). For example, Duarte et al. (2013) reported that the cost of genotyping pigs with a high-density commercial SNP panel (n = 60,000 SNPs) was more than twice that of genotyping the same individuals with a low-density commercial SNP panel (n = 9,000 SNPs). Cost-effective low-density SNP panels (n < 10,000) have been developed for routine breeding applications in threespine stickleback (*Gasterosteus aculeatus*; (Ferchaud et al., 2014)), Atlantic salmon (*Salmo salar L.*; (Holman et al., 2017)), brown trout (*Salmo trutta;* (Saint-Pé et al., 2019)), red tilapia (*Oreochromis spp*.; (Sukhavachana et al., 2021)), and Australian *P. monodon* (Guppy, et al., 2020); however, to date, no cost-effective low-density SNP panels have been identified that permit the accurate determination of genetic diversity, provenance, and pedigree within and among all Indo-Pacific *P. monodon* populations. Once identified, such low-density SNP panels could potentially advance the Indo-Pacific *P. monodon* aquaculture industry through 1) improved consistency in SNP genotype data across studies (e.g., direct comparison across common SNP panels), 2) development of a global genotype reference database that encompasses both wild and farmed populations, and 3)

establishment of robust and globally accepted analysis standards for genetic diversity, provenance, and pedigree determination.

1.5 Thesis overview

The overarching goal of this thesis was to provide new population genomics insights about and resources for Indo-Pacific *P. monodon* that help advance aquaculture (e.g., selective breeding programs) and fishery management (e.g., stock identification) practices. This goal was achieved through four interdependent studies (i.e., Chapters), which aimed to:

- Utilize both neutral and outlier SNP datasets to determine the fine-scale population structure, connectivity, and putative local adaptation of Australian *P. monodon* populations (Chapter 2).
- Utilize both neutral and outlier SNP datasets to determine the broad-scale population structure, connectivity, and local adaptation of Indo-Pacific *P. monodon* populations (Chapter 3).
- Evaluate the phylogenomic and phylogeographic relationships among Indo–Pacific *P. monodon* populations and sister taxa using genome-wide SNP and presence-absence variant (PAV) loci datasets (Chapter 4).
- Identify low-density SNP panels that permit cost-effective and accurate diversity, provenance, and pedigree determination for all Indo-Pacific *P. monodon* populations (Chapter 5).

1.6 Publication arising from this thesis

Peer reviewed scientific journal articles produced over the PhD candidate which are either published (Appendix 6.1) or in preparation at the time of writing are as follows:

- **Vu, N. T. T**., Zenger, K. R., Guppy, J. L., Sellars, M. J., Silva, C. N. S., Kjeldsen, S. R., & Jerry, D. R. (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. doi:10.1186/s12864-020-07084-x.

- **Vu, N. T. T**., Zenger, K. R., Silva, C. N. S., Guppy, J. L., & Jerry, D. R. (2021). Population structure, genetic connectivity, and signatures of local adaptation of the giant black tiger shrimp (*Penaeus monodon*) throughout the Indo–Pacific Region. Genome biology and evolution. doi:10.1093/gbe/evab214.
- **Vu, N. T. T**., Zenger, K. R., Jones, B. J., Edmunds, R. C., & Jerry, D. R. (*in final prep*). Identification of suitable SNP markers for rapid and accurate diversity, provenance, and pedigree analyses of Indo-Pacific giant black tiger shrimp (*Penaeus monodon*). In preparation for *Aquaculture*.

Chapter 2. Fine-scale population structure and evidence for local adaptation in Australia giant black tiger shrimp (*Penaeus monodon*) using SNP analysis

2.1 Introduction

Local adaptation can occur when geographically discrete environmental conditions impose selection pressure on resident populations of widespread species. Within heterogeneous populations, local adaptation can arise when strong environmental selection pressure outweighs the influence of genetic drift and gene flow towards homogeneity (Gandon & Michalakis, 2002). The marine environment provides sufficient opportunities for homogenizing gene flow through passive dispersal of eggs and larvae with ocean currents and adult mobility (Waples, 1998). Resident populations of discrete marine environments can initially differ genetically at a few sites in their genomes (e.g., allele frequency differences; (Beaumont & Balding, 2004), or can change rapidly to adapt to a new environment due to intense selection pressure (Gandon & Michalakis, 2002; Morin et al., 2004). Thus, understanding how environmental features shape the genetic structure of populations is crucial because it helps to determine how populations evolve, along with the extent and scale of local adaptation (Hecht et al., 2015; Jump et al., 2006; Rellstab et al., 2015). Recently, local adaptation of marine organisms to environmental conditions has been investigated in commercially important marine invertebrates, including greenlip abalone (*Haliotis laevigata*) (Sandoval-Castillo et al., 2018), eastern oyster (*Crassostrea virginica*) (Bernatchez et al., 2019), American lobster (*Homarus americanus*) (Benestan et al., 2016), and New Zealand scallop (*Pecten novaezelandiae*) (Silva & Gardner, 2015). However, to date, the influence of marine environmental conditions on the wild-type genetic diversity and structure of Australian giant black tiger shrimp (*Penaeus monodon;* Fabricius, 1798), is unknown.

Single nucleotide polymorphisms (SNP) have been demonstrated to be the most prevalent type of polymorphism within investigated genomes (Morin et al., 2004). Moreover, SNP are becoming the marker of choice in genetic diversity, population genetics, and seascape genomics studies (Nadeem et al., 2018). Additionally, genetic techniques offer a variety of methods using SNP datasets to detect putatively adaptive loci and assess how environmental parameters influence the extent of genetic variation within and among populations (Excoffier & Lischer, 2010; Foll & Gaggiotti, 2008; Frichot & François, 2015; Luu et al., 2017; Oksanen

et al., 2015; Whitlock & Lotterhos, 2015). The application of these analyses has been divided into two main approaches: population differentiation (PD) and environmental association (EA). More specifically, the PD approach addresses traditional population differentiation by estimating the inherent variance in pairwise genetic differentiation values (F_{ST}) of each locus across the genome without assuming the effect of environmental variables on outlier SNP generation (Jensen et al., 2016; Luu et al., 2017). Alternatively, the EA approach addresses environmental influence on outlier SNP generation by identifying candidate adaptive loci that covary with environmental factors based on allele frequency (Dalongeville et al., 2018; De Mita et al., 2013). EA applications are especially promising to detect loci under putative selection, because they do not require phenotypic nor experimental data, can define the contribution of environmental variables to adaptive genetic variation, and detect weak multilocus responses to environmental conditions (Dalongeville et al., 2018; De Mita et al., 2013; Hendricks et al., 2018; Xuereb et al., 2018). Given the unique advantages and disadvantages of each approach, unique combinations of PD and EA may help to reduce the number of false negatives and promote uncovering potential genomic footprints of selection (Dalongeville et al., 2018). For example, both PD and EA methods have been employed for the successful detection of putatively adaptive loci and assessment of environmental parameter influence on genetic variation within and among populations (Andrew et al., 2018; Benestan et al., 2016; Dalongeville et al., 2018; Rellstab et al., 2015).

Penaeus monodon is widely distributed throughout the Indo-Pacific and is found along the Western, Northern, and Eastern coasts of Australia (Brooker et al., 2000; Gribble et al., 2003; Motoh, 1985). The life history of *P. monodon* comprises an offshore planktonic larval phase, estuarine juvenile and adolescent phases, and an inshore adult phase (Motoh, 1985). Larvae and post-larvae can display daily vertical migration behaviours in synchrony with tidal current (Queiroga & Blanton, 2004; Rothlisberg, 1982). A tagging study by Gribble et al. (2003) revealed that shrimp in the Trinity Inlet closure in Eastern Australia moved an average distance of 31 km north (5 - 100 km) over an average of 74 days (10 - 111 days). Forbes et al. (1999) found evidence for high gene flow among *P. monodon* samples collected in Madagascar, Mozambique, and South Africa, despite geographic separations of up to 2,000 km. Thus, *P. monodon* exhibits the capacity for moderate to high-level gene flow (i.e., genetic connectivity) across large geographic distances. In Australia, previous genetic work on *P. monodon* determined the existence of population structure using microsatellites, allozymes and mitochondrial DNA (mtDNA) markers (Benzie et al., 2002; Benzie et al., 1993; Benzie et al.,

1992; Brooker et al., 2000). These studies suggested that there was no genetic differentiation between Northern and Eastern populations, but that *P. monodon* from Western Australia was a separate genetic stock with reduced a number of allelic due to colonisation by east coast *P. monodon* individuals. Additionally, another previous study of Indo-Pacific *P. monodon* presented that *P. monodon* from Northern Australia separated into a discrete cluster while Queensland and Western Australia grouped with Thailand, Palau, PNG, Taiwan, Philippines, and Vietnam (Bac Lieu and Can Tho) in one cluster (Waqairatu et al., 2012).

The main objective of this study was to resolve the population genetic structure of *P. monodon* in Australia using a high-resolution genome-wide SNP approach and identify any genomic signatures of local adaptation among geographically and environmentally discrete populations. More specifically, this study aimed to 1) assess the levels of gene flow, genetic diversity, and population structure among geographically discrete populations of Australian *P. monodon* for evidence of local adaptation, and 2) predict which environmental factors are most likely to influence Australian *P. monodon* population structure (i.e., drive local adaption). The implications of *P. monodon* genetic structure and local adaptation for management and conservation of Australian populations are discussed.

2.2 Methods

2.2.1 Sample collection and DNA extraction

Individual *P. monodon* ($n = 283$) were collected from seven locations around Australia between 2015 and 2017 (Figure 2.1, Table 2.1). More specifically, wild adult *P. monodon* were obtained by trawling from Bramston Beach ($n = 60$, BB), Etty Bay ($n = 50$, EB), Townsville ($n = 22$, TSV), Joseph Bonaparte Gulf ($n = 34$, JBG), Tiwi Island ($n = 56$, TIW), Gulf of Carpentaria ($n = 35$, GC), and Nickol Bay ($n = 26$, NKB) (Table 2.1). Pleopod tissue samples were collected from each adult individual using pre-sterilised scissors followed by preservation in 80% ethanol and -20˚C storage until DNA extraction.

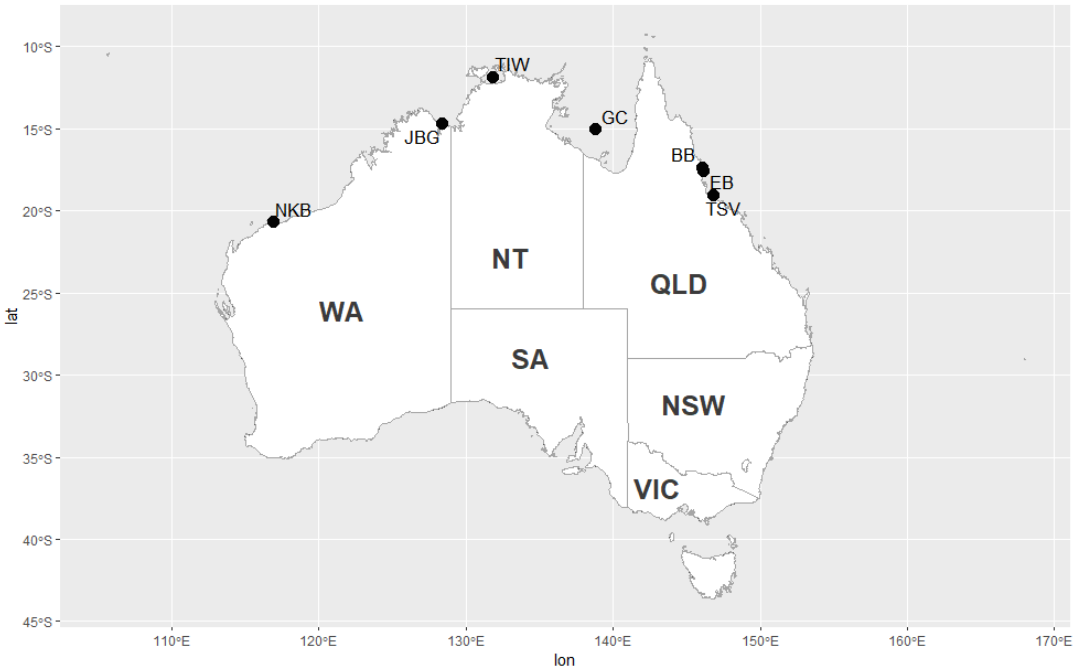


Figure 2.1 Map showing the seven localities where 283 wild Australian *Penaeus monodon* samples were collected (BB: Bramston Beach, EB: Etty Bay, Townsville: TSV, Gulf of Carpentaria: GC, Joseph Bonaparte Gulf: JBG, Tiwi Island: TIW, and Nickol Bay: NKB).

Table 2.1 Sampling codes and site locations with number of individuals (N).

Regions	Sampling locations	Code	N	Latitude	Longitude
Eastern Australia	Bramston Beach	BB	60	-17.353	146.0527
Eastern Australia	Etty Bay	EB	50	-17.5617	146.109
Eastern Australia	Townsville	TSV	22	-19.0821	146.8175
Northern Australia	Gulf of Carpentaria	GC	35	-14.9977	138.7808
Northern Australia	Joseph Bonaparte Gulf	JBG	34	-14.6818	128.3714
Northern Australia	Tiwi Island	TIW	56	-11.8668	131.8315
Western Australia	Nickol Bay	NKB	26	-20.697	116.8632

Genomic DNA (gDNA) was extracted from each pleopod tissue sample following a modified cetyl trimethyl ammonium bromide (CTAB) protocol (Adamkewicz & Harasewych, 1996; Lal et al., 2018). Briefly, 100 mg preserved tissue was digested in a CTAB buffer with 20 mg/mL proteinase K for ≥ 6 hr at 55 ˚C under gentle agitation followed by one phenol–chloroform–isoamyl alcohol phase separation, two chloroform – isoamyl alcohol phase separations, overnight isopropanol precipitation (4˚C), two 70% ethanol washes, and elution in

Tris-EDTA buffer (Gomes et al., 2007). All gDNA extractions were subsequently cleaned using Sephadex™ G-50 (GE, 2007). Extracted gDNA samples were assessed for quantity and quality using Nanodrop 1000 Spectrophotometer (260:230 nm and 260:280 nm ratios; ThermoFisher Scientific Pty Ltd, Australia) and 0.8% agarose gel containing GelGreen (ThermoFisher Scientific Pty Ltd, Australia). All gDNA samples were then diluted with Tris-EDTA buffer to a normalized final concentration of 50 ng/µl before shipping to Diversity Arrays Technology Sequencing (DArTseq™; Canberra Australia) for library preparation and DArTseq™ sequencing.

2.2.2 Genotyping and quality control

SNP discovery and genotyping were performed by Diversity Arrays Technology using DArTseq™ hybridization-based sequencing technology on next-generation sequencing (NGS) platforms as per (Adamkewicz & Harasewych, 1996; Kilian et al., 2012; Sansaloni et al., 2010). DArTSeq™ sequences were further filtered using custom *dartQC* python scripts (available at https://github.com/esteinig/dartqc). More specifically, SNP sequences matching the following criteria were removed from the dataset (Figure 2.2): (1) average read depth < 7, (2) average repeatability < 90%, (3) call rate < 80%, (4) similar sequence clusters < 0.95, and (5) minor allele frequency (MAF) < 0.02.

All individuals and remaining loci were subsequently tested for missing data and linkage disequilibrium (LD) using *PLINK 1.9* (Purcell et al., 2007). All individuals with high rate of missing data (> 40%) were excluded from the dataset. From the pairs of loci determined to be in LD (with a Pearson coefficient of determination (r^2) threshold of 0.2), SNPs with the lowest call rate value were removed. Lastly, remaining SNP dataset for each population was then assessed for deviation from Hardy–Weinberg equilibrium (HWE) using R package *dartR* (Gruber et al., 2017) and all SNPs that significantly deviated from HWE ($p < 0.0001$) were removed.

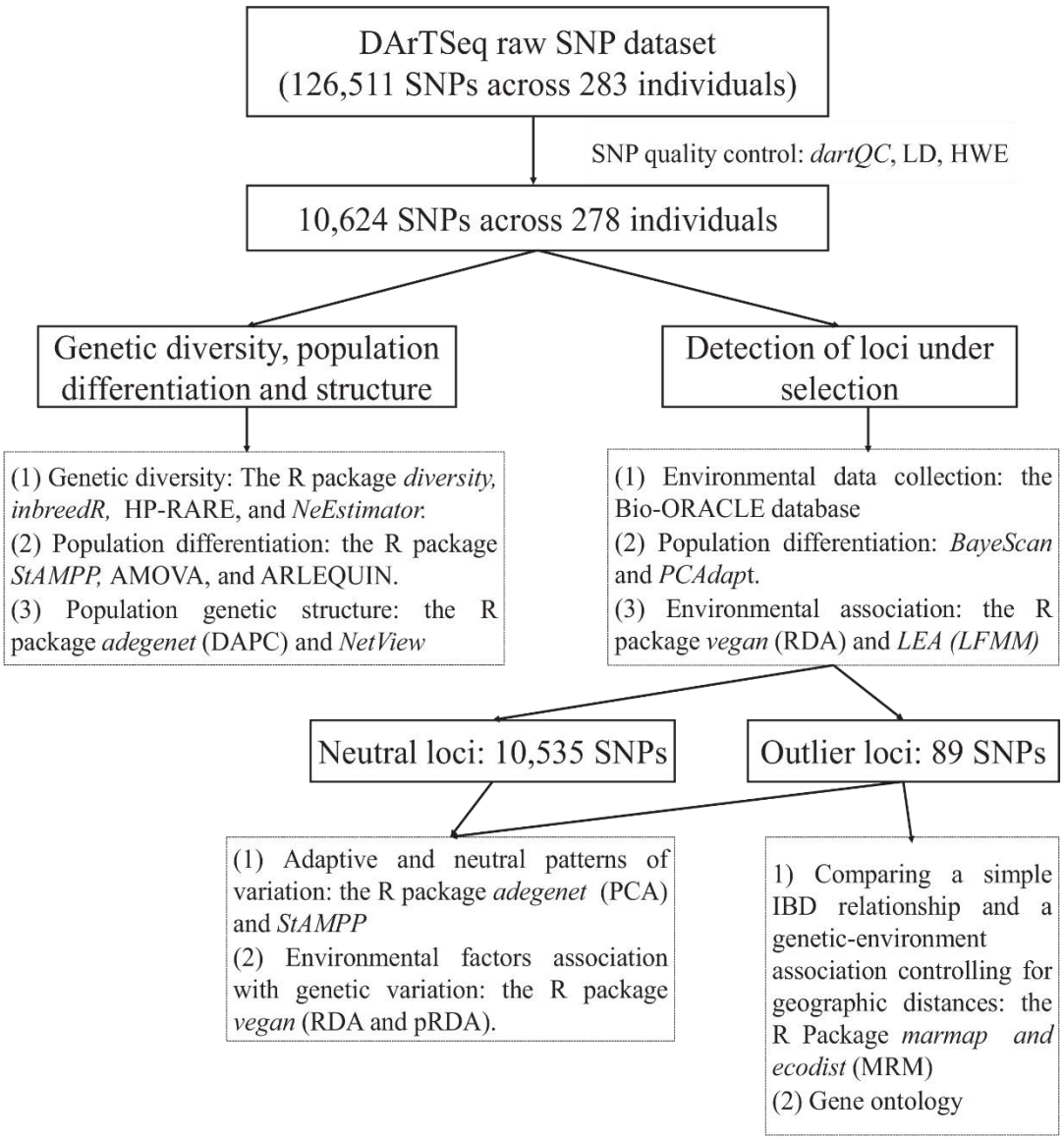


Figure 2.2 SNP data analysis workflow from raw (*n* = 125,511) to final neutral (*n* = 10,535) and outlier (*n* = 89) loci datasets.

2.2.3 Genetic diversity, population differentiation and structure

Several measures were used to estimate population genetic variation and differentiation (Figure 2.2). Observed heterozygosity (H_O), expected heterozygosity (H_E), and inbreeding coefficient (F_{IS}) were calculated using the *divBasic* function of R package *diveRsity* version 1.9.90 (Keenan et al., 2013). F_{IS} values were determined using 95 % confidence intervals and 1000 bootstrap replicates. HP-RARE version 1.1 (Kalinowski, 2005) was used to calculate allelic richness (A_R), private allelic richness (A_{RA}), and percentage of polymorphic loci (PPL) using rarefaction to avoid sampling size bias. Multilocus heterozygosity (MLH) was then calculated

for all individuals from each population using R package *inbreedR* version 0.3.2 (Stoffel et al., 2016)). Effective population size (N_{eLD}) was also estimated for each population using *NeEstimator* version 2.1 based on LD and random mating model (Do et al., 2014).

Genetic differences among populations and individuals were calculated using R package *StAMPP* version 1.5.1 (Pembleton et al., 2013). Pairwise genetic differentiation values (F_{ST}), along with confidence intervals and probability (p) values between populations, were calculated according to Wright (Wright, 1949) and updated by Weir and Cockerham (Weir & Cockerham, 1984). Nei (Nei, 1978) standard genetic distances (Ds) were calculated across loci in a pairwise comparison between populations and individuals. To evaluate hierarchical population genetic differentiation, an analysis of molecular variance (AMOVA) was implemented using ARLEQUIN version 3.5.2.2 (Excoffier & Lischer, 2010). Statistical significance of each variance component was assessed using 1000 permutations for each of the following hierarchical comparisons: 1) among groups (Eastern Australia, Northern Australia, and Western Australia), 2) among populations within groups, and 3) within populations (Table 2.1). To determine isolation by distance (IBD), a Mantel test was conducted for the full SNP dataset (10,624) using the *mantel.randtest* function in R package *ade4* (Chessel et al., 2004) with 999 permutations. Contemporary least-cost oceanographic distances between each pair of sampling sites were estimated using R package *marmap* (Pante & Simon-Bouhet, 2013).

To determine the number of genetic groups a Discriminant Analysis of Principal Components (DAPC) was used on individual genotypes in R package *adegenet* version 2.1.1, with number of possible clusters (cluster values (K) of 1 – 20) determined by running 60 iterations of '*find.clusters*' function (Jombart et al., 2018). Bayesian information criterion (BIC) values, which allow choosing the optimal K value, were averaged across all 60 iterations and standard deviation estimated for each K value. Additionally, genotypic relationships between individuals with no prior population assumptions were assessed using R package *NetView* (Neuditschko et al., 2012; Steinig et al., 2016). *NetView* was run through multiple K-nearest neighbour (kNN) values between 1 and 60 to assess both broad and fine scale population structure.

2.2.4 Environmental data collection and processing

Environmental factors considered ecologically relevant to *P. monodon* (Gribble et al., 2003; Motoh, 1985) were selected from the Bio-ORACLE database (http://www.bio-

oracle.org/downloads-to-email.php) (Assis et al., 2018; Tyberghein et al., 2012). QGIS version 2.18.11 (https://www.qgis.org/) was used to extract values of 23 marine environmental variables (Appendix 2.1) based on latitude and longitude of all seven sampling locations around Northern Australia (Figure 2.1; Table 2.1). Marine data layers were then produced using monthly averages of climate data from 2000 – 2014 for all seven sampling locations (Figure 2.1). Correlation estimates generated for marine data layers and filtered SNP datasets for each population using Pearson Test (R software *cor* and *cor.test* functions) identified seven marine environmental variables with Pearson coefficient of correlation (r) between -0.75 and 0.75. Marine environmental variables retained for further analysis were: 1) surface salinity mean (Env_1), 2) surface temperature maximum (Env_2), 3) surface temperature minimum (Env_3), 4) surface current velocity mean (Env_4), 5) surface phytoplankton mean (Env_5), 6) benthic temperature mean (Env_6), and 7) benthic current velocity mean (Env_7) (Appendix 2.1).

2.2.5 Detection of loci under selection

The possible effects of divergent selection on the overall pattern of genetic differentiation were assessed by combining the results of population differentiation (PD) and environmental association (EA) approaches (Figure 2.2). First, BayeScan version 2.1 (Foll, 2012; Foll & Gaggiotti, 2008) and R package *PCAdap*t (Luu et al., 2017) were used to identify loci under selection (i.e., two independent PD approaches). Second, to identify environmental variables associated with genetic variation, redundancy analysis (RDA) in the R package *vegan* version 2.5-2 (Oksanen et al., 2018; Oksanen et al., 2013) and latent factor mixed models (LFMM) in the R package *LEA* version 2.0.0 (Frichot & François, 2015) were explored (i.e., two independent EA approaches). All four approaches were run on the same SNP dataset using strict parameters (see below). Only candidate outlier SNP identified by all four approaches were retained for subsequent analyses (see below).

BayeScan version 2.1 (Foll, 2012; Foll & Gaggiotti, 2008) was run with 20 pilot runs of 5000 iterations followed by 100,000 iterations and an additional burn-in of 50,000 iterations. Alpha levels and F_{ST} values were ordered from largest to smallest and candidate loci were determined using false discovery rate (FDR) control levels of 0.001, 0.005, 0.01, 0.05, and 0.1 using Bayescan 2.01 function *plot_R.r* (Benjamini & Hochberg, 1995). The R package *PCAdap*t (Luu et al., 2017) with K = 2 and min.maf = 0.05 was used to detect outlier loci based on principal component analysis (PCA) by assuming that markers excessively related with

population structure are candidates for local adaptation. The candidate loci were determined using FDR control levels ranging from 0.01 to 0.1.

The linear model redundancy analysis (RDA) was run in R package *vegan* version 2.5.2 (Oksanen et al., 2013). The *vegan* function *ordistep*, which applies a stepwise per mutational ordination method, was used to perform the 'optimal' model that corresponded to the highest adjusted coefficient of determination (adj R^2). To test the significance of the final RDA model, the *vegan* function *anova.cca* was run with 999 permutations then outlier SNPs were identified on each of the first two constrained axes ($p = 0.001$) using the function *outliers*. Finally, LFMM models were run using R package *LEA* version 2.0.0 (Frichot & François, 2015) for each environmental variable with a total number of 10,000 iterations and a burn-in of 5,000 iterations. The specified result (K = 3) corresponded with the number of population clusters identified in DAPC from five repetitions. Z-scores for each locus were combined across five replicates and FDR were evaluated using FDR-adjusted *p* values. Candidate loci were determined for each marine environmental variable ($n = 7$; see above) and all candidate loci with FDR-adjusted $p \leq 0.05$ were retained for subsequent analyses (see below).

2.2.6 Environmental factors associated with genetic variation

To examine adaptive and neutral patterns of variation within SNPs obtained across all seven collection locations (Figure 2.1), PCA was performed on the retained environmental outlier ($n = 89$) and neutral ($n = 10{,}535$) loci using R package *adegenet* version 2.1.1 (Jombart et al., 2018). Retained outlier and neutral loci were then used to calculate genetic differences among populations and individuals using R package *StAMPP* version 1.5.1 (Pembleton et al., 2013).

To examine the relative contribution of geographical distances and environmental variables and select which environmental factors best explain Australian *P. monodon* genetic structure, redundancy (RDA) and partial redundancy (pRDA) analyses were conducted for both retained outlier and neutral locus allele frequencies using R package *vegan*. Analyses of variance (ANOVA) with 999 permutations were used to assess the significance of each environmental parameter within the RDA for all RDA and partial pRDA tests. The *vegan* function *ordistep* with 999 permutations was used to evaluate each environmental variable and perform the 'optimal' model (i.e., model that best explained genetic structure of each discrete population).

To determine whether isolation by distance (IBD) could explain geographic patterns of differentiation in retained outlier loci (see above), multiple regression on distance matrices (MRM) (Lichstein, 2007) was conducted in R package *ecodist* (Goslee & Urban, 2007) with 1,000 permutations. Pairwise population F_{ST} values calculated in *StAMPP* and least-cost geographic distances calculated in *marmap* were used as response values and explanatory variables, respectively, and then MRM models were generated using the marine environmental variables that significantly correlated with genetic variation of retained outlier loci. Note that MRM models focused on the relationship between geographic distance and environmental variables to determine the significance of environment factors on gene flow among populations (Jeffery et al., 2018).

2.2.7 Gene ontology

Each outlier SNP was assessed against an assembled *P. monodon* transcriptome (Huerlimann et al., 2018) using Geneious Prime version 2019.1.3 (Kearse et al., 2012). Contigs within assembled *P. monodon* transcriptome that exhibited strong pairwise similarity with any outlier SNP (E-value ≤ 1e−5) were compiled and reported along with contig annotations provided by Huerlimann et al. (2018). All SNP that exhibited transcriptome matches were translated into protein sequence using all six reading frames (three and three in 5' – 3' and 3' – 5', respectively) to assess if mutation caused synonymous or non-synonymous change.

2.3 Results

2.3.1 Genotyping and SNP quality control

A total of 126,511 unique SNPs were obtained for each *P. monodon* individual (n = 283) by DArTseq™ sequencing (Figure 2.2). Initial quality control assessment of DArTseq™ data using *dartQC* removed 115,811 SNPs or 91.5% of loci (Table 2.2). The remaining 10,700 loci were then tested for LD and HWE, which removed 76 and 0 SNPs, respectively. Following genotype filtering, five individuals (two and three individuals from Gulf of Carpentaria and Nickol Bay populations, respectively) were removed due to high rates of missing data (> 40%). A final set of 10,624 SNPs for each retained *P. monodon* individual (n = 278) were subjected to further analyses (Figure 2.2 and Table 2.2).

Table 2.2 Filtering steps and SNPs counts retained after each step

Steps	Retained SNP count
Initial potential SNPs	126,511
Average read depth of ≥ 7	126,511
Replication average ≥ 0.9	122,037
Call rate of ≥ 80%	23,845
Similar sequence clusters of ≥ 0.95	19,328
Minor allele frequency (MAF) of ≥ 0.02	10,700
LD filters with a correlation coefficient (r^2) threshold of 0.2	10,624
HWE filters ($p \geq 0.0001$)	10,624
Retained SNPs for further analysis	**10,624**

Linkage disequilibrium (LD) and Hardy–Weinberg equilibrium (HWE)

2.3.2 Population genetic diversity

For the retained 10,624 SNPs, mean allelic richness values (A_R) ranged from 1.61 ± 0.48 for Nickol Bay (Western Australia) population to 1.75 ± 0.34 for Tiwi Island population (Northern Australia) (Table 2.3). A similar pattern was observed for percentage of polymorphic loci (PPL), which ranged from 62% to 87% for Nickol Bay and Tiwi Island populations, respectively. Contrastingly, mean private allelic richness (A_{RA}) and mean MAF of polymorphic loci were slightly higher for Western Australia (0.03 ± 0.14 and 0.18) than for both Eastern Australia (Bramston Beach, Etty Bay, and Townville; 0.004 ± 0.04 and 0.13) and Northern Australia (Gulf of Carpentaria, Joseph Bonaparte Gulf, and Tiwi Island ; 0.03 ± 0.13 and 0.15) populations, respectively.

Mean observed heterozygosity (H_O) across all seven populations ranged from 0.13 ± 0.15 to 0.16 ± 0.2 and mean expected heterozygosity (H_E) ranged from 0.15 ± 0.16 to 0.17 ± 0.17 while average multi-locus heterozygosity (Av. MLH) within populations ranged from 0.13 ± 0.01 to 0.16 ± 0.04 (Table 2.3). Among populations, Western Australia displayed the highest Ho and Av. MLH values (0.16 ± 0.2 and 0.16 ± 0.04, respectively) whereas Eastern Australia showed the lowest levels (0.13 ± 0.15 and 0.13 ± 0.01, respectively). Moreover, mean H_O values were lower compared to mean H_E values in all Eastern Australia and Northern Australia

populations with the exception of Western Australia where mean H_O and H_E values were similar (Table 2.3). Mean F_{IS} values for both Eastern Australia and Northern Australia sites were positive and ranged from 0.10 to 0.14, while mean F_{IS} value for Western Australia population was negative, but close to zero (-0.04). Estimated N_{eLD} varied across populations and ranged from 165.4 in Western Australia (95% CI = 159.9 – 171.2) to 24,121 in Bramston Beach (95% CI = 11,024 - ∞), with Western Australia and Etty Bay populations having the lowest N_{eLD} across all seven populations (Table 2.3).

2.3.3 Population differentiation and genetic structure

DAPC and *NETVIEW* analyses suggested regional structure among Australian *P. monodon* populations, with evidence for three clusters: Eastern Australia group (Bramston Beach, Etty Bay, and Townville), Northern Australia group (Gulf of Carpentaria, Joseph Bonaparte Gulf, and Tiwi Island), and Western Australia (Nickol Bay) (Figure 2.3). The individual density distribution of the first retained discriminant function indicated separation of Eastern Australia shrimp from those in the Northern Australia and Western Australia groups (Figure 1B). Fine-scale population network analysis using *NetView R* provided greater resolution at the individual level between populations and demonstrated the same three genetic clusters at kNN = 60 (Figure 2.3C and 2.3D).

Table 2.3 Genetic diversity indices for the *Penaeus monodon* populations sampled.

Population	N	A_R (± SD)	A_{RA} (± SD)	PPL (Av. MAF)	H_o (± SD)	H_E (± SD)	Av. MLH (± SD)	F_{IS} ($p < 0.05$)	N_{eLD}[95% C.I.]
Bramston Beach (BB)	60	1.66 ± 0.39	0.004 ± 0.04	0.79 (0.13)	0.13 ± 0.15	0.15 ± 0.16	0.13 ± 0.01	0.12	24,121.6 [11,024.4- ∞]
Etty Bay (EB)	50	1.66 ± 0.40	0.01 ± 0.05	0.77 (0.14)	0.13 ± 0.15	0.15 ± 0.16	0.13 ± 0.02	0.14	147.5 [146 – 149]
Townsville (TSV)	22	1.68 ± 0.45	0.03 ± 0.13	0.69 (0.15)	0.13 ± 0.16	0.15 ± 0.17	0.13 ± 0.004	0.10	2,504.6 [1,7897.9 – 4,175.2]
Gulf of Carpentaria (GC)	33	1.74 ± 0.39	0.01 ± 0.07	0.80 (0.15)	0.15 ± 0.15	0.17 ± 0.17	0.15 ± 0.02	0.13	1,427.4 [1,263.1 – 1,640.5]
Joseph Bonaparte Gulf (JBG)	34	1.74 ± 0.38	0.01 ± 0.06	0.81 (0.14)	0.15 ± 0.16	0.17 ± 0.17	0.15 ± 0.004	0.11	7,065.4 [4,454.1 – 15,832]
Tiwi Island (TIW)	56	1.75 ± 0.34	0.01 ± 0.06	0.87 (0.13)	0.15 ± 0.15	0.17 ± 0.16	0.15 ± 0.004	0.12	8,125.2 [5,853.9 – 13,264.8]
Nickol Bay (NKB)	23	1.61 ± 0.48	0.03 ± 0.14	0.62 (0.18)	0.16 ± 0.20	0.16 ± 0.18	0.16 ± 0.04	-0.04	165.4 [159.9 – 171.2]

The parameters calculated number of samples (N), mean allelic richness values (A_R), private allelic richness (A_{RA}), percentage of polymorphic loci (PPL, mean MAF of polymorphic loci), mean observed heterozygosity (H_O), mean expected heterozygosity (H_E), average multi-locus heterozygosity (Av. MLH), inbreeding coefficient (F_{IS}), and effective population size (N_{eLD}).

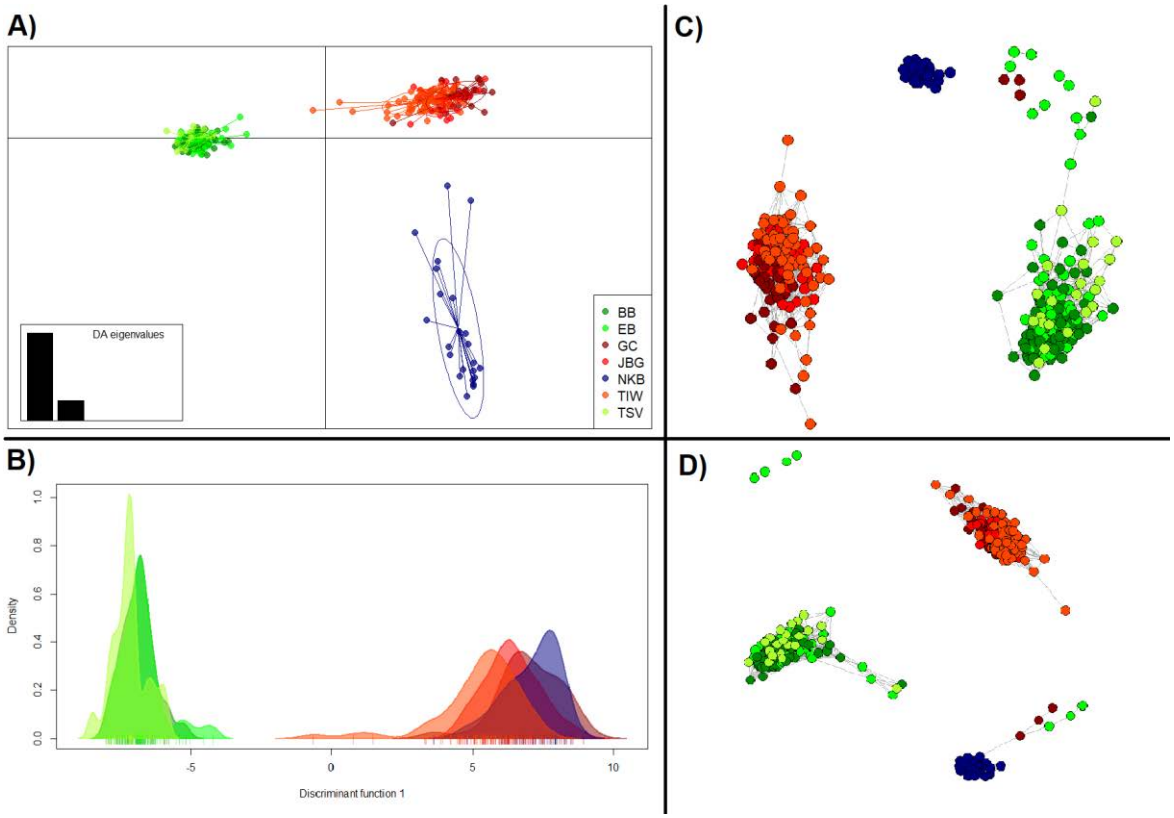


Figure 2.3 Population structure of 278 individuals of *Penaeus monodon* samples using 10,624 SNPs. (**A**) Discriminant Analysis of Principal Components (DAPC) scatterplot and (**B**) an individual density plot on the first discriminant function created through the R package '*adegenet* '; and (**C**) and (**D**) population networks constructed using the *NetView* P v.1.0 pipeline at kNN = 30 and 60. Dots represent individuals. Population names are defined in Figure 2.1.

Across all *P. monodon* populations pairwise F_{ST} values based on unbiased distances (Weir & Cockerham, 1984) ranged from 0.001 to 0.107, while pairwise standard genetic distances (Nei, 1978) similarly ranged from 0.002 to 0.028 (Table 2.4). All pairwise F_{ST} comparisons were significant ($p \leq 0.05$) except for pairwise comparison between Townsville and Etty Bay populations in Eastern Australia. Not surprisingly, the most geographically separated sites (Nickol Bay in Western Australia and three Eastern Australia populations at approximate 3,091 to 3,137 km apart) exhibited the largest significant pairwise F_{ST} (0.107) and standard genetic distance (0.028) values (Table 2.5). Subsequent regression-based analysis also demonstrated a significant linear relationship between pairwise geographic distance and pairwise genetic distance among all sampled populations ($r^2 = 0.86$, $p = 0.006$; Figure 2.4).

Table 2.4 Population differentiation estimates for *Penaeus monodon* populations sampled.

Populations	**BB**	**EB**	**TSV**	**GC**	**JBG**	**TIW**	**NKB**
Bramston Beach (BB)		0.002	0.003	0.010	0.009	0.008	0.026
Etty Bay (EB)	0.001		0.004	0.011	0.010	0.008	0.026
Townsville (TSV)	0.001	**0.001**		0.012	0.011	0.009	0.028
Gulf of Carpentaria (GC)	0.039	0.039	0.037		0.004	0.004	0.020
Joseph Bonaparte Gulf (JBG)	0.034	0.034	0.033	0.002		0.003	0.020
Tiwi Island (TIW)	0.028	0.028	0.026	0.003	0.001		0.019
Nickol Bay (NKB)	0.107	0.107	0.107	0.071	0.069	0.069	

Population pairwise F_{ST} (Weir & Cockerham, 1984) and Nei's genetic distance (Nei, 1972) estimates computed by using the R package StAMPP v.1.5.1 (Pembleton et al., 2013). Pairwise F_{ST} values are shown below the diagonal, Nei's genetic distance are reported above. All F_{ST} values were significant at $p \leq 0.05$ following 1000 bootstraps performed across loci to generate confidence intervals. The only non-significant F_{ST} value ($p > 0.05$) is in bold type.

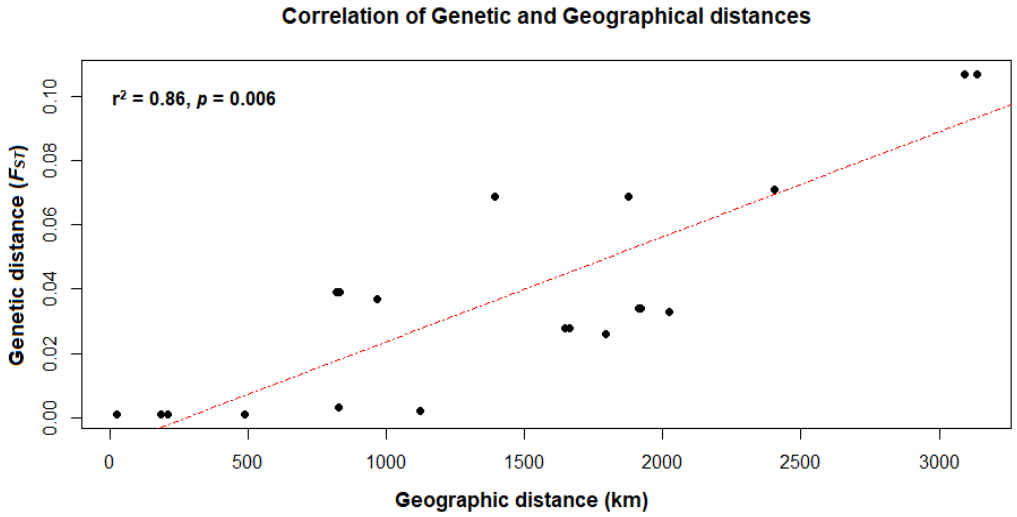


Figure 2.4 Plots of isolation by distance (IBD) with Mantel correlograms for the relationship between genetic distance (F_{ST}) and geographic distance (km) among *Penaeus monodon* populations (n = 7) using 10,624 SNPs.

AMOVA analysis also grouped populations by geographical region: Eastern Australia(Bramston Beach, Etty Bay, and Townsville), Northern Australia (Gulf of Carpentaria, Joseph Bonaparte Gulf, and Tiwi Island), and Western Australia (Nickol Bay). Hierarchical population genetic analyses indicated majority of variation (95.6%) being

explained among individuals within populations ($p < 0.01$; Table 2.5). Differences among geographically discrete populations accounted for only 4.2% of the total variance ($p < 0.01$), while 0.2% of the total variance ($p < 0.01$) was explained by variation among populations within each group.

Table 2.5 Analysis of molecular variance (AMOVA) of *Penaeus monodon* from seven Australian locations (groups were assigned from DAPC analysis (see Figure 1).

Variance partition	d.f.	Sum of squares	Variance component	% of variation	Fixation indices	*p*-value
Among groups	2	6,974	18.84	4.2	FCT: 0.042	< 0.01
Among populations within groups	4	1,999.1	0.9	0.2	FSC: 0.002	< 0.01
Within populations	549	234,808	427.7	95.6	FST: 0.042	< 0.01
Total	555	243,780.7	447.4			

Degrees of freedom (d.f.)

2.3.4 Detection of outlier loci

Initial Bayescan v.2.1 and *PCAdapt* analyses (FDR = 0.01 for both; see Methods) detected 582 and 514 outlier SNPs, respectively, among the 10,624 SNPs retained for each individual (Figure 2.2). LFMM analysis (K = 3 and FDR = 0.05; see Methods) identified 1,449 outlier SNPs that were significantly associated with at least one of seven environmental parameters tested. Full RDA model supported the role of environmental variables in shaping the distribution of SNP genotypes ($p < 0.01$; $R^2 = 0.048$; adjusted $R^2 = 0.031$) with 43.5% and 30.2% of inherent genetic variation explained by the first and second RDA axes, respectively. Based on these two significantly constrained axes the RDA model collectively yielded 963 candidate outlier SNPs. Lastly, the rigorous evaluation of both PD and two EA analyses (see Methods) identified 425 and 161 outlier SNPs respectively, of which 89 (21% and 55%) overlapped (Figure 2.5), leaving 10,535 SNPs as neutral. Both outlier and neutral SNPs were subjected to further analyses (Figure 2.2).

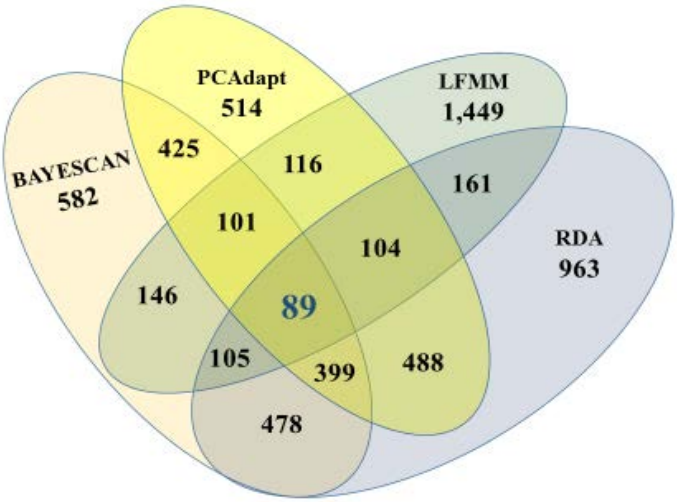


Figure 2.5 Venn diagram showing the total number of putative SNPs significantly associated with at least one environmental variable. The total number of SNPs is reported in each panel. The 89 outlier SNPs on the intersections (shaded area) were retained for further analysis.

2.3.5 Genetic structure based on outlier and neutral markers

PCA analysis of all individuals for both neutral and outlier SNP separated Australian populations into the same three clusters as DAPC and *NetView* analyses (Western Australia separate from Eastern Australia and Northern Australia populations; see above). Considering the 10,535 neutral SNPs, PC1 and PC2 explained 24.8% and 14.7% of the total genetic variance, respectively (Figure 2.6A). Considering the 89 outlier SNPs, PC1 and PC2 explained 5.8% and 1.0% of the total genetic variance, respectively (Figure 2.6B).

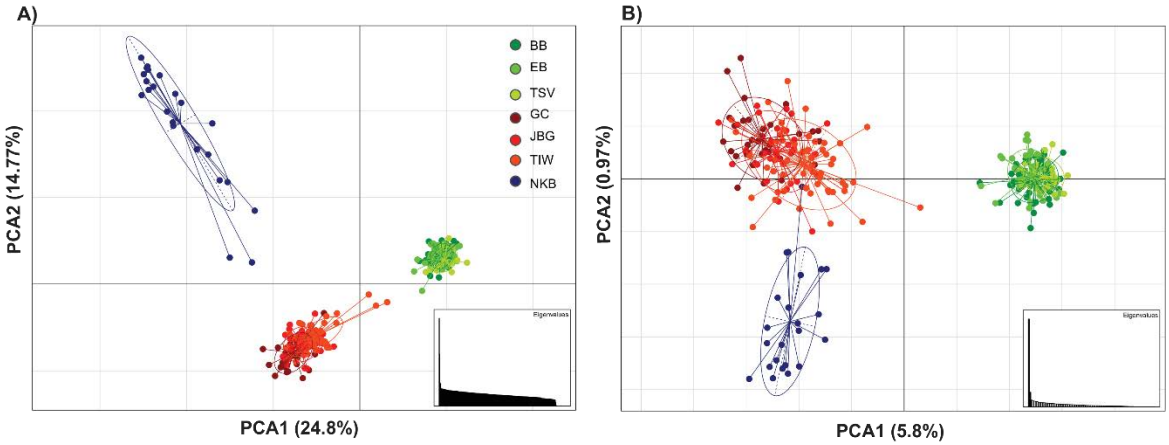


Figure 2.6 Principal components analysis based allele frequencies for (**A**) 10,535 neutral SNPs loci and (**B**) 89 outlier SNPs loci. Population names are defined in Figure 2.1

Pairwise F_{ST} estimates between populations differed depending on assessment of neutral or outlier loci (Table 2.6). Genetic differentiation was significant (p <0.05) for all pairwise comparisons when 10,535 neutral SNPs were assessed. Pairwise F_{ST} values ranged from 0.001 (Bramston Beach vs. Etty Bay and Joseph Bonaparte Gulf vs. Tiwi Island) to 0.1 (Eastern Australia populations vs. Western Australia population). In a similar way, using 89 outlier loci, F_{ST} values between Eastern Australia populations and Western Australia (F_{ST} = 0.60 - 0.64) were in all cases higher than those between Northern Australia populations and Western Australia (F_{ST} = 0.19 - 0.22; $p \leq 0.01$). F_{ST} values differed significantly between all populations within the Northern Australia ($p \leq 0.04$); however, no F_{ST} values differed significantly between populations within Eastern Australia ($p > 0.4$).

Table 2.6 Population differentiation estimates for *Penaeus monodon* populations sampled using neutral (n = 963) and outlier loci (n = 89).

	BB	**EB**	**TSV**	**GC**	**JBG**	**TIW**	**NKB**
Bramston Beach (BB)	0	0.001	0.001	0.03	0.03	0.022	0.10
Etty Bay (EB)	**-0.001**	0	0.001	0.03	0.03	0.02	0.10
Townsville (TSV)	**0.002**	**-0.005**	0	0.03	0.02	0.02	0.10
Gulf of Carpentaria (GC)	0.49	0.47	0.43	0	0.002	0.002	0.07
Joseph Bonaparte Gulf (JBG)	0.46	0.44	0.40	0.01	0	0.001	0.07
Tiwi Island (TIW)	0.38	0.36	0.33	0.03	0.02	0	0.07
Nickol Bay (NKB)	0.64	0.63	0.60	0.22	0.21	0.19	0

Population pairwise F_{ST} (Weir & Cockerham, 1984) estimates computed by using the R package *StAMPP* (Pembleton et al., 2013). F_{ST} values of 89 outlier loci are shown below the diagonal and F_{ST} values of 963 neutral loci are reported above. All F_{ST} values were significant at $p \leq 0.05$ following 1000 bootstraps performed across loci to generate confidence intervals. The only non-significant F_{ST} value ($p > 0.05$) is in bold type.

2.3.6 Environmental factors associated with genetic variation

ANOVA for RDA analyses demonstrated that five (Env_1, Env_2, Env_3, Env_5, and Env_7) and four (Env_2, Env_3, Env_5, Env_7) environmental factors (see Methods) were significantly associated with genetic variation in identified neutral and outlier SNP datasets (p = 0.001 and 0.001, adjusted R^2 = 0.03 and 0.2), respectively (Figure 2.7). When partitioning the relative importance of geographic proximity among sampled sites using partial RDA, the most important environmental predictors for neutral SNP were ranked as: Env_2 (F = 1.7, p =

0.001) > Env_1 (F = 1.5, p = 0.001) > Env_3 (F = 1.2, p = 0.001). For outlier SNP, Env_2 and Env_3 were both significant for putative adaptive genetic variation (p = 0.001); however, Env_2 explained 13% of inherent genetic variation while Env_3 explained 2.8%. When both neutral and outlier SNP datasets were combined, surface temperature max and min (Env_2 and Env_3 respectively) were the highest ranked predictors contributing to the genetic variation of Australian *P. monodon*. Variance partitioning based on neutral and outlier loci RDA models revealed that environmental effects explained 4.3% and 20.2% of genetic structure variation, while geographic location explained 12.9% and 2.8%, respectively. The remaining genetic structure variance was explained by the combination of environmental effect and geographic location.

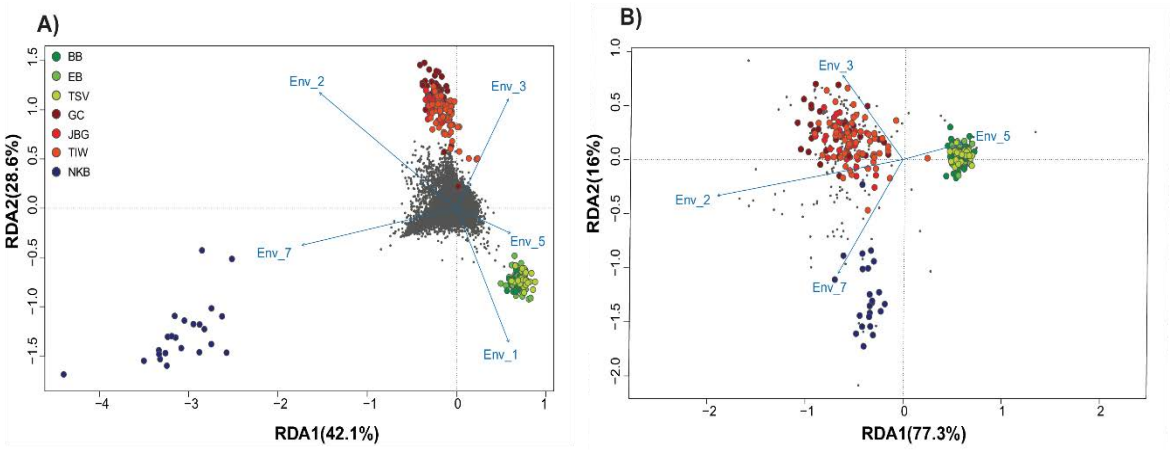


Figure 2.7 Redundancy analysis on (A) 10,535 neutral loci and (B) 89 outlier SNPs allele frequencies in seven populations of *Penaeus monodon* in Australia. Explanatory variables (arrows) were RDA axes retained as important variable selection accounting for genetic variation (Env_2: surface temperature maximum, Env_3: surface temperature minimum, Env_5: surface phytoplankton mean, Env_7: benthic current velocity mean, and Env_1: surface salinity mean).

MRM correlation analysis conducted on geographic distance and pairwise population F_{ST} values was significant (r^2 = 0.9, p = 0.008). MRM models including surface temperature maximum (Env_2) and surface temperature minimum (Env_3) exhibited significant relationships with pairwise F_{ST} values (p = 0.001 and 0.008, adjusted R^2 = 0.95 and 0.90), respectively (Table 2.7). Moreover, MRM models explained a large and significant proportion of genetic variability (> 90%), indicating that these environmental variables and geographic distance are significantly correlated and thus potentially constrain gene flow synergistically.

Table 2.7 Results of multiple regression of distance matrices (MRM) tests of geographic distance, genetic differentiation (F_{ST}) of 89 outlier loci, and significant environmental variables between *Penaeus monodon* populations in Australia.

MRM models	Adj-R^2	*p*-value
Distances ~ F_{ST} + surface temperature max	0.95	0.001
Distances ~ F_{ST} + surface temperature min	0.9	0.008
Distances ~ F_{ST} + surface temperature max + surface temperature min	0.95	0.002

2.3.7 Gene ontology

Among the 89 SNPs that were found associated with environmental variables, 11 (12.4%) yielded matches to 173 total contigs in the *P. monodon* transcriptome with E value ≤ 1e-5 (Appendix 2.2). Three of these 11 outlier SNPs exhibited BLAST hits in only one allele (ID: PM_10591, PM2065, and PM10621; Appendix 2.2) and were thus excluded from translated protein comparison. Protein translation of the remaining eight outlier SNPs demonstrated the following: 1) five contained synonymous and non-synonymous mutations in two and four reading frames, 2) two contained synonymous and non-synonymous mutations in one and five reading frames, and 3) one contained synonymous and non-synonymous mutations in zero and six reading frames, respectively (Appendix 2.3). However, only two of these eight outlier SNPs exhibited 100% pairwise identity with > 90% query cover to contigs within the *P. monodon* transcriptome (PM1958 and PM4714). More specifically, SNP PM1958 matched "calcium-activated chloride channel regulator", "calcium-activated chloride channel regulator 2-like", and "epithelial chloride channel-like" contigs while SNP PM4714 matched "uncharacterized protein APZ42_034504" or "Predicted: uncharacterized protein LOC108744461 isoform X4" contigs (Table 2.8). The remaining majority of outlier SNP (*n* = 78 or 87.6%) did not exhibit sufficiently strong matches to any contigs within the *P. monodon* transcriptome and were thus concluded to reside in non-coding (i.e., absent) genomic regions.

Table 2.8 Characterization of high-quality BLAST matches obtained in comparison with the transcriptome *Penaeus monodon* database

	PM1958	PM4714
Functional Annotation	Calcium-activated chloride channel regulator Calcium-activated chloride channel regulator 2-like Epithelial chloride channel -like	PREDICTED: uncharacterized protein LOC108744461 isoform X4 Uncharacterized protein APZ42_034504
E-value	2.01E-29	3.36E-27
% Identity	100	94.2
Bit-Score	128.5	121.2
Translation Frame 1	CRRVRRLPVPLPPRLQPRGAAPI CRRVRRLPIPLPPRLQPRGAAPI	CST*GTASWTGTRGRTCSTRSRD CSTSGTASWTGTRGRTCSTRSRD
Translation Frame 2	ADAYAAYPYPYHPGYSHEELPP ADAYAAYPYPYHPGYSHEELPP	AVHKAQRAGRALAAALVAHEAE AVHQAQRAGRALAAALVAHEAE
Translation Frame 3	QTRTPPTRTPTTQATATRSCPH QTRTPPTHTPTTQATATRSCPH	QYIRHSELDGHSRPHL*HTKPR QYIRHSELDGHSRPHL*HTKPR

Red characters indicate SNP location within translated protein sequences. Asterisks (*) indicate stop codons.

2.4 Discussion

The giant black tiger shrimp (*P. monodon*) is an important aquaculture species in Australia; however, to date only limited information is available regarding the influence of environmental pressures on the genetic structure of geographically discrete populations. Based on results from this study, three distinct genetic groups (i.e., stocks) were revealed across the geographic distribution range of Australian *P. monodon*: Eastern Australia (Bramston Beach, Etty Bay, and Townville), Northern Australia (Gulf of Carpentaria, Joseph Bonaparte Gulf, and Tiwi Island), and Western Australia (Nickol Bay). Of note is that the Western Australia population exhibited the lowest level of genetic variation of all assessed populations, which could be due to restricted gene flow between Western Australia and Northern Australia populations. Using multivariate analyses that considered both geographic distance and environmental factors it was determined that the majority of SNPs are neutral (n = 10,535), while a small subset of outliers are putatively adaptive (n = 89). Surface temperature maximum and minimum provided the strongest correlative explanation for the presence of these outliers (see below).

2.4.1 Population genetic diversity

Genetic diversity and stock assessments of Australian *P. monodon* were undertaken using all 10,624 SNP (i.e., neutral and outliers combined). Relative reduced genetic diversity (H_O, H_E, and Av.MLH) was observed for Australian *P. monodon* compared to other recently assessed crustacean species. More specifically, European green crab (*Carcinus maenas*) exhibited mean H_O and H_E of 0.254 and 0.256 (Jeffery et al., 2017) while European lobster (*Homarus gammarus*) had H_O and H_E that ranged between 0.049 – 0.63 and 0.179 - 0.504 (Jenkins et al., 2018), respectively and scalloped spiny lobster (*Panulirus homarus*) had H_O, H_E, and Av.MLH that ranged between 0.166 – 0.184, 0.226 – 0.233, and 0.168 – 0.186, respectively (Al-Breiki et al., 2018). The small reduction in genetic diversity observed for Australian *P. monodon* could be due to technical artefacts of the RADseq-based genotyping (i.e., null alleles) (Andrews & Luikart, 2014; Lal et al., 2016), sampling bias (i.e., Wahlund effect) (De Meeûs, 2017; Lal et al., 2018), and/or biological significance (i.e., colonization, bottleneck, or genetic drift) (Benzie et al., 2002; Benzie et al., 1992). H_O deficiency has also been observed in Bangladesh *P. monodon* using 10 microsatellite markers and 14 SNP loci (Alam & Pálsson, 2016), so the exact cause is yet to be fully resolved and both technical and biological could contribute to the small reduction in genetic diversity observed for Australian *P. monodon*.

In the Western Australia (Nickol Bay) population, heterozygosity (i.e., H_O and H_E) were equal or higher than other populations despite A_R and PPL being slightly lower (0.01 – 0.02) (see Results). This slight reduction in A_R and PPL is most likely due to the smaller sample size in my study for the Western Australia population (Kalinowski, 2004). SNP genotyping of this Western Australia population revealed equal or higher heterozygosity, despite previous allozyme and microsatellite-based demonstrations of reduced heterozygosity and number of alleles in *P. monodon* from the same Western Australia region (Benzie, 2000; Benzie et al., 1992). These previous *P. monodon* population genetics studies concluded that observed reductions in heterozygosity and number of alleles were most likely driven by founder effects or bottleneck events. As such, based on the entire Western Australia SNP dataset encompassing 10,624 genome-wide SNPs, Western Australia *P. monodon* appears from my dataset to not have undergone a bottleneck (i.e., heterozygosity not reduced), but rather has retained similar heterozygosity to Northern Australia and Eastern Australia populations. Differences between my dataset and those using microsatellites, allozymes and mtDNA may simply be due to the

level of genetic resolution of these markers when comparing allelic diversity, where my high-resolution sampling captured more of the true genetic diversity within the *P. monodon* genome. Likewise, Lemopoulos et al. (2019) demonstrated that SNPs were more informative than microsatellites for applications that required individual-level genotype information (i.e., estimating relatedness and genetic diversity with low sample sizes from small populations). Although Eastern Australia and Northern Australia *P. monodon* populations exhibited F_{IS} values > 10%, the large N_{eLD} estimates were determined for all Australian *P. monodon* populations. Consequently, the possible effect of inbreeding could be reduced (Frankham et al., 2002). Slightly positive F_{IS} values observed here are most likely a result of DArTSeq genotyping technical artifacts (i.e., null alleles) rather than of biological origin. It is also noteworthy, that F_{IS} values in this study are lower than observed in wild Pacific Ocean *L. vannamei* collected from Panama to Mexico (F_{IS}= 0.53) (Valles-Jimenez et al., 2004) and along the Mexican coast (F_{IS}= 0.36) (Vela-Avitúa et al., 2013).

2.4.2 Population differentiation and genetic structure

Pairwise F_{ST} analysis for Eastern and Northern Australia *P. monodon* populations revealed relatively weak genetic structuring (0 – 0.039), except for Western Australia (0.069 – 0.107), which had the highest levels detected (Table 3). Additionally, genetic differentiation increased proportionally with geographic distance and followed the IBD pattern across the entire sampled range, which is in agreement with previous observations based on mtDNA and microsatellite sequences (Benzie et al., 2002; Benzie et al., 1992; Brooker et al., 2000). However, this finding contrasts the F_{ST} values observed between Western Australia and Northern Australia (0.116) and Eastern Australia (0.032) (Waqairatu et al., 2012). As such, the significantly different genetic composition between Western Australia and all other populations may reflect the effects of restricted gene flow and genetic drift, which is not surprising for *P. monodon* given the relatively short offshore planktonic larval phase (approximately 20 days), during which larvae and post-larvae disperse and migrate to nursery areas (e.g., in-shore areas and mangrove estuaries) (Motoh, 1985). Moreover, the evidence for adaptive divergence among Australian *P. monodon* populations presented here (see below) suggests that divergent selection may be contributing to genetic divergence despite genetic drift (Funk et al., 2016). Regardless of the exact cause, these results suggest that there is a restriction in gene flow between geographically disparate Australian *P. monodon* populations.

Assessment of population structure using F_{ST}, DAPC and *NetView* analyses revealed regional structure in Australian *P. monodon* with three major population groupings: Eastern Australia (Bramston Beach, Etty Bay, and Townsville), Northern Australia (Gulf of Carpentaria, Joseph Bonaparte Gulf, and Tiwi Island), and Western Australia (Nickol Bay) (Figure 1). One possible explanation for this genetic structure is the presence of biogeographic barriers between Eastern, Northern, and Western Australia that caused restricted gene flow between *P. monodon* populations. Upwelling of deep cold water along the North-Western Australian coastline, which has existed since the Late Miocene period, is one biogeographical barrier known to prevent gene flow between Western Australia and other Australian regions (Williams & Benzie, 1998). Moreover, repeated declines in sea surface levels of 100 – 140 m during the Pleistocene caused biogeographic isolation between Northern and Eastern Australia due to repeated emergence of land bridges between Northern Australia, Torres Strait, and New Guinea (Chivas et al., 2001; Galloway, 1981; Sloss et al., 2018; Voris, 2000). Similar patterns in genetic structure between Australian populations have been observed for other marine species such as Australian Spanish mackerel (*Scomberomorus commerson*; Buckworth et al. (1998)), mud crab (*Scylla serrata*; Gopurenko and Hughes (2002)), brown tiger prawn (*Penaeus esculentus;* Ward et al. (2006)), and barramundi (*Lates calcarifer*; Chenoweth et al. (1998); Loughnan et al. (2019)). Previous *P. monodon* genetic differentiation investigations based on allozyme, mitochondrial DNA and microsatellite markers also revealed significant genetic partitioning among Australian and Southeast Asian *P. monodon* populations indicating significant bio-geographic barriers to dispersal (Benzie et al., 2002; Benzie et al., 1992; Brooker et al., 2000; Waqairatu et al., 2012). Accordingly, the current observed genetic differentiation among Australian *P. monodon* populations appears to be driven by the presence of biogeographic barriers between Eastern, Northern, and Western Australia that effectively limited or prevented gene flow over evolutionary timescales.

2.4.3 Evidence for local adaptation

PD and EA analyses collectively identified 89 overlapping outlier SNPs, which accounted for 21% and 55% of total outlier SNPs detected, respectively (Figure 2.5). This observation is in agreement with previous studies that also found EA approaches to perform better than PD approaches for detection of loci under putative divergent selection (Benestan et al., 2016; Dalongeville et al., 2018). Moreover, the observed levels of genetic structure based on these

89 outlier SNPs agreed with the genetic structure observed when all 10,535 neutral SNPs were considered.

Analysis of population structure based on 89 outlier SNPs using PCA presented a similar spatial pattern as to PCA based on 10,535 neutral SNPs in the form of three groups: Eastern Australia, Northern Australia, and Western Australia (Nickol Bay). Several causes may contribute to the same population structure being determined when either neutral or outlier loci were used (e.g., life history, natural selection, and environmental heterogeneity (Forester et al., 2016; Moore et al., 2014; Sexton et al., 2014)); however, determination of the exact cause requires further investigation. My findings agree with previous SNP-based studies conducted on other species with a planktonic larval phase (e.g., Eulachon (*Thaleichthys pacificus*), European green crab (*Carcinus maenas*), and Atlantic salmon (*Salmo salar*)) that were determined to have low level gene flow and genome-wide divergence (Candy et al., 2015; Jeffery et al., 2017; Moore et al., 2014), respectively. These outlier-based analyses suggest that key environmental factors (see Results) and geographic distance, synergistically or independently, contributed to the generation of the 89 outlier SNPs observed among Australian *P. monodon* populations (i.e., adaptive divergence drivers). This conclusion is supported by other studies that also demonstrated that association between environmental variables and outlier SNPs could be indicative of local adaptation (Benestan et al., 2016; Bernatchez et al., 2018; Jeffery et al., 2018; Wyngaarden et al., 2018; Xuereb et al., 2018).

2.4.4 Environmental variables contributing to genetic structure

RAD and pRAD analyses demonstrated a strong relationship between geographic distance, environmental distance, and genetic differentiation in both neutral (n = 10,535) and outlier (n = 89) SNPs inherent to wild Australian *P. monodon*. This significant positive correlation suggests that: 1) gene flow between sampled sites may be impacted by natural selection or 2) geographically discrete Australian *P. monodon* populations have adapted to region-specific environmental variables (e.g., temperature). Geographic isolation due to biogeographic barriers (see above) could have caused initial level of genetic differentiation between *P. monodon* populations with subsequent divergent selection imposed by environmental factors leading to increased differentiation across evolutionary time. This interpretation is supported by RDA analysis, which indicated that the effects of surface temperature maximum and minimum

(Env_2 and Env_3, respectively) were more pronounced than geographic distance for both neutral and outlier SNP.

RAD and partial RAD analyses conducted on outlier SNP demonstrated that surface temperature maximum (Env_2) explained five and six times more genetic variation than was explained by surface temperature minimum (Env_3). Therefore, it is plausible that surface temperature maximum (Env_2) could be the predominant driver of local adaption among Australian *P. monodon* populations. In a review of local adaptation among populations of marine invertebrates, Sanford & Kelly (2011) found that approximately 44% of surveyed studies provided evidence of local adaptation to temperature. Moreover, several studies targeting marine invertebrates demonstrated the role of sea surface temperature in shaping genetic differences among populations (e.g., American lobster (*H. americanus*), European green crab (*C. maenas*), eastern oyster (*C. virginica*), sea scallop (*Placopecten magellanicus*), and giant California sea cucumber (*Parastichopus californicus*);Benestan et al. (2016), Jeffery et al. (2018), Bernatchez et al. (2018), Wyngaarden et al. (2018), and Xuereb et al. (2018), respectively). Winter sea surface temperature was also demonstrated to be the most likely driver of local adaptation and limiter of gene flow among North American populations of *C. maenas* (Jeffery et al., 2018). In Australia, sea surface temperatures have become significantly warmer between 1950 - 2007 along North-Eastern and North-Western tropical coasts by 0.12˚C and 0.11˚C per decade, respectively (Lough, 2008), while sea surface temperature in south-western Australia has risen by 0.026 to 0.034˚C per year between 1985–2004 (Pearce & Feng, 2007). Moreover, surface temperature maximum and minimums occur during summer and winter, respectively, which is when *P. monodon* broodstock emigrate out of estuaries into foreshores for spawning and, thus, extreme swings could influence individual fitness (i.e., breeding success). Further studies are needed to elucidate the existence and extent of population-specific thermal adaptation among Australia *P. monodon* populations and the potential functional genomics implications of the identified 89 outlier SNPs.

2.4.5 Highlights from Gene Ontology

The combination of PD and EA approaches identified 89 outlier SNPs that are potential targets for local adaptation within the Australian *P. monodon* genome. Only 11 of these 89 outlier SNPs matched *P. monodon* transcriptome contigs (Huerlimann et al., 2018) with subsequent protein translation of eight outlier SNPs demonstrating more non-synonymous changes than

synonymous changes across all six reading frames. Despite the greater likelihood that non-synonymous mutations will have functional implications (Lawrie et al., 2013; Parmley et al., 2005), all 11 of these outlier SNPs with *P. monodon* transcriptome matches should be considered as candidates for future research into divergent selection driven local adaptation among these geographically discrete populations (Lebeuf-Taylor et al., 2019).

Of these 11 outlier SNPs, one matched a contig annotated as "calcium-activated chloride channel regulator" (*CLCA*) (Table 2.8). *CLCA* is involved in cellular physiology functions such as neuronal and cardiac action, muscle contraction, and epithelial secretion (Dingyuan Hu et al., 2018; Nyström et al., 2018; Zhang et al., 2014). In Pacific white shrimp *L. vannamei*, a similar calcium-activated chloride channel gene was shown to be expressed in gill cells and exhibit potential involvement in osmoregulation because of observed response to salinity challenge (Dongxu Hu et al., 2015). As such, future *P. monodon* gene-by-environment studies should consider investigating the potential role of this gene in local adaptation to population-specific environmental conditions.

The remaining 78 outlier SNPs are presumed to be located in non-coding regions absent from the transcriptome, or in genes with low-level expression that were not captured within initial transcriptome assembly (Harrisson et al., 2017; Harrisson et al., 2014; Tigano et al., 2017).

2.4.6 Implications for aquaculture management and future research directions

The black tiger shrimp is an important aquaculture and fishery species in Australia. Therefore, the identification of genetic stock units among wild populations is crucial for fishery and aquaculture broodstock management. Neutral and adaptive population structure findings suggested that Australian *P. monodon* should be managed as three separate stocks: Eastern Australia, Northern Australia, and Western Australia. The levels of genetic diversity revealed in the present study using both neutral and outlier SNP are useful for aquaculture purposes such as selective breeding programs, maintaining stock diversity, and distinguishing hatchery stocks from wild populations. Moreover, in addition to reduced gene flow, thermal adaptation is also likely to contribute to a greater divergence in the Western Australia population. Regardless, the strong genetic structure and presence of rare alleles in all seven *P. monodon* populations suggests that these populations should be managed accordingly to maintain genetic integrity. Environmental factors can influence genetic diversity and population structure in marine

species and an accurate understanding of this influence can both improve fisheries management and help predict responses to environmental change (Wyngaarden et al., 2018). This study suggests that both geographical distance and environmental factors interact to influence the genetic structure of Australian *P. monodon*; however, the magnitude of influence for each of these factors is hard to determine conclusively. This study also provides evidence for environmentally driven selection pressure on geographically discrete populations, which can be utilized to help ensure sustainable management of Australian *P. monodon* (e.g., guidance for future re-establishment of populations inhabiting similar thermal gradients). Practically, a population that is potentially under local adaptive pressures may be an important source of private or rare alleles that can enhance population resistance to future environmental change (e.g., naturally or via selective breeding programs) or assist natural migration (Golbuu et al., 2016; Ofori et al., 2017).

2.5 Conclusion

I utilized a SNP dataset containing 10,624 loci to determine genetic population structure and local adaptation across seven populations of Australian black tiger prawn *P. monodon* ($n = 278$ individuals). This study provides novel insights that assist the development and implementation of *P. monodon* aquaculture and fishery management practices within Australia. Analysis of population structure using both neutral ($n = 10,535$) and outlier ($n = 89$) SNPs suggest that Australian *P. monodon* should be managed as three separate stocks (Eastern Australia, Northern Australia, and Western Australia) and that geographically discrete *P. monodon* populations have likely undergone local adaptation to region-specific thermal regimes. Future studies should investigate the role that outlier SNPs potentially play in local adaptation in order to advance wild stock structure preservation and help facilitate selective breeding programs.

Chapter 3. Population structure, genetic connectivity, and signatures of local adaptation of the giant black tiger shrimp (*Penaeus monodon*) throughout the Indo–Pacific region

3.1 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 (Hemmer-Hansen et al., 2007; Matolweni et al., 2000). 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). Population genetics research also enhances the facilitation of appropriate fisheries management and aquaculture practices by providing novel insights into broodstock selection for breeding programs and discernment of hatchery stocks from wild populations (Bernatchez et al., 2017; Gjedrem, 2012; Lorenzen et al., 2012; Williams & Hoffman, 2009). In order to make accurate recommendations regarding aquaculture programs or fishery management plans for commercially important species, it is of particular importance to comprehensively understand the genetic connectivity among Indo-Pacific populations as it relates to their evolutionary and life history (e.g., demographics and connectivity).

The giant black tiger shrimp (*P. 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 migration to nursery areas, such as in-shore areas and mangrove estuaries (Motoh, 1985; Queiroga & 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 to 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; Waqairatu et al., 2012; You et al., 2008). These genetic markers can often have limited resolution power to detect fine-scale genetic structure and signatures of adaptation within and among populations (Allendorf et al., 2010; Nielsen et al., 2009). Despite limited resolution these previous studies demonstrated low-level of population differentiation 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 (Waqairatu et al., 2012; You et al., 2008).

The recent development of high-throughput sequencing techniques permits determination of both genome-wide neutral and adaptive genetic variation for non-model 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), while 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 (Anderson et al., 2019; Gagnaire et al., 2015; Nowland et al., 2019; Woodings et al., 2018).

Collectively, the high-resolution power of genome-wide single nucleotide polymorphisms (SNPs) genotyping offers the potential for heightened understanding of *P. monodon* population genetic structure, connectivity, and local adaptation that 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. More specifically, this study aimed to i) evaluate genetic diversity and determine the spatial population structure among *P. monodon* populations using a combined (i.e., neutral and outlier loci) dataset, ii) determine the genetic connectivity and evolutionary relationship among populations, and iii) identify putatively adaptive outlier SNP (i.e., candidate genes).

3.2 Material and Methods

3.2.1 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 southwestern coast of the African continent. In order to study the genetic relationships across the entire Indo-Pacific range I collected wild adult *P. monodon* (n = 532 total) from 16 discrete Indo-Pacific locations between 2015 and 2019 (Figure 3.1; Table 3.1). Pleopod tissue samples were excised from each wild adult *P. monodon* sample with scissors, preserved in ethanol, and stored at -20 ˚C until DNA extraction.

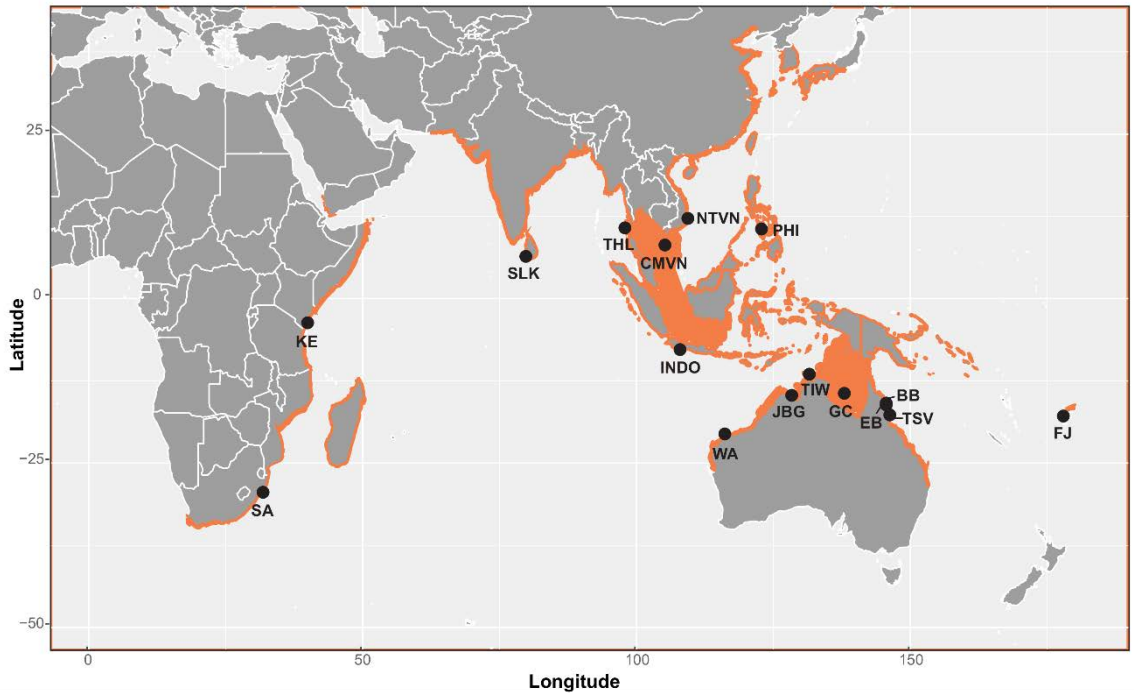


Figure 3.1 Sampling sites and the known distribution of *Penaeus monodon* (modified from (Kongkeo, 2005–2020)) sourced in this study. Orange areas denote the know distribution of *P. monodon*. See Schott et al. (2009) for detailed Indo-Pacific summer and winter monsoon currents maps. FJ: Fiji; Eastern 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.

Genomic DNA (gDNA) was extracted from each pleopod sample following a modified cetyl-trimethyl ammonium bromide (CTAB) protocol (Adamkewicz & Harasewych, 1996; Chapter 2). All gDNA extractions were subsequently cleaned using Sephadex™ G-50 (GE, 2007) and gDNA integrity was confirmed using 0.8% agarose gel containing GelGreen (ThermoFisher Scientific Pty Ltd, Australia). All gDNA samples were normalised to 50 ng/µl (20 µL final volume) before submission to Diversity Arrays Technology (DArTseq™; Canberra Australia) for library preparation and DArTseq™ sequencing.

Table 3.1 Fishing area, sampling location, population ID, code, number of individuals (*N*), and coordinates for each population of *Penaeus monodon* sampled.

Fishing Area	Regions	Sampling locations	Code	N	Latitude	Longitude
Oceania	Eastern Australia (EA)	Bramston Beach	BB	51	-17.561792	146.108953
		Etty Bay	EB	31	-17.353008	146.052726
		Townsville	TSV	22	-19.082107	146.817537
	Northern Australia (NA)	Gulf of Carpentaria	GC	30	-14.997655	138.780787
		Joseph Bonaparte Gulf	JBG	34	-14.681809	128.371358
		Tiwi Island	TIW	50	-11.866834	131.83147
	Western Australia (WA)	Nickol Bay	WA	23	-20.697014	116.863229
	Fiji	Fiji	FJ	49	-17.957068	178.608177
Southeast Asia	Philippines	Philippines	PHI	13	10.546042	122.447387
	Vietnam	Ca Mau	CMVN	39	8.5676	105.0526
		Nha Trang	NTVN	30	12.2608	109.2383
	Indonesia	Java	INDO	38	-7.702932	108.706541
	Thailand	Thailand	THL	46	11.693318	97.657631
					11.937713	101.465872
Northwest Indian Ocean (NWIO)	Sri Lanka	Sri Lanka	SLK	20	6.434407	79.974896
Western Indian Ocean (WIO)	Kenya	Kenya	KE	25	-3.583314	40.0142664
	South Africa	South Africa	SA	31	-28.978398	31.836385

3.2.2 SNP genotyping and quality control

All shrimp samples were sequenced and genotyped using DArTseq™ technology, as previously described (Anderson et al., 2019; Kilian et al., 2012; Chapter 2). Briefly, DArTSeq™ technology relies on a complexity reduction method in order to obtain genome sequences using next-generation sequencing technology. DArTSeq 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 °C for 1 min, then 30 cycles of 94 °C for 20 s, 58 °C for 30 s, 72 °C for 45 s and 72 °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 was obtained from DArT after initial SNP calling but without filtering by proprietary analysis pipelines (DArTsoft). SNP data was then initially filtered using custom *dartQC* python scripts (available at https://github.com/esteinig/dartqc). More specifically, SNPs matching the following stringent criteria were excluded from the dataset: 1) low read depth (--read_counts < 7), 2) average repeatability < 90%, 3) call rate < 80%, 4) duplicated SNPs base on similar sequence clusters (-- cluster > 0.95), and 5) minor allele frequency (MAF) < 0.02. Furthermore, all remaining loci and individuals were tested for linkage disequilibrium (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 Hardy–Weinberg equilibrium (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 SNPs remaining for population genetic analyses.

3.2.3 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 *diveRsity* 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 percentage of polymorphic loci (PPL) using rarefaction to avoid sampling size bias. Average multi-locus heterozygosity (Av. MLH) and standardized multilocus heterozygosity (sMLH) were calculated for all individuals using R package *inbreedR* version 0.3.2 (Stoffel et al., 2016) while internal relatedness (IR) was calculated for each population using the R package *Rhh* version 1.0.1 (Alho et al., 2010).

3.2.4 Determination of outlier loci

Loci under putative selection were identified using two independent population differentiation (PD) approaches: BayeScan version 2.1 (Foll, 2012; Foll & Gaggiotti, 2008) and R package *PCAdap*t (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 & 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 & Hochberg, 1995). To decrease false positives due to multiple testing, I applied the false discovery rate (FDR) correction of *q-values* at 0.01. Outlier detection using R package *PCAdap*t was based on principal components analysis (PCA) and identified loci that excessively related to population structure as being under putative selection. More specifically, *PCAdap*t was run using two principal components (K = 2) as they explained genetic structure of populations and a min.maf = 0.05; candidate loci then were determined at FDR of 0.01using R package *qvalue* (Dabney et al., 2010).

3.2.5 Population genetic structure: neutral vs. 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) Discriminant Analysis of Principal Components (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 10 iterations of the '*find.clusters'* function and the visualize Bayesian Information Criterion (BIC) values for all 10 iterations along with estimated standard deviation 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 maximum likelihood using K values that ranged from one 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.

Analysis of molecular variance (AMOVA) was used to test genetic diversity structure between populations and geographic regions. AMOVA was run with ARLEQUIN version 3.5.2.2 (Excoffier & 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 AMOVA analyses 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, INDO, and THL). Secondly, I 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; and 7) KE, SA.

Isolation by distance (IBD) 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 & Dufour, 2007). Four IBD analyses were estimated separately using both neutral and outlier loci for all 16 sampled sites and seven sampled groups as described in AMOVA analysis. Contemporary least-cost oceanographic distances between each pair of sampling sites were estimated using R package *marmap* (Pante & Simon-Bouhet, 2013).

3.2.6 TreeMix analysis

To quantify historical connectivity across the Indo-Pacific region, a maximum likelihood (ML) tree of the entire neutral dataset (n = 9,930 SNPs) was constructed using TreeMix version 1.12 (Pickrell & 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 fit. 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 10 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 *f3*-statistics were calculated using the three-pop approach (Reich et al., 2009) within TreeMix version 1.12. This approach calculated *f3*-statistics for all three possible populations in the form of (X; A, B) following a bifurcating tree. A significantly negative *f3*-statistic value means that X is the product of admixture between A and B (Reich et al., 2009).

3.2.7 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^{-5}. Contigs within the assembled *P. monodon* transcriptome that exhibited 100% pairwise identity with each queried SNP sequence (69 bp or each allele) were compiled and reported along with contig annotations provided by Huerlimann et al. (2018).

To ensure comprehensiveness I 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 environmental association (EA) and population differentiation (PD) analyses as per Chapter 2. 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^{-5}. Contigs within the assembled *P. monodon* transcriptome that exhibited 100% pairwise identity with each queried SNP sequence (69 bp or 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, 2018).

For Chapter 2 approach the whole 10,593 SNPs dataset was subjected to two discrete environmental association (EA) analyses to identify putatively adaptive loci associated with SST_max or SST_min: 1) redundancy analysis (RDA) in the R package *vegan* (Jari Oksanen et al., 2018; Jari Oksanen et al., 2013) and 2) latent factor mixed models (LFMM) in the R package *LEA* (Frichot & François, 2015). LFMM analysis used seven latent factors because this matched the number of population clusters identified in DAPC and ADMIXTURE (Figure 3.2). Each identified putatively adaptive SNP was then: 1) assessed against the assembled *P. monodon* transcriptome (Huerlimann et al., 2018b) using Geneious Prime version 2019.1.3 (Kearse et al., 2012) with 100% pairwise identity and E-value < 10^{-5} cut-offs and 2) assessed for correlation between MAF and SST_max and SST_min using R package *ggpubr* (Kassambara, 2018).

SST_max and SST_min data (Appendix 3.1) for each Indo-Pacific sampling location was obtained from the Bio-ORACLE database (http://www.bio-oracle.org/downloads-to-email.php) (Assis et al., 2018; Tyberghein et al., 2012) as described in Chapter 2.

3.3 Results

3.3.1 Genotyping and SNP quality control

A total of 131,929 unique SNPs were obtained for each of 532 *P. monodon* individuals (Figure 3.1, Table 3.2) using DArTseq™. Initial quality control assessment of DArTseq™ data using *dartQC* removed 119,514 SNPs, or 90.6% of loci (Table 3.2). The remaining 12,415 loci were then tested for linkage disequilibrium (LD) and Hardy–Weinberg equilibrium (HWE), which further removed 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 dataset comprised of 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.

Table 3.2 Filtering steps and SNPs counts retained after each step

Steps	Retained SNP count
Initial potential SNPs	**131,929**
Average read depth of ≥ 7	**131,929**
Replication average ≥ 0.9	**129,416** [2513 of 131929 (1.9%)]
Call rate of ≥ 80%	**25,716** [103700 of 129416 (80.1%)]
Similar sequence clusters of ≥ 0.95	**17,772** [7944 of 25716 (30.9%)]
Minor allele frequency (MAF) of ≥ 0.02	**12,415** [5357 of 17772 (30.1%)]
LD filters with a correlation coefficient (r^2) threshold of 0.2	**10,593** 1,822 of 12,415]
HWE filters ($p \geq 0.0001$)	**10,593** [0 of 10,593]
Retained SNPs for further analysis	**10,593**

Linkage disequilibrium (LD) and Hardy–Weinberg equilibrium (HWE)

3.3.2 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 3.3). The mean private allelic richness (A_{RA}) per population ranged from 0.002 in BB to 0.014 in SLK. The percentage of polymorphic loci ranged from 18% to 76% for KE and TIW populations, respectively. Mean observed heterozygosity (H_O) across populations ranged from 0.02 to 0.15 while mean 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), while two Western Indian Ocean (WIO) populations (KE and SA) exhibited a five-fold difference between their H_O and H_E values (0.02 and 0.10), respectively (Table 3.3). Samples from SA and KE populations exhibited the lowest values for H_E (0.10), average multi-locus heterozygosity (Av. MLH = 0.023), and standardized multilocus heterozygosity (sMLH = 0.20), whereas their mean 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 multi-locus heterozygosity (Av. MLH: 0.15), and standardized multi-locus heterozygosity (sMLH: 1.19 - 1.25) than other populations (Table 3.3).

3.3.2 Determination 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 Material 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 while the remaining SNPs (n = 9,930) were deemed the neutral loci dataset.

Table 3.3 Genetic diversity indices among 16 population of *Penaeus monodon* assessed using all 10,593 high quality filtered SNP loci.

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

Number of samples (N), mean allelic richness (A_R), private allelic richness (A_{RA}), percentage of polymorphic loci (PPL, mean minor allele frequency (MAF) of polymorphic loci), mean observed heterozygosity (H_O), mean expected heterozygosity (H_E), average multi-locus heterozygosity (Av. MLH), standardised multi-locus heterozygosity (sMLH), internal relatedness (IR), and significant ($p < 0.05$) inbreeding coefficient (F_{IS}). SD: standard deviation.

3.3.3 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 (Appendix 3.2). I visualized clustering patterns for K values 6 through 8 to identify hierarchal clustering patterns as K increases (Figure 3.2 A 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 (Figure 3.2A and B). Also, 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, while 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) (Figure 3.2A and 3.4A). 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 (Figure 3.2B). 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 (Figure 3.2C 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 (Figure 3.2E 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 (Appendix 3.2). 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 (Figure 3.2).

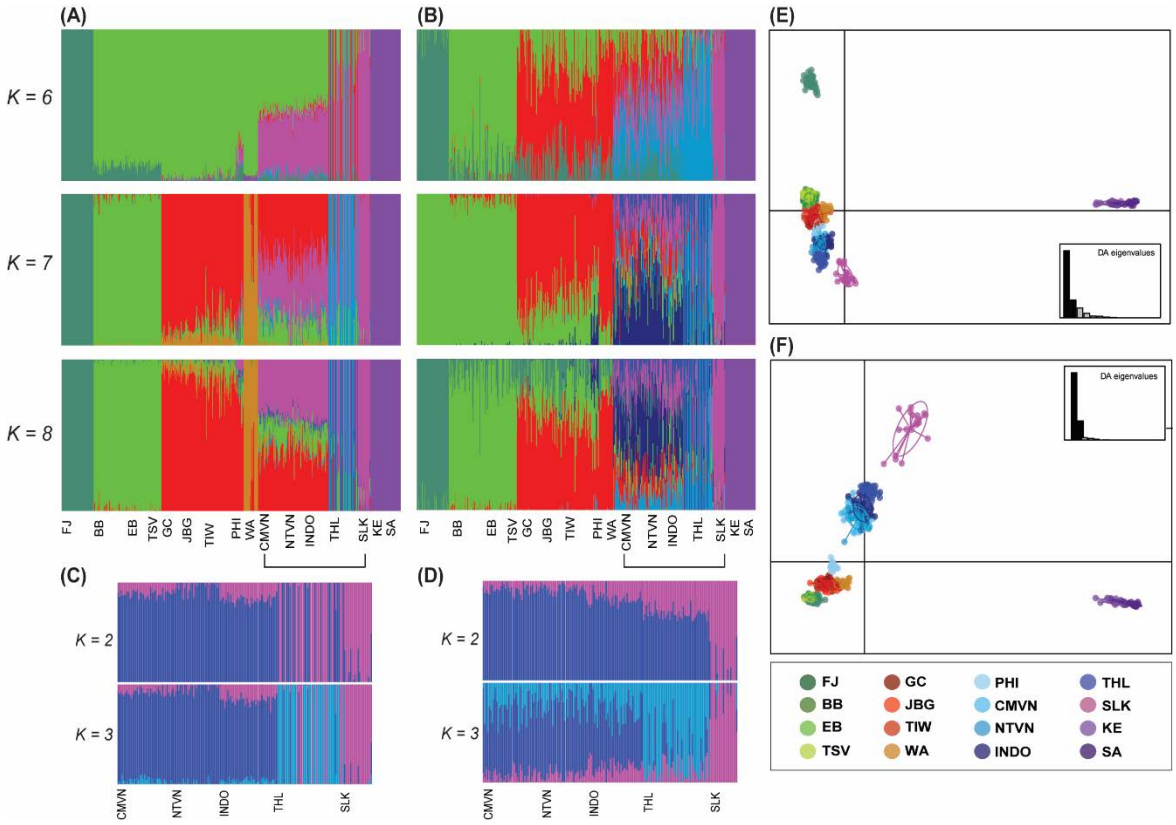


Figure 3.2 Population structure results for *Penaeus monodon* populations (n = 16) using neutral loci (**A, C, E**) and outlier loci (**B, D, F**). Plots of individual admixture determined using ADMIXTURE with cross-validation error using three unique K values (K = 6, K = 7, and K = 8) for both neutral (**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 3.1.

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 (Table 3.4). 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 ($p < 0.01$). In both neutral and outlier datasets, pairwise 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). I also observed that geographically proximal populations exhibited the

lowest F_{ST} values (e.g., between KE and SA and among EA populations; Table 3.4). Overall, mean F_{ST} was higher for outlier loci (F_{ST} = 0.40 ± 0.29) than neutral loci (F_{ST} = 0.16 ± 0.16) (Table 3.4).

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 < .001, respectively; Figure 3.3). 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 < .001, respectively; Figure 3.3).

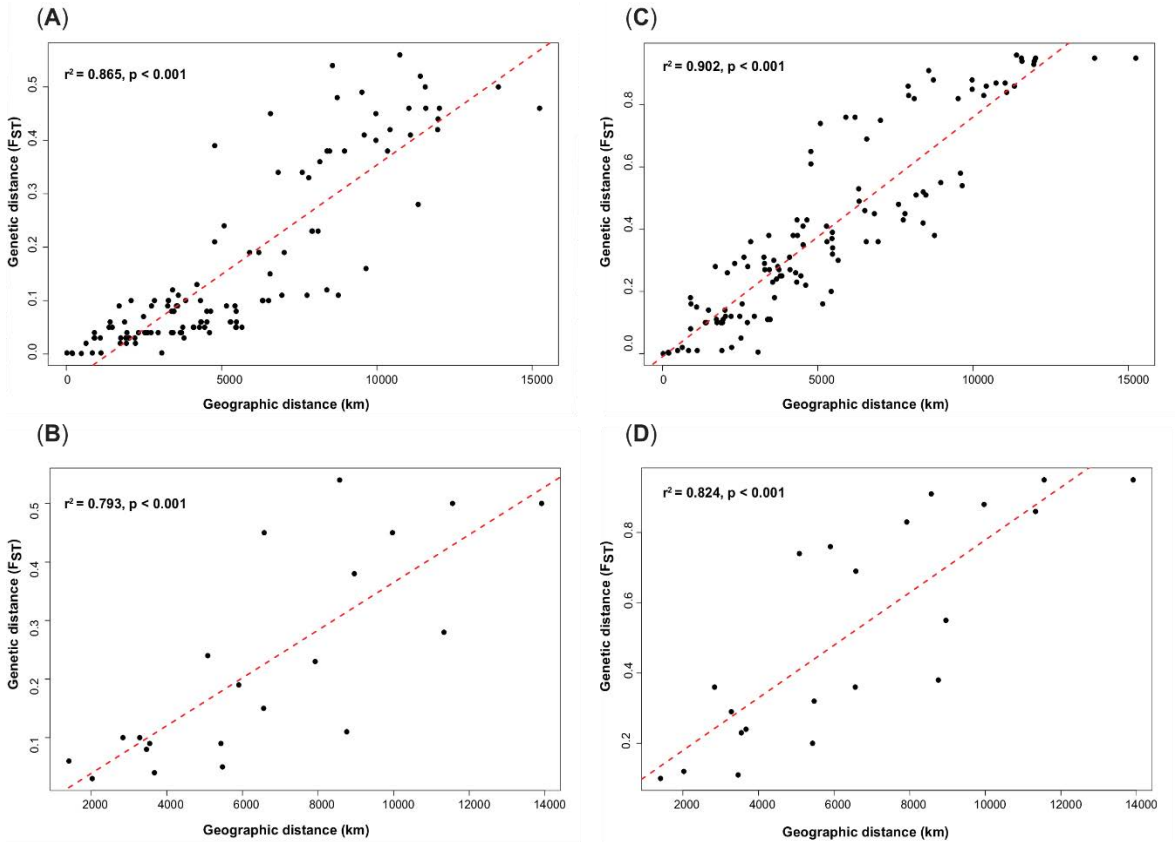


Figure 3.3 Plots of isolation by distance (IBD) with Mantel correlograms for the relationship between genetic (F_{ST}) and geographic (km) distances among *Penaeus monodon* populations (n = 16; **A**, **C**) and among clusters (n = 7; **B**, **D**) using 9,930 neutral SNPs (**A**, **B**) and 663 outlier SNPs (**C**, **D**).

Table 3.4 Population differentiation estimates for *Penaeus monodon* populations sampled using both neutral (n = 9,930) and outlier (n = 663) loci. The only non-significant F_{ST} value ($p > 0.05$) is in bold type.

	FJ	**BB**	**EB**	**TSV**	**GC**	**JBG**	**TIW**	**PHI**	**WA**	**NTVN**	**CMVN**	**INDO**	**THL**	**SLK**	**KE**	**SA**
Fiji (FJ)		0.08	0.08	0.08	0.10	0.09	0.09	0.11	0.15	0.12	0.11	0.11	0.16	0.28	0.46	0.50
Bramston Beach (BB)	0.11		0.002	0.001	0.03	0.03	0.02	0.05	0.09	0.06	0.05	0.06	0.10	0.23	0.42	0.46
Etty Bay (EB)	0.11	**0.000**		0.002	0.03	0.03	0.03	0.05	0.10	0.06	0.05	0.06	0.10	0.23	0.44	0.50
Townsville (TSV)	0.11	**0.002**	**0.004**		0.03	0.002	0.02	0.05	0.10	0.06	0.05	0.06	0.10	0.23	0.46	0.52
Gulf of Carpentaria (GC)	0.23	0.18	0.16	0.15		0.002	0.002	0.04	0.07	0.05	0.04	0.04	0.08	0.19	0.41	0.46
Joseph Bonaparte Gulf (JBG)	0.20	0.14	0.12	0.12	0.01		**0.001**	0.04	0.06	0.05	0.04	0.04	0.08	0.19	0.40	0.45
Tiwi Island (TIW)	0.16	0.11	0.10	0.10	0.01	0.01		0.04	0.06	0.04	0.03	0.04	0.08	0.19	0.38	0.42
Philippines (PHI)	0.36	0.31	0.27	0.26	0.11	0.12	0.10		0.11	0.05	0.04	0.04	0.09	0.21	0.49	0.56
Western Australia (WA)	0.36	0.31	0.29	0.27	0.12	0.10	0.10	0.18		0.10	0.09	0.09	0.13	0.24	0.48	0.54
NhaTrang Vietnam (NTVN)	0.42	0.41	0.36	0.34	0.25	0.27	0.28	0.14	0.25		0.02	0.02	0.05	0.12	0.36	0.41
CaMau Vietnam (CMVN)	0.38	0.37	0.32	0.30	0.22	0.24	0.25	0.11	0.23	0.02		0.00	0.04	0.10	0.34	0.38
Java Indonesia (INDO)	0.43	0.43	0.38	0.35	0.27	0.29	0.31	0.16	0.28	0.02	0.01		0.04	0.09	0.33	0.38
Thailand (THL)	0.54	0.53	0.49	0.46	0.39	0.41	0.43	0.28	0.38	0.10	0.08	0.05		0.10	0.34	0.38
SriLanka (SLK)	0.86	0.86	0.83	0.82	0.75	0.76	0.76	0.65	0.74	0.38	0.36	0.30	0.26		0.39	0.45
Kenya (KE)	0.95	0.93	0.94	0.95	0.84	0.85	0.83	0.82	0.88	0.51	0.48	0.45	0.45	0.61		**0.002**
SouthAfrica (SA)	0.95	0.94	0.95	0.96	0.87	0.88	0.86	0.87	0.91	0.58	0.55	0.52	0.51	0.69	**0.005**	

Population pairwise F_{ST} (Weir & Cockerham, 1984) estimates computed by using the R package *StAMPP* v.1.5.1 (Pembleton et al., 2013). Pairwise F_{ST} values of 663 outlier loci are shown below the diagonal, and pairwise F_{ST} values of 9,930 neutral loci are reported above. All F_{ST} values were significant at $p \leq 0.01$ following 1000 bootstraps performed across loci to generate confidence intervals.

AMOVA analyses for the two oceans (Indian and Pacific) using neutral loci showed a weak, but significant population structure (Table 3.5A; FSC = 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 analysis 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 3.5B). For neutral loci, the majority of genetic variability (83.0%) 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 3.5). AMOVA analysis also yielded significant FSC values for both neutral and outlier datasets (FSC = 0.01711 and 0.04685), respectively, indicating significant genetic division among these seven groups ($p < 0.01$).

Table 3.5 Analysis of molecular variance (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
A) Between the Indian and Pacific Oceans							
Neutral dataset (9,930 SNPs)	Among groups	1	17,168.99	69.54	22.75	FCT: 0.22748	< 0.01
	Among populations within groups	14	19,378.48	17.78	5.82	FSC: 0.07529	< 0.01
	Among individuals within populations	504	112,215.79	4.27	1.4		
	Within individuals	520	111,340.50	214.12	70.04	FST: 0.29960	< 0.01
B) Among seven regional groups							
Neutral dataset (9,930 SNPs)	Among groups	6	32,347.27	35.74	13.86	FCT: 0.13857	< 0.01
	Among populations within groups	9	4,200.20	3.80	1.47	FSC: 0.01711	< 0.01
	Among individuals within populations	504	112,215.79	4.27	1.65		
	Within populations	520	111,340.50	214.12	83.01	FST: 0.16986	< 0.01
Outlier dataset	Among groups	6	15,306.66	17.93	46.68	FCT: 0.46681	< 0.01
	Among populations within groups	9	727.41	0.96	2.50	FSC: 0.04685	< 0.01

(663 SNPs)	Among individuals within populations	504	9,691.87	-0.29	-0.76		
	Within individuals	520	10,304.5	19.82	51.58	FST: 0.48416	< 0.01

3.3.4 TreeMix analysis

The generated maximum likelihood (ML) phylogenetic tree, which considered the entire neutral loci dataset (*n* = 9,930 SNPs), was rooted to Fiji (FJ) given the consistent log-likelihood and geographic separation of this population from all other sampled populations. Five migration events (m = 5) provided the best explanation of sample covariance (Figure 3.4B) out of all migration events tested (m = 0 – 10; Appendix 3.3). 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 (Appendix 3.4). 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 (Figure 3.4B, see Appendix 3.3).

There were three additional migration events included in the ML tree from the WIO to Southeast Asia with strong migration weights (> 0.35; Figure 3.4B). 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 North-Western Australian populations (TIW, JBG, GC, and WA). The last migration event estimated a strongly weighted migration of NTVN and CMVN to THL (Figure 3.4B). 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 *f3*-statistics were also calculated, which resulted in 212 significant *f3*-scores with Z-scores ranging from -32.35 to -0.07 (Appendix 3.5). These significant *f3*-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 (Appendix 3.5). Sixteen *f3*-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 *f3*-statistics included SLK as one of the two source populations.

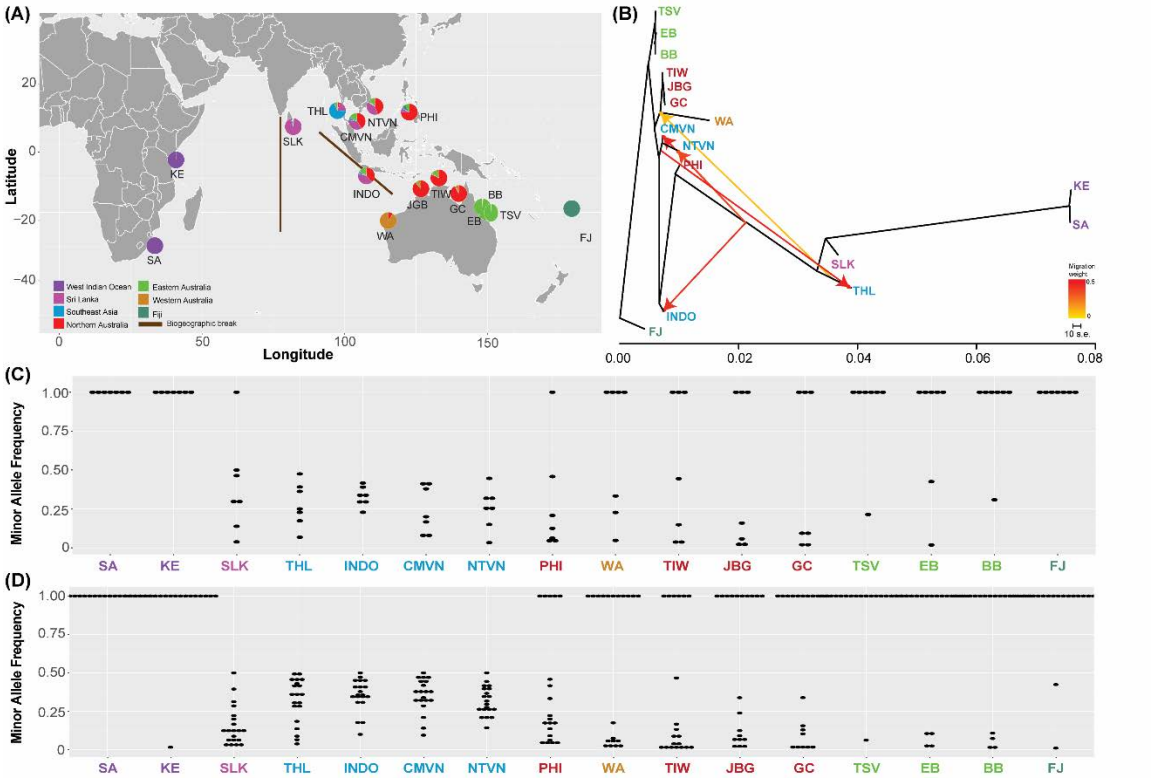


Figure 3.4 (**A**) Map of biogeographic breaks and the proportion of each of the 16 *Penaeus monodon* populations assigned to seven clusters identified in the program ADMIXTURE using 9,963 neutral loci. The proportion of individuals assigned to each cluster per site depicted in colored pie charts. (**B**) Population relationships and migration edges of *Penaeus monodon* inferred by TreeMix for neutral SNP dataset at five migration events (m = 5). Fiji (FJ) was use 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 3.1.

3.3.5 Identification of putatively adaptive SNPs

The transcriptome BLAST approach (see Methods) revealed that 20.5% of outlier loci (n = 136) matched coding contigs (n = 1,408 total) with an E-value < 10^{-5} (Appendix 3.6). Of these only 19.5% (n = 27) exhibited 100% pairwise identity to 62 transcriptome contigs (Appendix 3.7). 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 (Appendix 3.7). 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 significantly correlated ($p < 0.05$) with sea surface temperature maximum (SST_max) and minimum (SST_min), respectively (Table 3.6A, Appendix 3.8). Of these SNP 2019PM_4058 (C > G) and SNP 2019PM_6523 (T > G) matched transcriptome contigs annotated as "AChain Apocrustacyanin C1 Crystals" and "C-type lectin", respectively (Appendix 3.7), while the other five temperature correlated putatively adaptive SNPs matched contigs that were annotated as "Predicted: uncharacterized protein", "nucleic acid binding", or "N/A" (Appendix 3.7).

The combined environmental association (EA) and population differentiation (PD) approaches 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 (Figure 3.5). 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 3.6B, Appendix 3.9). Of these only 2019PM_5152 exhibited 100% pairwise identity with two transcriptome contigs annotated as "transcriptional regulatory–like" and "N/A" (Appendix 3.7).

MAF distribution for the seven temperatures 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 (Figure 3.4C). Similarly, MAF distribution for the 20 temperatures correlated putatively adaptive SNPs identified by the combined EA 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 (Figure 3.4D).

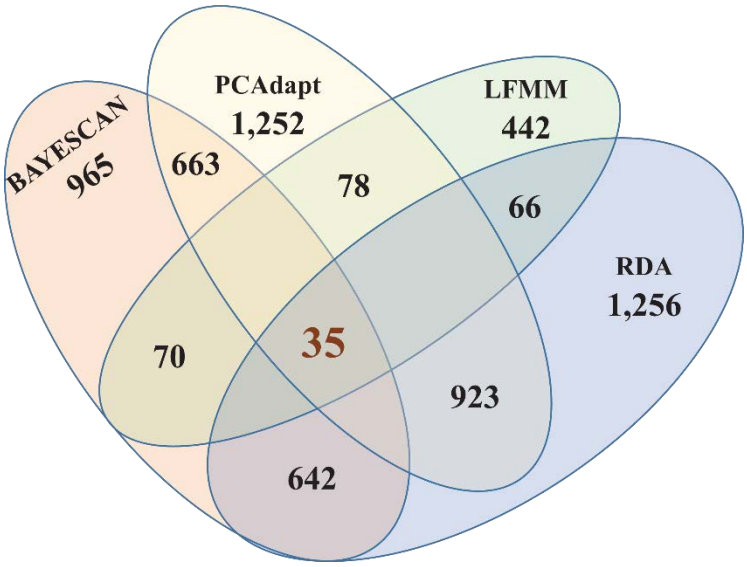


Figure 3.5 Venn diagram showing the overlapped putative adaptive SNPs identified by both environmental association and population differentiation approaches.

Table 3.6 Pearson correlations for putatively adaptive SNPs that exhibited a significant relationship between minor allele frequency and maximum or minimum sea surface temperature.

SNP ID	SST_max		SST_min	
	R	*p*-value	R	*p*-value
A) Candidate SNPs identified by the first analysis				
2019PM_4058	**-0.666**	**0.005**	0.005	0.986
2019PM_4727	**-0.667**	**0.005**	-0.162	0.550
2019PM_6523	**-0.629**	**0.009**	-0.248	0.354
2019PM_7307	**-0.541**	**0.030**	-0.325	0.220
2019PM_3144	-0.209	0.437	**-0.504**	**0.047**
2019PM_5152*	-0.305	0.252	**-0.650**	**0.006**
2019PM_6229	-0.249	0.352	**-0.763**	**0.001**
B) Candidate SNPs identified by the second analysis				
2019PM_665	**-0.580**	**0.020**	-0.080	0.770
2019PM_4130	**-0.600**	**0.010**	-0.060	0.840
2019PM_1552	**-0.639**	**0.008**	-0.237	0.378
2019PM_2811	**-0.655**	**0.006**	-0.251	0.349
2019PM_3848	**-0.670**	**0.005**	-0.222	0.409

2019PM_6807	**-0.596**	**0.015**	-0.296	0.266
2019PM_7282	**-0.588**	**0.017**	-0.177	0.513
2019PM_10077	**-0.633**	**0.008**	-0.098	0.719
2019PM_10517	**-0.564**	**0.023**	-0.143	0.596
2019PM_1233	-0.313	0.238	**-0.708**	**0.002**
2019PM_1287	-0.244	0.362	**-0.760**	**0.001**
2019PM_2814	-0.220	0.413	**-0.561**	**0.024**
2019PM_3748	-0.471	0.065	**-0.635**	**0.008**
2019PM_5152*	-0.304	0.252	**-0.650**	**0.006**
2019PM_6872	-0.209	0.437	**-0.685**	**0.003**
2019PM_6979	-0.208	0.440	**-0.686**	**0.003**
2019PM_8233	-0.263	0.324	**-0.500**	**0.048**
2019PM_9353	-0.289	0.278	**-0.653**	**0.006**
2019PM_9861	-0.254	0.343	**0.540**	**0.031**
2019PM_10081	-0.315	0.234	**-0.632**	**0.009**

Sea surface temperature maximum: SST_max; sea surface temperature minimum: SST_min. Significant values ($p < 0.05$) are bolded. *Overlapping SNP across both utilized approaches for putatively adaptive SNPs identification.

3.4 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, I 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 adaption 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.

3.4.1 Population genetic structure and gene flow

3.4.1.1 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 (Table 3.4). Outlier loci revealed an increase in genetic divergence across sample sites, which could be due to i) less variation within populations at outlier loci thereby leaving more to be partitioned between populations and ii) directional selection because an outlier locus is one that is not shared among all populations thus increasing differentiation among locations. 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 3.5).

My results indicated significant genetic differences among *P. monodon* populations. This finding is supported by similar previous observations for *P. monodon* (Benzie et al., 2002; Waqairatu et al., 2012; You et al., 2008), giant red shrimp *Aristaeomorpha foliacea* (Fernández et al., 2013), skunk clownfish *Amphiprion akallopisos* (Huyghe & 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 (Amaral et al., 2017; Hui et al., 2016; Huyghe & Kochzius, 2017; Vogler et al., 2012). 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; Waqairatu et al., 2012; You et al., 2008), as well as other Indo-Pacific marine species (Gaither et al., 2010; Horne et al., 2008; 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 (Crandall et al., 2019; Randall, 1998; Rocha et al., 2007).

3.4.1.2 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, while outlier SNPs found evidence for four clusters (Figure 3.2). 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, Chapter 2 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 environmental association analyses (n = 89 SNPs). I believe that the difference between the present study and Chapter 2 is most likely due to outlier determination being based on population differentiation verses environmental association analyses (n = 663 verses 89), respectively. Other non-SNP studies also reported *P. monodon* from Western Australia as being genetically distinct from other Australian *P. monodon* populations (Benzie et al., 2002; Brooker et al., 2000). Taken together, I 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 (Figure 3.2). 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 while Southeast Asia populations (e.g., Indonesia, Vietnam, and Thailand) exhibited the highest among-population genetic diversity (Table 3.3). This elevated genetic diversity among Southeast Asia populations could be driven by the migration of genetically divergent Indian Ocean shrimp into Southeast Asia (Figure 3.4 A 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 to 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 & 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, my 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.

3.4.1.3 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 $F_{ST} \geq 0.39$). Vogler et al. (2012) presented the following three hypotheses as potential explanations for the observed differences in genetic structure of crown-of-thorns starfish *A. planci* populations collected throughout the Indian Ocean that were believed to have diverged during the late-Pliocene to early-Pleistocene (1.86 – 2.89 million years ago): 1) landmass distribution differences, 2) geographical barriers (lack of coral reefs between these regions), and 3) ocean surface currents. 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 (Amaral et al., 2017; Hui et al., 2016; Huyghe & Kochzius, 2017; Vogler et al., 2012). In the present study, TreeMix analysis revealed that there was a very low level of gene flow between WIO populations and Sri Lanka (Figure 3.4B), as well as the lack of any significant negative *f3*-statistics for three-population tests for Sri Lanka, South Africa and Kenya (Appendix 3.5). 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 approximate 120 m given the reported re-colonisation (by shelf areas or populations further north/east) of lost habitat following sea level rise following last glacial maximum (Benzie et al., 2002; Forbes et al., 1999; Silva et al., 2010). 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 (Huyghe & Kochzius, 2018; Otwoma & Kochzius, 2016; Schott et al., 2009). 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 (Kaka et al., 2019; Sk et al., 2015). Although I 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 (Farhadi et al., 2017; Hui et al., 2016; Huyghe & Kochzius, 2017; Vogler et al., 2012). 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 Western Indian Ocean (Huyghe & Kochzius, 2018; Schott et al., 2009). 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 3.3). 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 Meeûs (2017)) 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 in the neutral loci dataset; Appendix 3.10). 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 common SNPs across all 16 populations (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 3.3), respectively, while H_O and H_E only differed by

0.01 – 0.02 across all 16 populations (Table 3.7). 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).

Table 3.7 Genetic diversity indices among 16 populations of *Penaeus monodon* assessed using 2,155 SNP loci (> 99% concurrence in technical replicates with missing data per SN*P* < 2%)

Population	A_R (± SD)	H_o (± SD)	H_E (± SD)	F_{IS}
Fiji	1.34 ± 0.43	0.09 ± 0.16	0.10 ± 0.16	0.01
Bramston Beach (BB)	1.43 ± 0.42	0.10 ± 0.15	0.10 ± 0.14	0.01
Etty Bay (EB)	1.43 ± 0.43	0.10 ± 0.15	0.10 ± 0.15	-0.01
Townsville (TSV)	1.43 ± 0.44	0.10 ± 0.15	0.11 ± 0.15	0.01
Gulf of Carpentaria (GC)	1.47 ± 0.43	0.11 ± 0.15	0.11 ± 0.15	0.03
Joseph Bonaparte Gulf (JBG)	1.47 ± 0.43	0.11 ± 0.15	0.10 ± 0.15	0.01
Tiwi Island (TIW)	1.48 ± 0.41	0.11 ± 0.15	0.12 ± 0.15	0.03
Nickol Bay (NKB)	1.39 ± 0.45	0.11 ± 0.18	0.11 ± 0.16	-0.08
Philippines (PHI)	1.44 ± 0.50	0.12 ± 0.18	0.11 ± 0.15	-0.04
NhaTrang Vietnam (NTVN)	1.50 ± 0.44	0.13 ± 0.17	0.13 ± 0.16	0.01
CaMau Vietnam (CMVN)	1.55 ± 0.40	0.13 ± 0.16	0.13 ± 0.15	0.02
Java Indonesia (INDO)	1.55 ± 0.41	0.13 ± 0.16	0.13 ± 0.16	0.03
Thailand (THL)	1.47 ± 0.46	0.14 ± 0.18	0.13 ± 0.16	-0.03
SriLanka (SLK)	1.44 ± 0.46	0.10 ± 0.16	0.10 ± 0.15	0.02
Kenya (KE)	1.12 ± 0.32	0.03 ± 0.09	0.03 ± 0.09	0.03

SouthAfrica (SA)	1.11 ± 0.30	0.03 ± 0.09	0.03 ± 0.09	0.02

3.4.2 Signatures of local adaptation

In chapter 2, I recently demonstrated the potential role of environmental factors (sea surface temperature) in shaping the genetic structure and local adaptation of Australian *P. monodon* populations. Based on the results of Chapter 2, I hypothesized that *P. monodon* may also exhibit signatures of local adaption to sea surface temperature on a larger scale (e.g., across entire Indo-Pacific region). Using two discrete analytical approaches I 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 EA and PD (all SNPs) approaches ensures that all putatively adaptive SNPs associated with temperature were identified. My 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 (*C. maenas*) (Benestan et al., 2016; Jeffery et al., 2018; 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 3.6; Appendix 3.7). In *P. monodon* crustacyanin protein is known to be involved in temperature sensitive hypodermis pigmentation (Budd et al., 2017; Fan et al., 2021; Wade et al., 2015; Wade et al., 2014) and substrate color adaptation (Wade et al., 2012), which have important implications for *P. monodon*, *P. japonicus*, and *L. vannamei* aquaculture (Rodríguez et al., 2017). Another identified temperature correlated putatively adaptive SNP was 2019PM_6523, which matched transcriptome contigs annotated as "C-type lectin" (Table 3.6; Appendix 3.8; and Appendix 3.7). 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 (*PmChi5*). *PmChi5* 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). While 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).

3.4.3 Implications for aquaculture and fishery management

Within the Indo-Pacific *P. monodon* is considered to be 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, a multi-jurisdictional approach to management of these wild stocks is recommended as opposed to independent management by each country, because local aquaculture farmers from many locations source wild individuals as broodstock (Coman et al., 2006; Marsden et al., 2013; Wong et al., 2016) which can result in over-exploitation of natural *P. monodon* populations. The sustainability of *P. monodon* fisheries in this area can be problematic if wild-sourced broodstock is from degraded (e.g., introducing disease into the farm) and overfished populations. Additionally, *P. monodon* aquaculture development in terms of selective breeding programs and characterisation of commercially important traits remains largely under-developed (Guppy et al., 2020; Norman-López et al., 2015). Thus, a better understanding of the distribution of genetic diversity and genetic differentiation within and

among Indo-Pacific populations presented in this study can advance the efficiency and success of selective breeding programs through improved selection of genetically diverse broodstock. Lastly, future research should aim to elucidate which outlier SNPs (or regions of the genome) are directly linked to phenotype(s) of interest for selective breeding programs (e.g., thermal adaptation, enhanced growth rate, altered pigmentation, improved disease resistance, etc) (Lebeuf-Taylor et al., 2019).

3.5 Conclusion

We utilized a SNP dataset containing 10,593 genome-wide SNPs to determine the population structure, connectivity, and local adaptation of *P. monodon* from throughout the Indo-Pacific region. 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. I also identified 26 putatively adaptive SNPs that exhibited significant Pearson correlation between minor allele frequency and sea surface temperature maximum and minimum. 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.

Chapter 4. Genome-wide SNP markers reveal biogeographic impacts and confirm the sister taxa relationships in giant black tiger shrimp (*Penaeus monodon*) across the Indo–Pacific region

4.1 Introduction

Members of the family Penaeidae represent over 200 shrimp species, several of which are commercially farmed and fished (Chan, 1998; Dall et al., 1990; Grave & Fransen, 2011; Hurzaid et al., 2020). The life cycle of penaeid shrimp is characterized by a pelagic larval phase that permits oceanic dispersal and migration between offshore and inshore waters, which, collectively, expose larvae to a variety of environmental factors. Due to the wide geographical distribution and economic importance of some species (e.g., particularly species from the genus *Penaeus*) the phylogenetic relationships among species in this family are relatively well understood through the use of traditional genetic markers, including amplified fragment length polymorphism (AFLP), cytochrome oxidase subunit I (COI), and mitochondrial DNA marker (mtDNA) datasets (Calo-Mata et al., 2009; Chan et al., 2008; Hualkasin et al., 2003; Hurzaid et al., 2020; Lavery et al., 2004; Quan et al., 2004; Voloch et al., 2005, 2009; Wang et al., 2004). Chan et al. (2008) consolidated species into a three-tribe scheme within the Penaeidae represented by eight genera, however, more recently three studies (Hurzaid et al., 2020; Lavery et al., 2004; Ma et al., 2011) indicated that the Penaeidae appears to be comprised of six genera divided into two clades (i: the *Melicertus* and *Marsupenaeus* and ii: the *Fenneropenaeus, Farfantepenaeus, Litopenaeus* and *Penaeus*).

Within the genus *Penaeus*, the giant black tiger shrimp (*P. monodon*) is native to and considered to be one of the most important aquaculture species throughout the Indo-Pacific region. The life history of *P. monodon,* which includes a pelagic marine larval phase, highlights potential for dispersal between geographically discrete populations (Chapter 3). Indeed, studies which have focused on genetic diversity and population structure of *P.monodon* using different genetic markers show significant population structure across the Indo-Pacific region and collectively suggest the existence of a deep-seated genetic division between Indian and Pacific Ocean populations (Benzie et al., 2002; Chapter 2; Chapter 3; Waqairatu et al., 2012; You et al., 2008). More broadly, the main barriers between the Indian and Pacific Oceans (including Sunda and Sahul shelves) and between the East and West Indian Ocean have effectively

restricted gene flow between these oceans and thus, strongly influenced the evolutionary history of not only *P. monodon*, but many other Indo-Pacific marine species (Ahti et al., 2016; Liu et al., 2014; Ludt & Rocha, 2015; Wainwright et al., 2018; Waqairatu et al., 2012).

Genomic phylogeography plays an important role in describing evolutionary processes and to identify historical events including geographic and ecological drivers that have shaped the current population genetic structure within species (Avise, 2000; Bowen et al., 2016; Van Cise et al., 2019). Relevant to *P. monodon,* Waqairatu et al. (2012) and You et al. (2008) generated phylogeographic trees based on mitochondrial haplotype and microsatellite sequences for the Indo-Pacific *P. monodon*, which revealed that *P. monodon* separated into two divergent lineages (Pacific and Indian Oceans lineages); however, *P. monodon* samples in these studies did not cover the large sample sites of the Indian Ocean. Additionally, mtDNA exhibits a non-Mendelian mode of inheritance and provides a single perspective which may not be a reliable indicator of species relationships or demographic history (Ballard & Whitlock, 2004), while microsatellites fail to capture critical components of the original spirit of phylogeography (Edwards et al., 2015). Thus, these two markers have limited resolution power to genetic structure and phylogeographic patterns (Jeffries et al., 2016; Zarraonaindia et al., 2012). Recently, single nucleotide polymorphisms (SNP) have been demonstrated to be the most prevalent type of polymorphism within investigated genomes d becoming the marker of choice in phylogeographic, phylogenetic, and population genetics studies (Hodel et al., 2017; Jeffries et al., 2016; Morin et al., 2004). Furthermore, phylogeographic studies using SNP loci have demonstrated that this approach provides more (cost-effective) data for resolving phylogenetic relationships at deep evolutionary timescales as well as alleviates concerns regarding intragenic recombination (Harvey et al., 2016; Leaché & Oaks, 2017).

The utility of high-density genome-wide markers, SNP and presence-absence variants (PAV) loci as molecular markers for population genetic and phylogenetic reconstruction research have been demonstrated in previous studies (Kjeldsen et al., 2019; Lal et al., 2018; Morse et al., 2018). These two markers were generated independently by DArTseq™ genotyping technology. While SNPs were used for both population and phylogenetic analyses, whereas PAVs were only used in phylogenetic reconstructions (Kjeldsen et al., 2019; Morse et al., 2018). In the present study, I obtained wild *P. monodon* from 16 sites throughout their natural Indo-Pacific distribution to infer phylogeographic relationships using both SNPs and PAV markers.

To clarify the phylogenetic position of *P. monodon* and the relationships among penaeid shrimp, six other penaeid species occurring in the Indo-Pacific were also included. The enhanced phylogenetic resolution provided by these SNP- and PAV-based phylogeographic trees will help to identify Evolutionary Significant Units (ESUs) and improve the effectiveness of *P. monodon* aquaculture and fisheries management strategies (i.e., identifying areas of high genetic diversity).

4.2 Methods

4.2.1 Sample collection and DNA extraction

Wild adult *P. monodon* were sampled from 16 sites throughout their natural Indo-Pacific distribution, which spans from coastal waters of Fiji in the central Pacific, throughout Australia and Southeast Asia, to the Eastern Indian Ocean and south-western coast of Africa (Figure 4.1; Table 4.1). Specific sampling locations for *P. monodon* were the same as described in Chapter 3. Briefly, twenty *P. monodon* were sampled from each Indo-Pacific population, except for the Philippines (PmonoPHI), Sri Lanka (PmonoSLK), and Kenya (PmonoKE) populations (n = 12, 19, and 16), respectively (Figure 4.1).

To validate the phylogenetic relationships among sister taxa of *P. monodon* using genome-wide SNP markers, six other Penaeidae species were used (Figure 4.1; Table 4.1). These six penaeids were included in analyses because they were either closely related relatives within the genus *Penaeus*, or more distantly related (Hurzaid et al., 2020; Voloch et al., 2005, 2009; Wang et al., 2004). Whiteleg shrimp (*Litopenaeus vannamei*, n = 8) samples were obtained from an aquaculture farm in Indonesia, while wild adult samples for banana shrimp (*Fenneropenaeus merguiensis*; n = 13), brown tiger shrimp (*Penaeus esculentus*, n = 4), green tiger shrimp (*Penaeus semisulcatus*, n = 4), western king shrimp (*Penaeus latisulcatus*, n = 3) and red-spotted shrimp (*Penaeus longistylus*, n = 1), were obtained from commercial sources harvested off the Northern Australian coast.

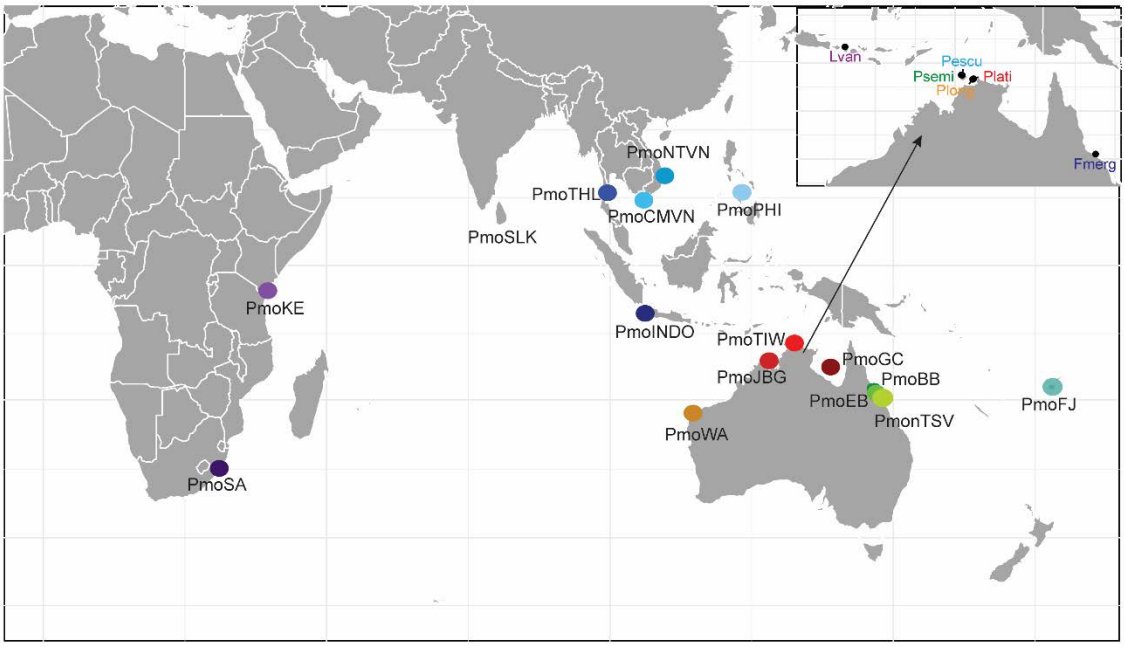


Figure 4.1 Sampling locations for seven Penaeidae species (*n* = 340 individuals) used in this study. *Penaeus monodon* samples: Eastern Australia (PmoBB, Bramston Beach; PmoEB, Etty Bay; PmoTSV, Townsville); Northern Australia (PmoGC, Gulf of Carpentaria; PmoJBG, Joseph Bonaparte Gulf; PmoTIW, Tiwi Island); Western Australia (PmoWA); Vietnam (PmoCMVN, Ca Mau; PmoNTVN, Nha Trang); Fiji (PmoFJ); Indonesia (PmoINDO); Thailand (PmoTHL); Philippines (PmoPHI); Sri Lanka (PmoSLK); Kenya (PmoKE); South Africa (PmoSA). Non-*P. monodon* samples: *Fenneropenaeus merguiensis*: Fmerg; *Litopenaeus vannamei*: Lvan; *Penaeus esculentus*: Pescu; *Penaeus semisulcatus*: Psemi; *Penaeus longistylus*: Plong; and *Penaeus latisulcatus*: Plati. Map generated using R version 4.0.3.

Pleopod tissue was excised from shrimp individuals with DNA-free scissors, preserved in 80% ethanol, and stored at -20˚C until DNA extraction. Genomic DNA (gDNA) was extracted from each pleopod tissue sample following a modified cetyl-trimethyl ammonium bromide (CTAB) protocol (Adamkewicz & Harasewych, 1996; Chapter 2, Chapter 3). All gDNA extractions were subsequently cleaned using Sephadex™ G-50 (GE, 2007). Concentration and purity of extracted gDNA (260:230 nm and 260:280 nm ratios), respectively, were measured using a Nanodrop 1000 Spectrophotometer (ThermoFisher Scientific Pty Ltd, Australia) with gDNA integrity confirmed by 0.8% agarose gel containing GelGreen (ThermoFisher Scientific Pty Ltd, Australia). All gDNA samples were normalised to 50 ng/µl (20 µL final volume) before submission to Diversity Arrays Technology (DArTseq^{TM}; Canberra Australia) for library preparation and DArTseq^{TM} sequencing.

Table 4.1 Species code, species, sampling location, and total number of individuals (*N1*) obtained for each of the seven shrimp species included in this study. Number of individuals (*N2* and *N3*) used for a) phylogeny of non-*Penaeus monodon* species and b) phylogeography of *P. monodon* analyses, respectively.

Code	Species		Sampling location	*N1*	*N2*	*N3*
Fmerg	Banana shrimp	*Fenneropenaeus merguiensis*	Queensland, Australia (wild)	**13**	**13**	**0**
Pescu	Brown tiger prawn	*Penaeus esculentus*	Northern Territory, Australia (wild)	**4**	**4**	**4**
Psemi	Green tiger prawn	*Penaeus semisulcatus*	Northern Territory, Australia (wild)	**4**	**4**	**4**
Plati	Western king prawn	*Penaeus latisulcatus*	Northern Territory, Australia (wild)	**3**	**3**	**0**
Plong	Red-spotted prawn	*Penaeus longistylus*	Northern Territory, Australia (wild)	**1**	**1**	**0**
Lvan	Whiteleg shrimp	*Litopenaeus vannamei*	Global Gen farm, Indonesia (farm)	**8**	**8**	**0**
PmoBB	Giant black tiger shrimp	*Penaeus monodon*	Bramton Beach, Australia (wild)	**20**	**1**	**6**
PmoEB			Etty Bay, Australia (wild)	**20**	**0**	**0**
PmoTSV			Townsville, Australia (wild)	**20**	**1**	**6**
PmoGC			Gulf of Carpentaria, Australia (wild)	**20**	**0**	**6**
PmoJBG			Joseph Bonaparte Gulf, Australia (wild)	**20**	**1**	**6**
PmoTIW			Tiwi Island, Australia (wild)	**10**	**0**	**0**
PmoWA			Western Australia (wild)	**20**	**1**	**6**
PmoFJ			Fiji (wild)	**20**	**1**	**6**
PmoPHI			Philippines (wild)	**12**	**1**	**6**
PmoCMVN			Ca Mau, Vietnam (wild)	**20**	**1**	**6**
PmoNTVN			Nha Trang, Vietnam (wild)	**20**	**1**	**6**
PmoINDO			Indonesia (wild)	**20**	**1**	**6**
PmoTHL			Thailand (wild)	**20**	**1**	**6**
PmoSLK			Sri Lanka (wild)	**19**	**3**	**6**
PmoKE			Kenya (wild)	**16**	**2**	**6**
PmoSA			South Africa (wild)	**20**	**3**	**6**
Total				**340**	**51**	**92**

4.2.2 Sequence quality control and marker filtering

All samples were sequenced and genotyped using DArTseq™ technology at Diversity Arrays Technology, Canberra, Australia as previously described (Chapter 2 and Chapter 3). DArTseq™ generated two independent marker types based on restriction-site-associated (RAD) fragments: single nucleotide polymorphisms (SNP) and presence-absence variants (PAV).

Genomic DNA extracted from non-*P. monodon* samples was amplified during DArTseq™ protocol using primers designed for *P. monodon*, which notably increased the total number of raw SNP sequences from 131,929 SNPs (obtained for *P. monodon* sample only; Chapter 3) to 166,946 SNPs (obtained for both *P. monodon* and the other six Penaeidae samples) (Table 4.2). This aspect of the utilized DArTseq™ protocol resulted in a high rate of missing SNP and PAV data for all non-*P. monodon* samples. Accordingly, all raw SNP sequences missing from any non-*P. monodon* samples after initial SNP calling were removed from each *P. monodon* sample and then the remaining (i.e., normalized) SNP datasets for each of the Penaeidae species were filtered using custom *dartQC* python scripts (Chapter 2 and Chapter 3) with removal of all SNPs that matched any of the following stringent criteria: 1) low read depth (--read_counts < 7), 2) average repeatability < 90%, 3) call rate ≤ 50%, and 4) duplicated SNPs based on similar sequence clusters (-- cluster > 95%). The resulting high-quality SNP datasets (n = 4,496 each) were included in subsequent phylogenetic analyses.

For PAV filtering, due to the missing data of non-*P. monodo*n samples (n = 33 individuals), I first identified PAV markers missing of all these samples from the raw PAV dataset (n = 75,359). These missing PAV markers (with a score of zero) then were removed from the dataset. After that, PAV markers were filtered manually as a common dataset to retain the most informative marker set across all *P. monodon* and non-*P. monodon* species (Lal et al., 2018; Kjeldsen et al., 2019). Accordingly, the remaining PAV markers were filtered based on call rate, minor allele frequency (MAF), and technical reproducibility. Only PAV markers with a call rate ≥ 82%, MAF ≥ 0.02 across the dataset, and technical reproducibility ≥ 95% were retained. The resulting high-quality PAV datasets (n = 7,054) were included in subsequent phylogenetic analyses.

4.2.3 Phylogenetic analysis

Phylogenetic relationships were inferred based on both SNP and PAV datasets by using three independent approaches: 1) maximum-likelihood (ML), 2) Bayesian inference (BI), and 3) neighbour-joining (NJ).

For phylogenetic inference the species relationships among all seven penaeid species were determined for *F. merguiensis*, *L. vannamei*, and five *Penaeus* species (*P. monodon*, *P. semisulcatus*, *P. esculentus*, *P. latisulcatus*, and *P. longistylus*). For computationally viability, the number of included *P. monodon* samples in analyses was reduced from 307 (column *N1*, Table 4.1) to 18 by choosing 18 samples at random across all Indo–Pacific sampling locations (column *N2*, Table 4.1). Phylogenetic relationships included individuals for all seven species (n = 51 total; Table 4.1 column *N3*) and were based on both SNP and PAV datasets using independent NJ, ML, and BI analyses. *P. longistylus* was chosen as the outgroup by following the phylogenetic results of Hurzaid et al. (2020) to root the inferred trees.

For phylogeographic analyses of Indo–Pacific *P. monodon*, given the large number of *P. monodon* samples (n = 307; column *N1*, Table 4.1) and computational intensity of phylogenetic analyses, the number of samples utilized was reduced. Maximum-likelihood trees were constructed to trim the *P. monodon* dataset (n = 307; column *N1*, Table 4.1) down to six individual samples each from 14 discrete sampling locations from across the Indo–Pacific region (n = 84; column *N3,* Table 4.1) using IQ-TREE v2.1.0 (Minh et al., 2020) with best-fit substitution model (K3Pu+F+R6) calculated by *ModelFinder* (Kalyaanamoorthy et al., 2017) under the Bayesian information criterion (Appendix 4.1). The *P. monodon* samples that were retained in the trimmed dataset (n = 84; column *N3,* Table 4.1) were chosen based on their genetic distance to ensure that overall genetic distance across all sampling locations was retained. Given the closest known relatives and the least genetic distance between *P. monodon*, *P. esculentus,* and *P. semisulcatus* (Hurzaid et al., 2020; Voloch et al., 2005, 2009), all *P. esculentus* (n = 4) and *P. semisulcatus* (n = 4) individuals were included in phylogeographic relationship analyses as the outgroup for each independent analysis. Phylogeographic relationships among Indo–Pacific *P. monodon* samples within the trimmed dataset (column *N2,* Table 4.1) were determined based on both SNP and PAV loci using three independent approaches (NJ, ML, and BI).

ML trees were constructed based on both SNP and PAV markers in IQ-TREE v2.1.0 (Minh et al., 2020) using best-fit substitution models, which were identified using Bayesian

information criterion (BIC), as determined by *ModelFinder* (Kalyaanamoorthy et al., 2017). Branch support was assessed by running 10,000 ultrafast bootstrap replicates (UFBoot; Hoang et al. (2018)) and 10,000 bootstrap replicates of the SH-like approximate likelihood-ratio test (SH-aLRT; Guindon et al. (2010)). The BI of phylogenetic relationships was determined based on the PAV dataset using MrBayes v3.2.7 (Ronquist et al., 2012). MrBayes analyses consisted of two independent runs with the following parameters: 1) six chains, 2) 10 million generations, 3) temperature = 0.1, and 4) sample frequency = 1000 (first 15% considered burn-in and discarded). To monitor for parameter convergence during Bayesian analyses Tracer v1.7.1 (Rambaut et al., 2018) was used to calculate log-likelihood scores, which permitted accurate determination of additional run requirements and stationary phase obtainment. A NJ tree based on Nei's (1972) genetic distances was constructed based on the SNP dataset in MEGA-X (Kumar et al., 2018) with 1000 bootstrap replicates. All consensus phylogenetic trees were visualized and edited using Fig-Tree v1.4.4 (Rambaut, 2018).

4.3 Results

4.3.1 Genotyping and SNP/PAV filtering

The raw dataset contained 166,946 SNPs genotyped across all samples for all shrimp species (n = 340; Table 4.1 column *N1*). The first filtering step (Section 4.2.2) removed 105,393 SNP loci missing from all 33 non-*P. monodon*, or 63.1% (n = 61,553 loci remaining). The second filtering step (Section 4.2.2) in *dartQC* removed a further 57,057 SNP loci, or 34.2% (Table 4.2). The remaining 4,496 SNP loci for all samples (n = 340) were used for further analyses. Of note is that no individuals were lost due to poor call rates.

The raw PAV dataset contained 75,359 variant scores across all samples for all shrimp species (n = 340; Table 4.1 column *N1*). The first filtering step (Section 4.2.2) removed 33,686 PAV loci with a score of zero across all 33 non-*P. monodon*, or 44.7% of loci (n = 41,673 loci remaining). Following filtering, a total of 26,632 PAV loci, or 63.9% of PAVs were removed due to a call rate < 82%, MAF < 0.02 across the dataset, and technical reproducibility < 95%. The remaining 7,054 PAV loci were used for further analyses. Of note is that no duplicate genomic SNP loci are represented in the PAV dataset.

Table 4.2 Filtering steps and SNPs counts retained after each step

Steps	Retained SNP count
Initial potential SNPs	**166,946**
SNP missing in all 33 individuals of non- *P. monodon* samples	**61,553** [105,393 of 166,946 (63.1%)]
Average read depth of ≥ 7	**61,553**
Replication average ≥ 0.9	**61,553** [340 of 61,553 (0.6%)]
Call rate of ≥ 50%	**6,281** [54,932 of 61,213 (89.7%)]
Similar sequence clusters of ≥ 0.95	**4,496** [1,785 of 6,281 (28.4%)]
Retained SNPs for further analysis	**4,496**

4.3.2 Phylogenetic relationships among seven Penaeidae species

ModelFinder identified the HKY+F and GTR2+FO+R2 as the best-fitting substitution models for SNP-based and PAV-based ML trees, respectively (Figure 4.2A and 4.3B). ML, BI, and NJ phylogenetic trees showed nearly identical topologies, although node supports and relationships among divergent species differed (Figure 4.2 and 4.3). SNP-based ML and NJ trees contained similar branching pattern and reflected some, but not all, of the branching patterns present in PAV-based ML and NJ trees; however, both SNP- and PAV-based ML and NJ trees were monophyletic for each of the seven *Penaeidae* species.

4.3.2.1 SNP-based ML and NJ trees

SNP-based ML and NJ reconstructions (Figure 4.2) both contained seven well supported clades (Bootstrap Support, BS > 89); however, node support for the division between *P. esculentus*, *P. semisulcatus* and *P. monodon* was low in the SNP-based ML tree (BS = 42). Both ML and NJ reconstructions separated four species (*F. merguiensis*, *P. esculentus*, *P. semisulcatus*, and *P. monodon*) into one large species-level clade (BS ≥ 89; Figure 4.2). Both ML and NJ reconstructions recovered *L. vannamei* as a sister taxon to *F. merguiensis*, *P. esculentus*, *P. semisulcatus*, and *P. monodon* (BS = 100). Within the *L. vannamei* lineage, both ML and NJ trees recovered *P. esculentus* and *P. semisulcatus* as sister taxon (BS = 100), respectively.

Moreover, the SNP-based NJ reconstruction placed *P. esculentus*, *P. semisulcatus*, *P. monodon*, and *F. merguiensis* into one clade (i.e., sister taxa; BS = 100), wherein *P. monodon* appeared to be more closely related to *F. merguiensis* than to *P. esculentus* or *P. semisulcatus*. The SNP-based ML reconstruction placed *P. esculentus*, *P. semisulcatus,* and *P. monodon* into one clade (BS = 42) wherein *F. merguiensis* was a sister species (BS = 89). Interestingly, the monophyletic *P. monodon* in the SNP-based ML reconstruction revealed a deep split between individuals from the Pacific Ocean and the Sri Lankan clade and individuals from the WIO clade (BS = 100). At the base of the tree, *P. longistylus* and *P. latisulcatus* were recovered as sister species to all other species (BS > 94).

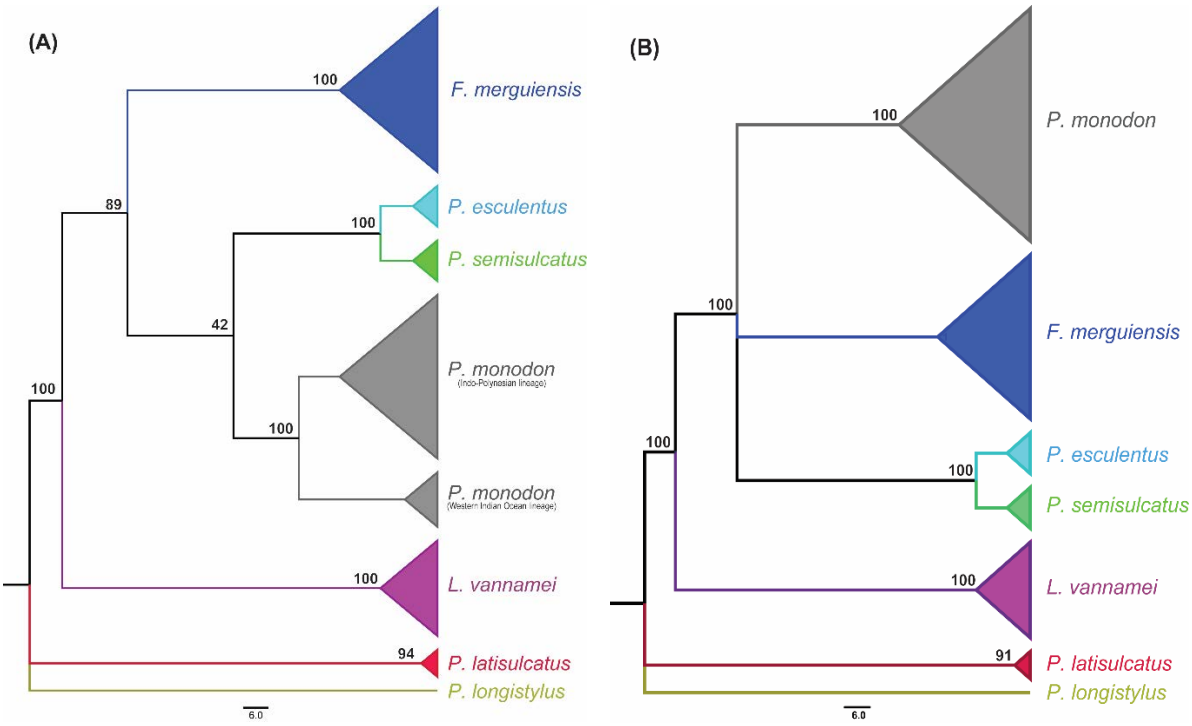


Figure 4.2 Phylogenetic relationships of seven species from the Family Penaeidae using 4,496 SNP loci (**A**) and 7,054 PAV loci (**B**). (**A**) The maximum likelihood reconstruction was conducted using IQ-TREE v2.1.0 under HKY+F model with 10,000 bootstrap replicates and log likelihood scores of ln L = -20,273.726. (**B**) neighbour-joining (NJ) based on Nei's (1972) genetic distances in MEGA-X with 1000 bootstrap replicates. *Penaeus longistylus* was used as the outgroup taxon. Numbers at nodes represent bootstrap values in percentage.

4.3.2.2 PAV-based ML and BI trees

PAV-based ML and BI phylogenetic reconstructions showed similar phylogeny with *F. merguiensis* was monophyletic in BI tree and paraphyletic in ML tree (Figure 4.3). A total of 18,002 trees were sampled from both runs for the BI tree, and following discard of the burn-in set, 5,953 credible trees remained for calculation of posterior probabilities (Figure 4.3A). The final average standard deviation of split frequencies achieved was 0.015120, with an average potential scale reduction factor (PSRF) for parameter values of 1.00. PAV-based ML analysis produced trees with log likelihood scores of ln L = -52,413.995 (Figure 4.3B). The main difference between SNP-based and PAV-based analyses was the clade topology for *F. merguiensis* and *L. vannamei* in the PAV-based reconstructions, which was less resolved than the SNP-based reconstructions.

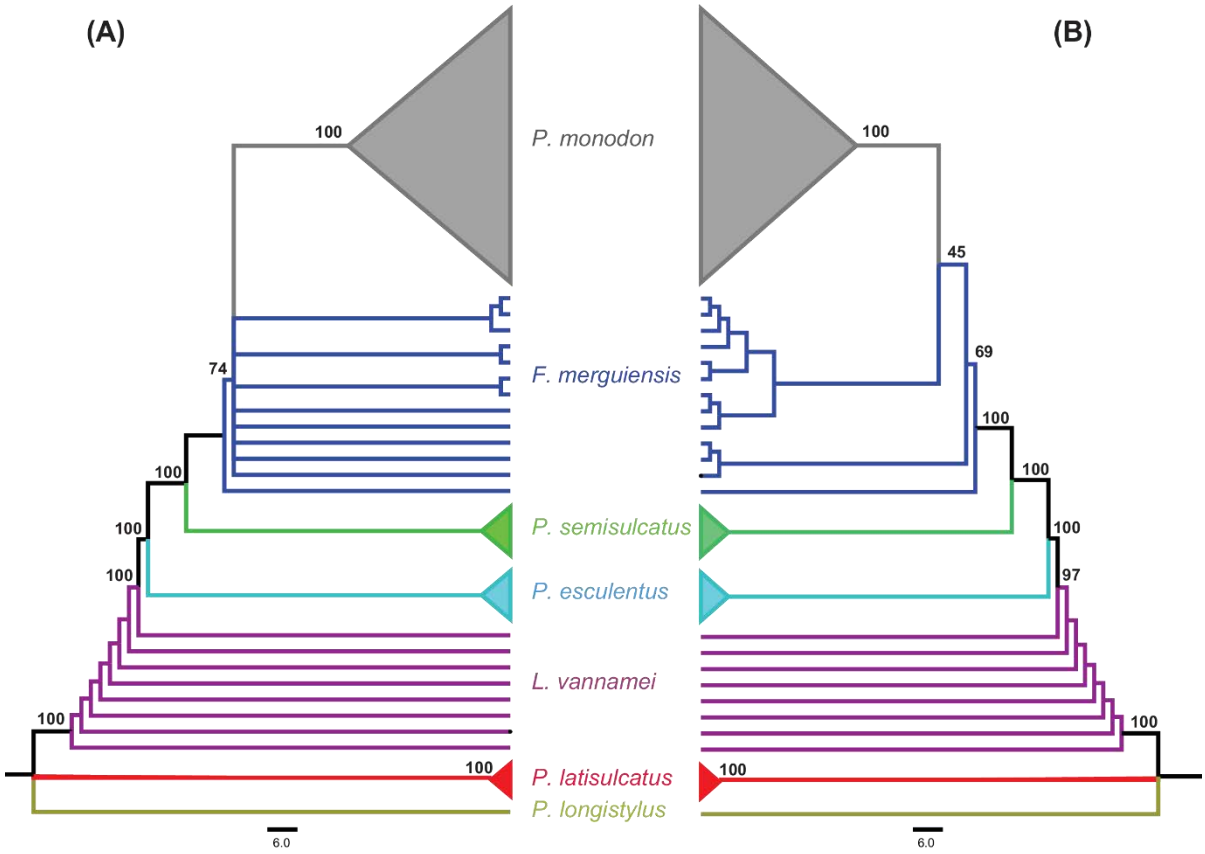


Figure 4.3 Phylogenetic relationships of seven Penaeidae species using 7,054 PAV loci. **(A)** Bayesian reconstruction generated in MrBayes v3.2.7 from 5,953 of the most credible set of trees. **(B)** The maximum likelihood reconstruction was conducted using IQ-TREE v2.1.0 under the GTR2+FO+R2 model with 10,000 bootstrap replicates. *Penaeus longistylus* was used as the outgroup. Numbers at nodes represent posterior probabilities and bootstrap values in percentage.

As a monophyletic taxa, all *P. monodon* individuals grouped into one clade (Posterior Probability, PP and BS = 100) with *F. merguiensis* being the closest related sister taxa (Figure 4.3). The clade containing these two species resolved to be genetically proximal to the *P. semisulcatus* clade (PP and BS = 100; Figure 4.3). One after another, the *P. monodon*, *F. merguiensis*, and *P. semisulcatus* clade resolved as the sister group to both *P. esculentus* and *L. vannamei* (PP = 100 and BS ≥ 97), respectively. Similar to SNP-based phylogenetic reconstructions, at the root of the tree *P. longistylus* and *P. latisulcatus* were recovered as sister taxa to the five remaining species (*P. monodon*, *F. merguiensis*, *P. semisulcatus*, *P. esculentus*, and *L. vannamei*; PP and BS = 100; Figure 4.3).

4.3.3 Phylogeographic relationships among Indo–Pacific *P. monodon* samples

To evaluate the relationships among individual Indo-Pacific *P. monodon* samples, phylogeographic analyses based on both SNP and PAV loci datasets (n = 4,496 and 7,054 respectively) were performed, with all *P. esculentus* and *P. semisulcatus* samples used as outgroup (Figure 4.4 and 4.5). ML analyses produced trees with log likelihood scores of ln L = -14,761.999 and ln L = -82,706.424 for SNP and PAV loci datasets, respectively. Moreover, *ModelFinder* identified KY+F+R2 and GTR2+FO+R4 as the best-fitting substitution models for the SNP and PAV loci dataset ML trees, respectively (Figure 4.4A and 4.5A). A total of 17,002 trees were sampled from both runs for the BI tree and, following discard of the burn-in set, 11,971 credible trees remained for calculation of posterior probabilities. The final average standard deviation of split frequencies achieved was 0.093129, with an average PSRF for parameter values of 1.004. Branch lengths were generally shorter among Australian and Fijian populations in all ML trees based on SNP and PAV loci datasets.

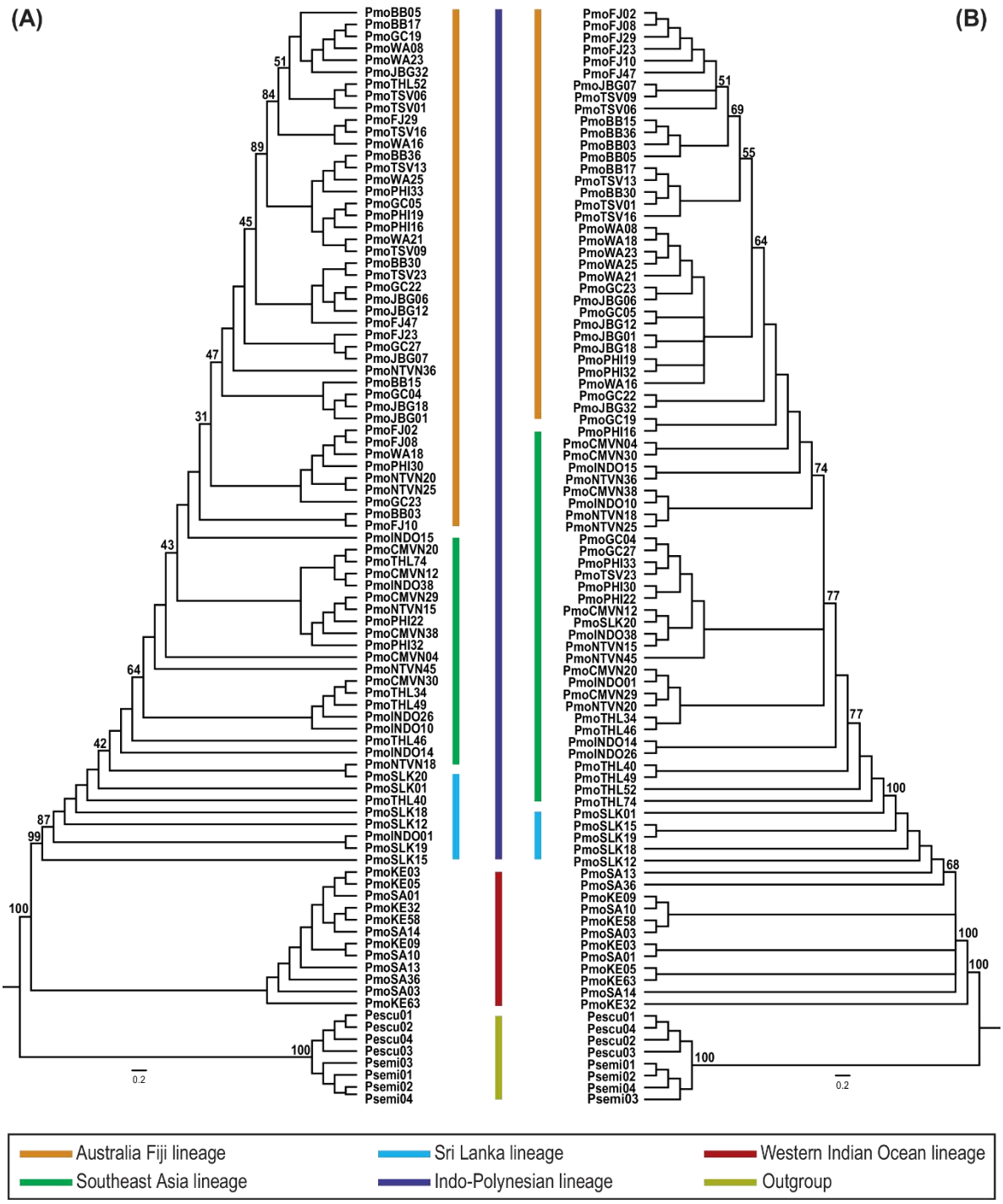


Figure 4.4 Phylogeographic analyses of *Penaeus monodon* using SNP and PAV loci datasets. Tree topology corresponds to the best ML tree. (**A)** ML tree generated in IQ-TREE v2.1.0 using 4,496 SNP loci under KY+F+R2 model. (**B)** BI tree generated in MrBayes using 7,054 PAV loci from 11,971 of the most credible set of trees. Numbers at nodes represent posterior probabilities and bootstrap values in percentage. Out-group Pescu: *Penaeus esculentus*; and Psemi: *Penaeus semisulcatus*. Legend for locality codes is the same as defined in Figure 4.1.

The phylogenetic trees constructed by all three independent analysis approaches (ML, BI, NJ) displayed similar topologies and showed geographic structure with strong support between the West Indian Ocean populations (WIO; containing SA and KE populations) and the Indo-Polynesian populations (containing the Pacific Ocean, Western Australia, and Sri Lanka populations) (Figure 4.4 and 4.5). All WIO *P. monodon* populations nested together in a WIO lineage, while all Pacific Ocean *P. monodon* populations clustered together with Sri Lankan individuals, which was located in the Northwest Indian Ocean in a single overarching Indo-Polynesian lineage (Figure 4.4 and 4.5). In PAV-based trees, although WIO populations did not form a separate clade, they still grouped together with moderate support (ML: BS = 25 - 100; BI: PP = 51 - 100). All ML, BI, and NJ trees also identified Sri Lanka as a transition site between the West Indian and Pacific Oceans.

As the largest clade, the Indo-Polynesian lineage was clearly divided between Sri Lanka and Southeast Asia to Australia and Fiji (Figure 4.4 and 4.5). A high degree of population mixing was observed across all SNP-based trees. More specifically, Sri Lanka nested with Southeast Asia and some Southeast Asian samples clustered with Australia and Fiji samples. The PAV-based trees revealed that most individuals from proximal localities tended to group together, except some individuals assigned to one location clustered with another locality (e.g., neighboring population). At a finer scale, relationships within the geographically disjunct Indo-Polynesian lineage (Sri Lanka, Southeast Asia, Australia, and Fiji) varied from low to high resolution.

Taken together, phylogeographic analyses of Indo-Pacific *P. monodon* samples based on SNP and PAV loci datasets strongly supports the existence of pronounced genetic divergence between WIO and the Indo-Polynesian (Sri Lanka, Southeast Asia, Australia, and Fiji) populations.

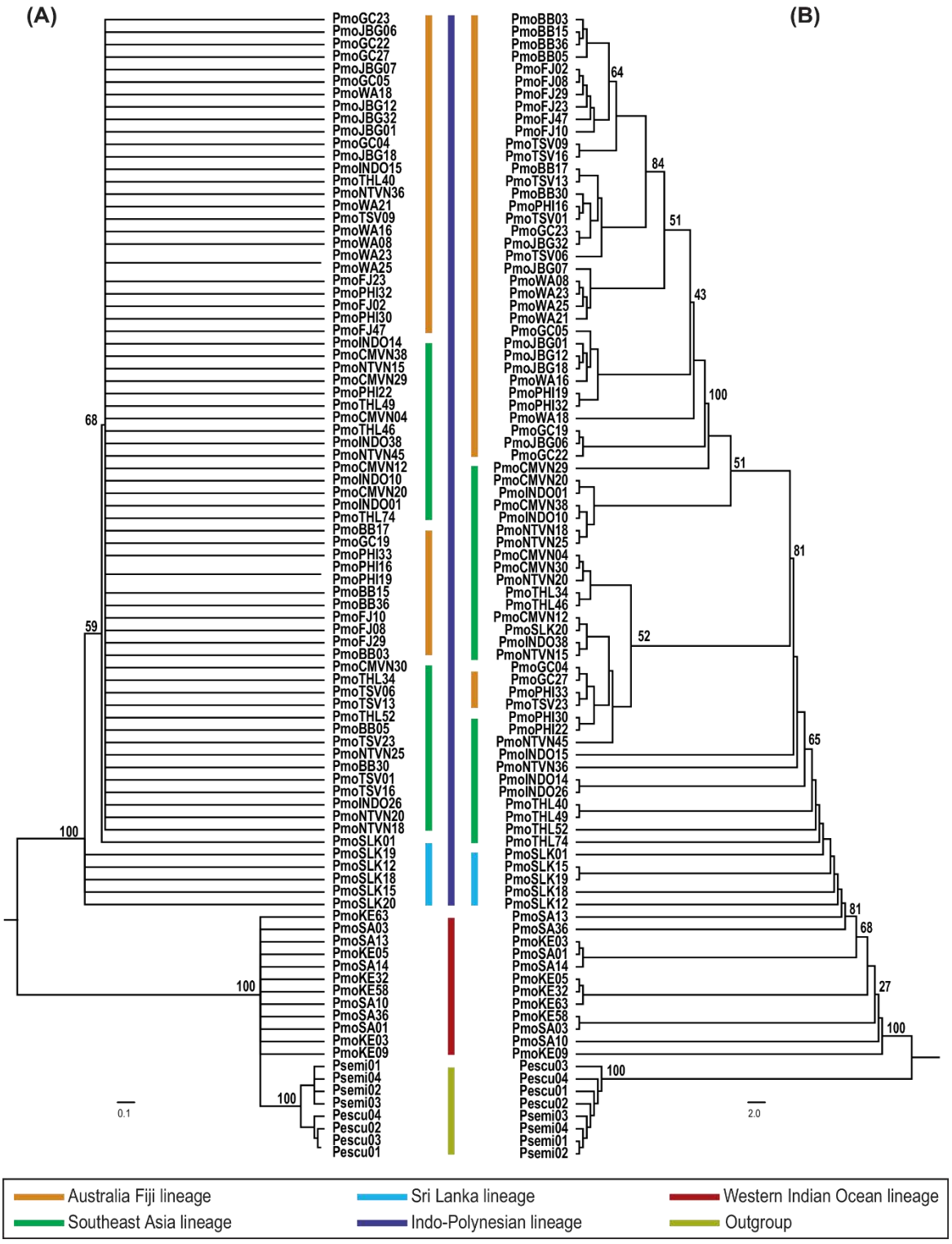


Figure 4.5 Phylogeographic analyses of *P. monodon* using SNP and PAV loci datasets. (**A**) NJ tree generated in MEGA-X using 4,496 SNP loci based on Nei's (1972) genetic distances. (**B**) ML tree generated in IQ-TREE v2.1.0 using 7,054 PAV loci under GTR2+FO+R4 model. Numbers at nodes represent posterior probabilities and bootstrap values in percentage. Both *P. esculentus* (Pescu) and *P. semisulcatus* (Psemi) were used as outgroup. All abbreviations are the same as defined in Figure 4.1.

4.4 Discussion

In this study, novel insights regarding the phylogenetic relationships among Indo–Pacific *P. monodon* and its sister taxa using two independent genome-wide marker datasets (SNP and PAV loci) were reported. Both SNP- and PAV-based phylogeographic structures for Indo-Pacific *P. monodon* demonstrated a clear pattern of regional connectivity among all Indo-Pacific *P. monodon* populations with two divergent lineages: 1) the Indo-Polynesian (containing Sri Lanka, Southeast Asia, Australia, and Fiji) and 2) the WIO (containing Kenya and South Africa). Among all seven Penaeidae species sampled, my SNP- and PAV-based phylogenetic analyses supported the monophyletic positioning of *P. monodon,* as well as the position of six sister taxa to *P. monodon*.

4.4.1 Phylogeny of seven Penaeidae species

Both SNP- and PAV-based phylogenetic analyses of seven Penaeidae species (Table 4.1) provided clear, consistent, and well-supported relationship patterns (Figure 4.2 and 4.3). More specifically, these seven species divided into three distinct and well-supported clades (BS and PP = 100), with the first clade containing *F. merguiensis*, *P. monodon, P. semisulcatus*, *P. esculentus*, and *L. vannamei*, while the second clade contained *P. latisulcatus* and the third clade contained *P. longistylus*. The presence of a close evolutionary genetic relationship between the *Fenneropenaeus*, *Penaeus*, and *Litopenaeus* genera within Penaeidae was similar to the results reported in other non-SNP genetic marker studies (Hurzaid et al., 2020; Lavery et al., 2004; Ma et al., 2009; Voloch et al., 2009), which grouped *Fenneropenaeus, Penaeus,* and *Litopenaeus* into one clade.

In the present study, the most notable difference between the SNP-based and PAV-based phylogenetic reconstructions was the specific placement of *F. merguiensis*. PAV-based reconstruction nested *F. merguiensis* between *P. monodon* and *P. esculentus* and *P. semisulcatus* (PP = 74 and BS = 69), while SNP-based reconstruction included *F. merguiensis* along with all three *Peneus* species in one clade. These differing patterns may have been due to the reduced species-level discriminatory power of PAV-based analyses compared to SNP-based analyses, which resulted in these more distal relationships being less resolved (Lal et al., 2018; Morse et al., 2018). This difference in the species-level discrimination power of PAV- versus SNP-based analyses might also explain the similar branch tip lengths observed for *F.*

merguiensis and *L. vannamei*. Moreover, as a genome-wide "dominant" marker, the PAV classification of "0" or "1" (i.e., absence or presence), respectively, can lead to a loss in total number of informative sites throughout the genome (Lischer et al., 2014; Morse et al., 2018). Additionally, Voloch et al. (2005) reported the nesting of five *Fenneropenaeus* species between the two *Penaeus* lineages (*P. monodon, P. esculentus,* and *P. semisulcatus*) with low support (NJ BS = N/A; ML BS = 44; BI PP = N/A) and suggested that future studies should aim to verify this result. Taken together, the reduced species-level discrimination power of PAV-based analyses appears to be the primary driver of the observed *F. merguiensis* position within Penaeidae clade presented herein.

Lavery et al. (2004), Voloch et al. (2005), and Voloch et al. (2009) reported that species within the genera *Melicertus* were paraphyletic which aligns with my SNP- and PAV-based phylogenetic analyses in that the two former species assigned to the genus *Melicertus* (*P. latisulcatus* and *P. longistylus*) were consistently placed together in separate clades with relatively strong support (BS and PP > 90). A recent study of Hurzaid et al. (2020) using four gene datasets (mitochondrial cytochrome oxidase I, mitochondrial 16S ribosomal RNA, NaK, and PEPCK) also presented that genera *Melicertus* (including *P. latisulcatus* and *P. longistylus*) appear to be non-monophyletic. Thus, the present study confirmed the non-monophyly of the genera *Melicertus*.

4.4.2 Phylogeographic relationships of Indo–Pacific *P. monodon* samples

In order to construct the phylogeographic relationships among Indo–Pacific *P. monodon* two independent and high-density genome-wide marker sets (SNP and PAV loci) were utilized for a representative sub-set of Indo-Pacific *P. monodon* samples (n = 84; Table 4.1; Figure 4.4 and 4.5). Both SNP- and PAV-based phylogenetic analyses revealed that two divergent lineages of *P. monodon* exist within the Indo-Pacific (BS = 100, PP = 51 – 100). Rather than directly conforming to ocean basins, the two lineages were the Indo-Polynesian (containing Sri Lanka, Southeast Asia, Australia, and Fiji) and the WIO (containing Kenya and South Africa). Similar phylogeographic patterns have recently been reported for other Indo-Pacific marine species, including: 1) bottlenose dolphin *Tursiops aduncus* and humpback whale *Sousa spp*. using mtDNA markers (Amaral et al., 2017), 2) spiny lobster *Panulirus homarus* using mitochondrial control region and microsatellite markers (Farhadi et al., 2017), and 3) dugong *Dugong dugon* using mtDNA markers (Plön et al., 2019). Additionally, the deep genetic division between *P.*

monodon from WIO and Indo-Polynesian lineages were indicated in Benzie et al. (2002) and Chapter 3 using mtDNA sequences and whole genome SNP sequencing, respectively. Several factors have been proposed as explanations of the genetic divergence of the WIO *P. monodon* populations, including life history characteristics, environmental transitions, large geographical distances, geographical barriers, and/or ocean surface currents (Amaral et al., 2017; Benzie et al., 2002; Huyghe & Kochzius, 2017; Vogler et al., 2012; Chpater 3).

In this study all ML, BI, NJ trees identified the WIO *P. monodon* as ancestral to the Indo-Polynesian *P. monodon*, with the Sri Lankan population being a transition zone between these ocean basins. In the Indo-Polynesian lineage, as a transition between the West Indian and Pacific Oceans, *P. monodon* from Sri Lanka is genetically similar to Southeast Asia and grouped together with medium to high moderate support. This result is consistent with the recent findings of Chapter 3, which reported three migration events from the WIO to Southeast Asia (Thailand, Vietnam, and Indonesia) via Sri Lanka with strong migration weights, as well as Sri Lanka forming a single cluster. However, two previous studies reported a divergent NJ reconstruction based on mtDNA haplotypes sequences (You et al., 2008) and microsatellites (Waqairatu et al., 2012) for Indo-Pacific *P. monodon*. While You et al. (2008) grouped Kenya and Madagascar in the WIO lineage and India in a separate lineage, Waqairatu et al. (2012) grouped *P. monodon* from Madagascar, India, and Sri Lanka into one lineage. Differences between the genome-wide SNP dataset herein and the findings reported by these previous studies may simply be due to the level of genetic resolution provided by genome-wide SNPs. Lemopoulos et al. (2019) recently demonstrated that SNP-based analyses were more informative than microsatellite-based analyses for applications that require individual-level genotype information, such as estimating genetic diversity and relatedness for small populations with low sample sizes. Future studies should further elucidate the degree of genetic divergence among *P. monodon* populations within the Indian Ocean by undertaking extensive sampling within this particular region of their Indo-Pacific distribution.

Within the Indo-Polynesian lineage SNP-based phylogenies revealed that *P. monodon* from Western Australia (located in the Eastern Indian Ocean) nested within other Australian and Fijian *P. monodon* populations (located in the western Pacific Ocean). This is despite both Benzie et al. (2002) and Waqairatu et al. (2012) reporting that Western Australia *P. monodon* was genetically distinct from all other Australian *P. monodon* populations based on mitochondrial and microsatellite datasets. Chapter 2 demonstrated that Western Australia *P. monodon* are genetically distinct from all other Australian *P. monodon* populations, but that

Western Australia *P. monodon* do share an evolutionary ancestry with Northern Australia *P. monodon*. Additionally, a high degree of population mixing was also observed across SNP-based trees, which supports the presence of an admixture zone in the Java Sea and the Strait of Malacca within the Pacific Ocean. This finding is consistent with the ADMIXTURE and TreeMix analysis reported in chapter 3.

4.5 Conclusion

In this study, both genome-wide single nucleotide polymorphism (SNP) and presence-absence variant (PAV) loci datasets were used to determine the evolutionary relationships of Indo–Pacific *P. monodon* and validate sister taxa relationships of species within the Penaeidae. SNP- and PAV-based NJ, ML, and BI phylogenetic analyses in the present study confirm the monophyletic structure of *P. monodon*, as well as the position of six sister taxa to *P. monodon*. Both SNP- and PAV-based phylogeographic structures of Indo-Pacific *P. monodon* revealed that Indo-Pacific *P. monodon* appear to be divided into two lineages: 1) Indo-Polynesian (containing Sri Lanka, Southeast Asia, Australia, and Fiji) and 2) Western Indian Ocean (containing Kenya and South Africa). All ML, BI, NJ trees identified the WIO *P. monodon* populations as ancestral to the Indo-Polynesian *P. monodon*, with Sri Lankan *P. monodon* being a transition zone between these geographic regions. Future studies should consider undertaking a comprehensive sampling of wild *P. monodon* within the Indian Ocean in order to increase the resolution of the phylogeographic structures reported here and elucidate the degree of and potential reasons for genetic divergence among populations within the Indian Ocean.

Chapter 5. Development of a global SNP resource for diversity, provenance, and parentage analyses in the Indo-Pacific giant black tiger shrimp (*Penaeus monodon*)

5.1 Introduction

The giant black tiger shrimp (*Penaeus monodon*) is an important aquaculture and fishery species and is the second most farmed Penaeid shrimp species globally. Between 2010 and 2018 *P. monodon* farm production increased from 562,900 to 750,000 tonnes; however, its overall proportion of global crustacean aquaculture has decreased over this time period from 10% to 8% (FAO, 2020). This observed trend is mostly attributed to the comparatively large increase in farm production of whiteleg shrimp (*Litopenaeus vannamei*), which more than doubled from 2,648,500 tonnes in 2010 to 4,966,200 tonnes in 2018 (FAO, 2020). The rapid growth of the global *L. vannamei* industry has been attributed, in large part, to the success achieved in the domestication, selective breeding, and translocation (i.e., introduced to Asia) of this species (Briggs, 2005; Moss et al., 2012). In contrast, the increase in *P. monodon* is limited by the lack of genetically improved stocks, inconsistent seedstock quality, and/or numerous devastating disease outbreaks (Guppy et al., 2020; Nguyen et al., 2020). Consequently, research priorities within the *P. monodon* industry are now focusing on the development of breeding programs to address these issues and increase overall production. Access to informative genetic marker suites is an integral resource for breeding programs as well as fisheries since they permit: 1) accurate assessment of genetic diversity in wild and farmed populations, 2) accurate tracing of broodstock/individuals back to their source population, 3) accurate parentage assignment throughout the hatchery and production systems

The giant black tiger shrimp (*Penaeus monodon*) is widely distributed throughout the Indo-Pacific from the Eastern coast of Africa, the Arabian Peninsula, Southeast Asia, the North Pacific Ocean, South Pacific islands, and Australia (Alfaro-Montoya et al., 2015; Brooker et al., 2000; Waqairatu et al., 2012). As an important aquaculture species, *P. monodon* has been introduced to establish new aquaculture products in many countries outside their natural distribution, including Americas (e.g., USA, Ecuador, Costa Rica, Colombia, Brazil; (Funge-Smith & Briggs, 2003; Wakida-Kusunoki et al., 2016)) and West Africa (Zink et al., 2018). The

introduction or importation of wild *P. monodon* broodstock is also commonly practised among the major producing countries in Southeast Asia because local supplies are insufficient and domestication technology has not yet been commercially developed (FAO, 2005-2021). The escape of *P. monodon* individuals from aquaculture farms has been reported in several countries (Xu et al., 2001; Aguirre-Pabón et al., 2015; Alfaro-Montoya et al., 2015; de Souza Junior et al., 2015; Sandoval et al., 2014), which led to detectable genetic changes in wild populations (Xu et al., 2001; de Souza Junior et al., 2015; Sandoval et al., 2014; Zink et al., 2018). Taken together, new genetic tools are needed for monitoring genetic diversity, population traceability (i.e., provenance), and parentage assignment (i.e., pedigree) within and among global Indo-Pacific *P. monodon* populations.

Over the past few decades, there have been major advances in the genetic tools available for aquaculture and fishery management that have led to significant improvements in efficiency (Robledo et al., 2018). Single nucleotide polymorphisms (SNPs) have recently become the predominant molecular markers utilized in aquaculture research (e.g., genotyping) because they are the most prevalent type of polymorphism throughout the genome, exhibit co-dominant inheritance, and permit low-cost genotyping per locus (Flanagan & Jones, 2019; Wenne, 2018; Zenger et al., 2019). Following genome-wide SNP genotyping (e.g., Chapter 3), informative marker subsets (i.e., panels) can be identified for rapid, robust, and cost-effective use in diversity, provenance, and/or pedigree determination (Dierens et al., 2014; Guppy et al., 2020; Henshall et al., 2014; Sellars et al., 2014). For example, Duarte et al. (2013) reported that the cost of genotyping pigs with a high-density commercial SNP panel (n = 60,000) was more than twice that of genotyping the same individuals with a low-density commercial SNP panel (n = 9,000). One way to reduce SNP genotyping costs is to utilize low-density SNP panels (n < 10,000) instead that have the same analytical power in associated genetic analyses (Kriaridou et al., 2020; Saint-Pé et al., 2019; Sukhavachana et al., 2021). For example, Kriaridou et al. (2020) showed that similar results of genetic selection were obtained when a low-density SNP panel (n = 1,000 – 2,000) was compared to a high-density SNP panel (ranging from 12,000 – 27,000 SNPs) across four aquaculture species (*S. salar*, *Cyprinus carpio*, *Sparus aurata*, and *Crassostrea gigas*).

Determination of diversity, provenance, and pedigree are the three most commonly implemented analyses within aquaculture systems (Holman et al., 2017; Robledo et al., 2018; Symonds et al., 2019). Genetic diversity evaluations of wild populations are of key importance when developing sustainable breed improvement strategies because high genetic diversity

within founder/base populations is essential (D'ambrosio et al., 2019; Lind et al., 2019; Mastrochirico-Filho et al., 2019). The fine-scale and broad-scale genetic diversity of Australian and Indo-Pacific populations have been examined using more than ten thousand high-quality SNP markers (Chapter 3 and 4), respectively; however, use of this entire dataset as a high-density SNP panel (n = 10,593; Chapter 4) would not be cost-effective for commercial applications as it contains more than the minimum number of markers required for accurate genetic diversity determination. Provenance determination is necessary for monitoring the source of aquaculture stocks, assessing impacts of farm escapees on wild stocks, and allow testing of marketed origin labels on products for food quality and safety purposes (Leal et al., 2015; Ogden, 2011). For example, low-density SNP panels (n = 58 and 94) have recently been demonstrated to provide cost-effective and reliable provenance determination in, for example, Florida bass *Micropterus floridanus* (Zhao et al., 2018) and Atlantic salmon *S. salar* (Holman et al., 2017), respectively. Pedigree determination minimizes inbreeding and improves breeding value estimates, which collectively allow for better decisions in selective breeding programs (Nazari & Pourkazmi, 2021; Xu et al., 2017). Accurate pedigree determination also allows fishery managers to track broodstock sources, minimize broodstock inbreeding, and estimate insightful genetic parameters (Liu et al., 2016), which are all especially important when broodstocks are sourced from the wild. In a review of 58 publications that used SNP-based parentage analyses, Flanagan and Jones (2019) concluded that a relatively small number of SNP markers (e.g., 100 – 500) could completely resolve pedigree in most situations. For *P. monodon*, two low-density SNP panels (n = 112, (Sellars et al., 2014); and n = 4,194, (Guppy et al., 2020)) have been developed for pedigree analysis; however, these panels were designed for utilization within the Australian industry only. To date, no low-density SNP panels have been identified that permit diversity, provenance, and/or pedigree determination in all Indo-Pacific *P. monodon* populations despite the industry desire for advanced selective breeding programs, ability to accurately trace the source of commercial products, and robust analysis standards.

Using the previously described Indo-Pacific *P. monodon* DArTseq™ SNP dataset (Chapter 4) this chapter identified low-density SNP panels that permit accurate diversity, provenance, and pedigree determination within and among all Indo-Pacific *P. monodon* populations. Low-density SNP panels were targeted to minimize costs associated with sequencing (e.g., fewer reactions; (Gebrehiwot et al., 2021; Herry et al., 2018; Song & Hu, 2021) and computation (e.g., quality assurance (QA), quality control (QC), and analyses; (Dominik et al., 2021) as well as potentially provide the following long-term advancements to

the Indo-Pacific *P. monodon* aquaculture industry: 1) improved consistency in SNP genotype data across global studies (e.g., direct comparison across common SNP panels as opposed to indirect comparison between uncommon SNP panels or between SNP and microsatellite panels), 2) development of a global reference database of genotypes collected from wild and farmed populations without the need for additional and expensive genome-wide sequencing, and 3) establishment of robust global genetic diversity, provenance, and pedigree analysis standards. This study identified discrete low-density SNP panels for provenance and pedigree determination because provenance determination benefits from using SNP markers with high population differentiation while pedigree determination benefits from using SNP markers with high minor allele frequencies (Anderson, 2010; Heaton et al., 2014; Holman et al., 2017; May et al., 2020). The identified low-density SNP panels for diversity (n = 1) and provenance (n = 1) determination were tested for accuracy using the previously described high-quality Indo-Pacific *P. monodon* SNP dataset (n = 10,593; Vu, et al., 2021) while the low-density SNP panels for region-specific pedigree determination (n = 3) were tested *in silico* using the larger raw Indo-Pacific *P. monodon* SNP dataset (n = 133,929; Vu, et al., 2021).

5.2 Methods

5.2.1 Tissue sample collection and genomic 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 Southwestern coast of the African continent (Figure 5.1). This study obtained a low coverage ddRAD genotype by sequencing (DArTseq™; n = 131,929) for each of 520 *P. monodon* individuals from 16 discrete Indo-Pacific Ocean locations (Figure 5.1, Table 3.1) as described in Chapter 3. The information of sampling locations, numbers of individuals, and coordinates for each sampled Indo-Pacific *Penaeus monodon* population were described in Chapter 3. These samples did not cover the entire global distribution of *P. monodon*, however, because I was unable to get *P. monodon* samples from the North Pacific Ocean (e.g., China, Taiwan, and Japan).

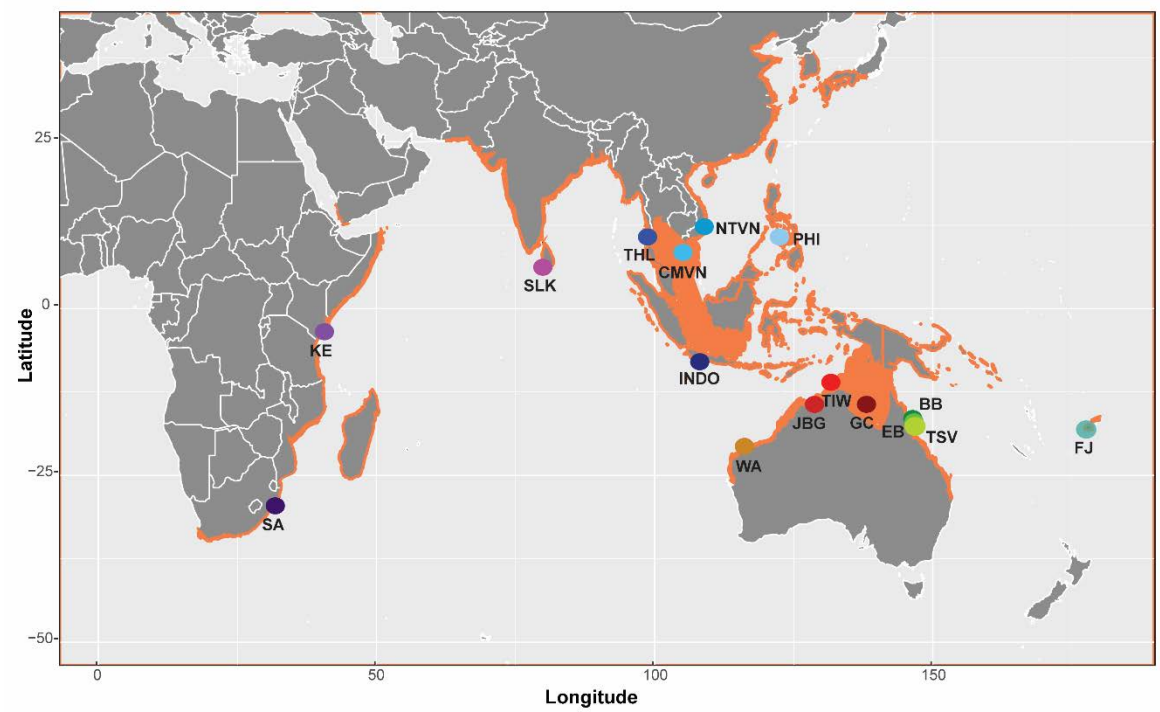


Figure 5.1 Sampling sites and the known distribution of Indo-Pacific *Penaeus monodon* (modified from Kongkeo (2005–2020)) sourced in this study. FJ: Fiji; Eastern Australia (EA; BB: Bramston Beach, EB: Etty Bay; TSV: Townsville); Northern Australia (NA; GC: Gulf of Carpentaria, JBG: Joseph Bonaparte Gulf, TIW: Tiwi Island); Western Australia (WA); Vietnam (VN; 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.

5.2.2 SNP genotyping and quality control

Raw SNP sequence data (n = 131,929) was obtained from DArT after initial SNP calling but without filtering by proprietary analysis pipelines (DArTsoft). Due to the five discrete SNP panels needed for genetic diversity (n = 1), provenance (n = 1), and parentage assignment (n = 3) the raw SNP dataset was subjected to two independent filtering approaches, each of which used custom *dartQC* python scripts (available at https://github.com/esteinig/dartqc; Figure 5.2). Full details regarding how the high-quality SNP dataset (n = 10,593) was obtained for each Indo-Pacific *P. monodon* (n = 520) from the raw DarTseq™ dataset can be found in Chapter 3 and below (Sections 2.3. -2.5).

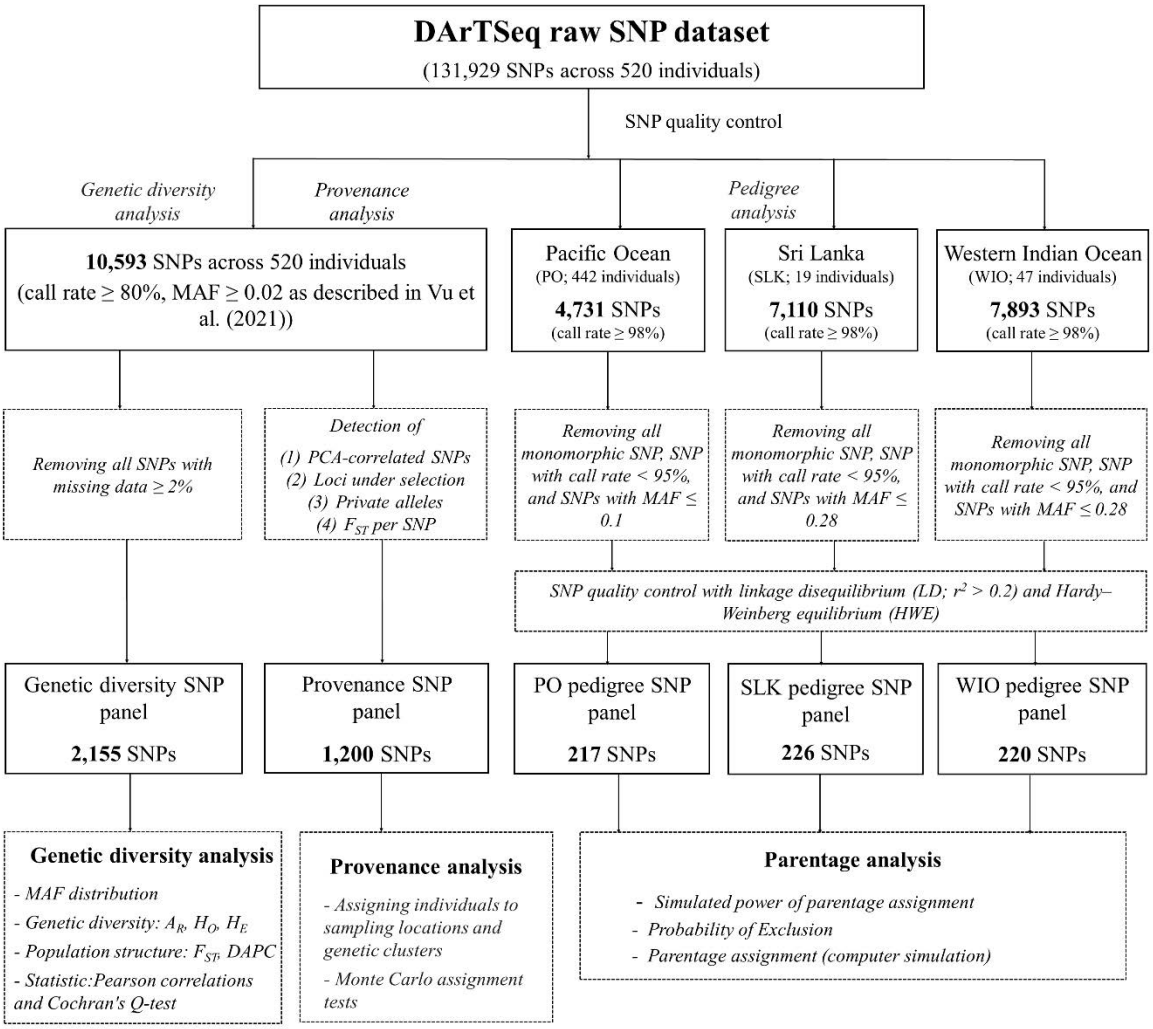


Figure 5.2 SNP filtering workflow from raw (n = 131,929) to SNP panels for determining genetic diversity, provenance, and pedigree analyses of *Penaeus monodon*. MAF: minor allele frequency, A_R: mean allelic richness, H_O: mean observed heterozygosity, H_E: mean expected heterozygosity, F_{ST}: pairwise population difference value.

5.2.3 Diversity SNP panel

Diversity SNP panel for Indo-Pacific *P. monodon* was based on targeting SNPs with the most informative and repeatable loci for allele frequency determination (i.e., lowest missing data; (Fu et al., 2014)). Furthermore, unbiased SNP allele frequencies were required to reflect the original high-density SNP dataset (Chapter 4) allele frequency spectrums in each population. To identify a robust and broadly applicable diversity SNP panel the high-quality SNP dataset for Indo-Pacific *P. monodon* (n = 10,593; Chapter 4) was mined. In brief, the raw DArTseq™ dataset (n = 131,929 SNPs) obtained for each Indo-Pacific *P. monodon* individual collected from 16 discrete Indo-Pacific sampling sites (n = 532 total) was initially filtered using custom

dartQC python scripts (read counts < 7, average repeatability < 90%, call rate < 80%, retention of one SNP from sequence similarity clustering at > 0.95, and MAF < 0.02). The remaining SNPs (n = 12,415) and individuals (n = 520) were tested for linkage disequilibrium (LD; r^2 > 0.2) and deviation from Hardy–Weinberg equilibrium (HWE), which yielded a high-quality SNP dataset consisting of 10,593 SNPs for each retained *P. monodon* individual (n = 520; 12 individuals removed due to high missing data). Given missing data introducing a large bias in the estimation of allelic richness (Fu et al., 2014), the high-quality SNP dataset (n = 10,593) was then filtered for the most informative and repeatable loci by removing all those with missing data ≤ 2% per SNP to identify the final diversity SNP panel (n = 2,155 SNPs; Figure 5.2).

To confirm the minimum number of SNP markers necessary for accurate and consistent genetic diversity determination three smaller subsets (n = 1,500, n = 1,000, and n = 500) were populated with randomly selected loci from the full diversity SNP panel (Figure 5.2). All four subsets were then used to estimate minor allele frequency (MAF) distribution and genetic diversity factors for all Indo-Pacific *P. monodon* individuals and populations (n = 520, n = 16 respectively) to compare: 1) MAF distribution metrics, 2) genetic diversity metrics, and 3) population genetic differentiation metrics.

For consistency, this study used the same genetic diversity metric estimation methodology described in Chapter 3 to estimate genetic diversity factors. In brief, the R package *adegenet* (Jombart, 2008) was used to calculate MAF for each of the full diversity SNP panel and its three subsets while key diversity estimates (H_O, H_E, and F_{IS}) were computed for each of the four panel/subsets using the *divBasic* function of R package *diveRsity* 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) for each of the four panel/subsets using rarefaction to avoid sampling size bias while pairwise genetic differentiation values (F_{ST}) and their significance between populations were calculated for each of the four panel/subsets according to Wright (1949) and Weir and Cockerham (1984) using the R package *StAMPP* version 1.5.1 (Pembleton et al., 2013).

The full diversity SNP panel and its three subsets were then assessed against the entire high-quality SNP dataset (n = 10,593; Chapter 3) using discrete Pearson correlations, which were each calculated using the R package *ggpubr* (Kassambara, 2018) to ensure that key diversity estimates were not skewed. Heterogeneity was also assessed between each of the full diversity SNP panel and its three subsets, and the entire high-quality SNP dataset (n = 10,593)

based on *p*-values from discrete Cochran's Q-tests (Cochran, 1954) determined using the R package *nonpar* (Sweet, 2020).

5.2.4 Provenance SNP panel

Provenance determination for Indo-Pacific *P. monodon* was based on targeting SNP markers with high population differentiation capability (Anderson, 2010; May et al., 2020). Similar to the diversity SNP panel, the provenance SNP panel was also identified from the high-quality SNP dataset (n = 10,593; Chapter 4). Following initial filtering, pairwise genetic differentiation values (F_{ST}) were calculated for each SNP in the entire high-quality dataset (n = 10,593) using *basic.stats* in the R package *hierfstat* version 0.5.7 (Goudet, 2005). Three subsets were then identified: 1) outlier SNPs using two independent population differentiation (PD) approaches: *BayeScan* version 2.1 (Foll, 2012; Foll & Gaggiotti, 2008) and R package *PCAdapt* (Luu et al., 2017) as described in Chapter 3; 2) SNPs that captured population structure in Principal Component Analysis (PCA) using Discriminant Analysis of Principal Components (DAPC) in the R package *adegenet* version 2.1.1 (Jombart, 2008); and 3) private alleles using HP-RARE version 1.1 (Kalinowski, 2005). The final provenance SNP panel was determined by merging all three subsets together, removing all duplicate loci, and removing all SNPs with pairwise F_{ST} value ≤ 0.1.

The first provenance test involved two independent analyses to determine the effectiveness of the identified provenance SNP panel for tracing individuals back to their original population with and without *a priori* assumptions. First, each *P. monodon* individual was assigned to one of the seven previously determined Indo-Pacific genetic clusters (i.e., with *a priori* population assumptions; Chapter 4) using DAPC analysis based on the entire provenance SNP panel (Figure 5.2) in the R package *adegenet* version 2.1.3 (Jombart, 2008). The seven Indo-Pacific genetic clusters identified in Chapter 3 were: 1) Fiji (FJ); 2) Eastern Australia (EA; containing BB: Bramston Beach, EB: Etty Bay, and TSV: Townsville); 3) Northern Australia (NA; containing GC: Gulf of Carpentaria, JBG: Joseph Bonaparte Gulf, TIW: Tiwi Island); 4) Western Australia (WA); 5) Southeast Asia (SEA; containing CMVN: Ca Mau, NTVN: Nha Trang, PHI: Philippines, INDO: Indonesia, and THL: Thailand); 6) Sri Lanka (SLK); and 7) West Indian Ocean (WIO; containing KE: Kenya, SA: South Africa). Second, each individual was assigned to one of the 16 sampling locations (Figure 5.1) based on individual genomic relationships (i.e., without *a priori* population assumptions) using the R

package *NetView* pipeline (Neuditschko et al., 2012; Steinig et al., 2016) with multiple K-nearest neighbor (kNN) values between 10 and 100.

The second provenance test utilized a Monte Carlo analysis to determine the effectiveness of the identified provenance SNP panel for tracing individuals back to their original basin (either Pacific Ocean: PO or Indian Ocean: IO), original region (one of seven genetic clusters; Chapter 4), and original sampling location (one of eight PO genetic units or one of three IO genetic units) using the R package *assignPOP* v1.2.2 (Chen et al., 2018). The PO basin contained 13 populations (FJ, BB, EB, TSV, GC, JBG, TIW, WA, PHI, CMVN, NTVN, INDO, THL; $n = 459$ total; Table 3.1) from five genetic clusters (FJ, EA, NA, WA, and SEA; Chapter 4), which were divided into eight genetic units for population traceability analyses: 1) FJ, 2) EA (containing BB, EB, and TSV), 3) NA (containing GC, JBG, TIW), 4) WA, 5) PHI, 6) VN (containing CMVN and NTVN), 7) INDO, and 8) THL. The IO basin contained three populations (SLK, KE, and SA; $n = 66$ total) from two genetic clusters (SLK and WIO; Chapter 4), which were divided into three genetic units for provenance analyses: 1) SLK, 2) KE, and 3) SA. To assess ocean basin traceability, 5 - 10 individuals from each of the PO populations were randomly selected ($n = 99$ total) and 15 individuals from each of the IO populations were randomly selected ($n = 45$ total). To assess region traceability, 30 individuals were randomly selected from the FJ, EA, NA, SEA, and WIO clusters while 19 and 22 individuals were selected from the SLK and WA clusters due to their lower sample size, respectively. To assess sampling location traceability, 30 individuals were randomly selected from each PO unit except WA and PHI, which included all individuals ($n = 22$ and 12), respectively ($n = 214$ total) while all individuals were selected from each IO unit ($n = 66$ total).

The Monte Carlo analysis in the R package *assignPOP* v1.2.2 (Chen et al., 2018) was based on Monte Carlo cross-validations (the function *assign.MC*) followed by PCA to evaluate assignment accuracy and membership probabilities. To determine the minimum number of SNP markers necessary for accurate and consistent provenance analysis, I tested for three proportions of individuals in the training group (50%, 70%, and 90%) and fourth different proportions of loci ranked in decreasing F_{ST} values (25%, 50%, 75%, and 100%) for three levels of population assignment in this study. Each traceability test was replicated 30 times with the support vector machine (SVM) model used to build predictive models.

5.2.5 Pedigree SNP panels

Pedigree determination for Indo-Pacific *P. monodon* was based on targeting SNP markers with high probability of exclusion (PE) (Liu et al., 2017) and MAF (Heaton et al., 2014; Holman et al., 2017; May et al., 2020) values (i.e., across and within all populations) thus assignment accuracy increases when both values increase. Unlike the diversity and provenance SNP panels, the pedigree SNP panels were identified from the raw DArTseq dataset (*n* = 131,929 SNPs) for each Indo-Pacific *P. monodon* individual except those from the PHI population (*n* = 12 individuals removed). Retained Indo-Pacific *P. monodon* individuals (*n* = 508 total) were separated into three genetic groups (PO, SLK, and WIO) based on the genetic divergence reported in Chapter 3. The PO group contained 12 populations (FJ, BB, EB, TSV, GC, JBG, TIW, WA, CMVN, NTVN, INDO, THL; *n* = 447 total; Table 3.1), which were divided into the following seven genetic units for pedigree analyses: 1) FJ, 2) EA (containing BB, EB, and TSV), 3) NA (containing GC, JBG, TIW), 4) WA, 5) VN (containing CMVN and NTVN), 6) INDO, and 7) THL (Table 5.2). The SLK group contained only one population (SLK, *n* = 19; Table 3.1), which was considered its own genetic unit (SLK) for pedigree analyses (Table 5.2). The WIO group contained two populations (KE and SA, *n* = 47 total; Table 3.1), which were grouped into one genetic unit (WIO) for pedigree analyses (Table 5.2).

Initial filtering was run independently for the PO, SLK, and WIO groups using custom *dartQC* python scripts (available at https://github.com/esteinig/dartqc). First, genotype silencing was conducted where a genotype call made with < 7 read counts was silenced to minimise potential genotyping errors followed by removal of all SNPs with an average repeatability < 90% and call rate < 98%. SNPs with minor allele frequency (MAF) ≤ 0.1 were then removed from PO group individuals while SNPs with MAF ≤ 0.28 were removed from SLK and WIO group individuals (Table 5.2). All retained SNPs were then subjected to a quality control filter to ensure linkage disequilibrium (LD; $r^2 > 0.2$) and Hardy-Weinberg equilibrium (HWE). Finally, CD-HIT was utilised to cluster the allele sequences of all retained SNPs with 95% sequence identity, of which only the SNP with the highest MAF was retained. This filtering pipeline retained 4,731 SNPs, 7,110 SNPs, and 7,893 SNPs for the PO, SLK, and WIO groups, respectively (Figure 5.2; Table 5.2).

Following these initial filtration steps, metrics including MAF, expected heterozygosity (H_E), observed heterozygosity (H_O), and missing data per SNP were calculated for each genetic unit (*n* = 10; Table 5.2) using the R package *adegenet* (Jombart, 2008). All monomorphic loci,

SNPs with call rate < 0.95, and SNPs with MAF ≤ 0.1 (PO group) or MAF ≤ 0.28 (SLK and WIO groups) were then removed from each genetic unit. Of note is that the MAF threshold used for the PO group was lower than that used for SLK and WIO groups because this resulted in each genetic group having a similar number of SNP markers in their pedigree SNP panel. Given that the PO and WIO groups contained multiple genetic units (n = 7 and 2), respectively, the retained SNPs from each genetic unit were merged to create one pedigree SNP panel for each group. This identified 218, 433, and 233 SNP markers for the PO, SLK, and WIO genetic groups, respectively (Table 5.2). These SNP markers were then tested for LD ($r^2 > 0.2$) using *PLINK 1.9* (Purcell et al., 2007) and population-level deviation from HWE using Arlequin version 3.5 (Excoffier et al., 2005) before running an exact test with 10,000 steps in the Markov Chain and 100,000 dememorizations (Figure 5.1). For SNP pairs in LD, the SNP with the lower call rate value was removed. Likewise, if a SNP significantly deviated from HWE in all populations ($p < 0.0001$) it was removed from further analyses. This identified the final pedigree SNP panel for the PO, SLK, and WIO groups (n = 217, 226, and 220), respectively (Figure 5.2).

Indo-Pacific *P. monodon* individuals from across the PO, SLK, and WIO groups (n = 508) were then used as the theoretical broodstock (F0) for *in silico* validations. Given the close genetic relatedness between KE and SA (Chapter 4) these two populations were merged together into one WIO unit for *in silico* validations. In order to determine the number of SNP markers required to confidently determine pedigree for individuals within a theoretical *P. monodon* breeding program, 508 individuals were divided into nine genetic units (seven PO, one SLK, and one WIO) and progeny genotypes were simulated *in silico* independently for each genetic unit. A total of 18 individuals (least common denominator) were then randomly selected from each of the nine genetic units and set as candidate parents (F0) used to simulate F1 generation genotypes. These 18 individuals were randomly assigned as nine sires and nine dams and randomly paired at a 1:1 ratio to create 18 unrelated full-sib families containing 100 F1 progenies each. Progeny genotypes were simulated *in silico* based on parental genotypes (Massault et al., 2021) or using custom R scripts (Appendix 5.1).

To confirm the accurate determination of parent-offspring pairs over multiple generations, the EA genetic unit was used to simulate 10 generations (F1 to F10) using only 30, 35, 40, and 50 randomly selected SNPs from the PO pedigree SNP panel (n = 217; Figure 5.2). To simulate progeny genotypes across 10 generations, a pool of 50 progenies were randomly selected from the previous generation (starting with F1 progenies from the initial simulation),

randomly assigned as sires or dams (n = 25 each), and randomly paired at a 1:1 ratio, and then simulated out to generation F10. Each single-pair family produced 100 progenies for the subsequent generation and pedigree records were retained. The simulation of progeny genotypes over these 10 generations were replicated 10 times with all computations undertaken using custom R scripts (Appendix 5.1).

In silico simulated datasets were then assessed for pedigree determination accuracy using exclusion- and likelihood-based approaches in CERVUS version 3.0.7.7 (Kalinowski et al., 2007). To assign parent-pairs to progenies, CERVUS uses multi-locus parental exclusion probabilities and pairwise likelihood, which consists of three continuous modules: (1) allele frequency analysis, (2) simulation of pedigree analysis, and (3) pedigree analysis. For the exclusion-based approach, CERVUS was used to determine the ability of polymorphic SNP markers to distinguish first-degree relatives based on probability of exclusion (PE) calculations. These PE calculations were performed in the first CERVUS module using multiple (n = 14) incremental SNP marker subsets (n = 5, 10, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, and 100), which were each generated ten times (i.e., 10 replicates), to test three independent scenarios for each of the three pedigree SNP panels (Figure 5.2). Scenario 1 (PE1) calculated PE for each incremental SNP marker subset derived from the PO, SLK, and WIO pedigree SNP panels (n = 217, 226, and 220), respectively, following manual high-to-low MAF-ranking. Scenario 2 (PE2) calculated PE for each incremental SNP marker subset derived from the PO, SLK, and WIO pedigree SNP panels (n = 217, 226, and 220), respectively, following random MAF-ranking by R package *dplyr* version 1.0.7 (Wickham et al., 2021). Scenario 3 (PE3) calculated PE for each incremental SNP marker subset derived from the filtered polymorphic datasets (MAF ≤ 0.01 and missing data per SNP ≥ 5%) for each of the nine genetic units (FJ, n = 1,129; EA, n = 1,472; NA, n = 1,732; WA, n = 1,026; VN, n = 1,955; INDO, n = 1,933; THL, n = 1,585; SLK, n = 2,668; and WIO, n = 1,287; Table 5.2) following random MAF-ranking by R package *dplyr* version 1.0.7 (Wickham et al., 2021).

The likelihood-based approach was also undertaken in CERVUS version 3.0.7.7 (Kalinowski et al., 2007) using multiple (n = 14) incremental SNP marker subsets (n = 5, 10, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, and 100) of PO, SLK, and WIO pedigree SNP panels, which were each generated five times (i.e., five replicates) by R package *dplyr* version 1.0.7 (Wickham et al., 2021). More specifically, CERVUS calculates the pairwise parentage likelihood of each progeny with each possible dam, sire and parent-pair trio at each SNP to determine the threshold log-likelihood (LOD) scores for the true parental pair by simulating the

parents and progeny. Candidate sires and dams were set to 25 each and the proportion of candidate parents were set to 100% with the proportion of loci typed derived from allele frequency analysis. A total of 10,000 offspring simulations were run for each tested population with a default genotyping error rate of 0.01 and default confidence levels set to 95% and 80% for strict and relaxed confidence, respectively. Lastly, the pedigree analysis module assigned the most likely candidate parent-pair to each offspring tested with a pre-determined population wide assignment confidence set to 95% for strict confidence.

5.2.6 Genomic distribution of SNP panels

All five SNP marker panels (Figure 5.2) were assessed individually and together (n = 3,433 common SNP markers) for their location within the chromosome-scale genome assembly of *P. monodon* (Uengwetwanit et al., 2021) using Geneious Prime version 2021.2.2 (Kearse et al., 2012) with a maximum E-value of 10^{-5}. Contigs within the *P. monodon* reference genome that exhibited > 93% pairwise identity and grade > 80% with each queried SNP sequence (69 base pairs) were compiled and reported (including chromosomal distribution).

5.3 Results

5.3.1 Diversity SNP panel

Filtering of the previously reported high-quality SNP dataset (*n* = 10,593) for Indo-Pacific *P. monodon* (Chapter 4) removed 8,438 SNPs (79.7%) due to missing data per SNP ≥ 2%. This yielded a diversity SNP marker panel that consisted of 2,155 SNPs with call rate per SNP > 98% and per individual > 88.9% for each retained *P. monodon* individual (*n* = 520; Figure 5.2).

Cochran's Q-test demonstrated that MAF distribution across all 11 frequency categories (MAF ≤ 0.05 to MAF = 1) did not differ significantly (*p*-value = 1) between the entire SNP dataset (*n* = 10,593) and the full diversity SNP panel (*n* = 2,155) or any of the three smaller diversity subsets (1,500 SNPs, 1,000 SNPs, and 500 SNPs; Figure 5.3A). This consistency between the entire SNP dataset (*n* = 10,593; Chapter 4), full diversity SNP panel (*n* = 2,155), and all three smaller diversity subsets (1,500, 1,000, and 500 SNPs) was further supported by significant pairwise Pearson correlations (R = 1, $p < 0.00001$; Figure 5.3B-E).

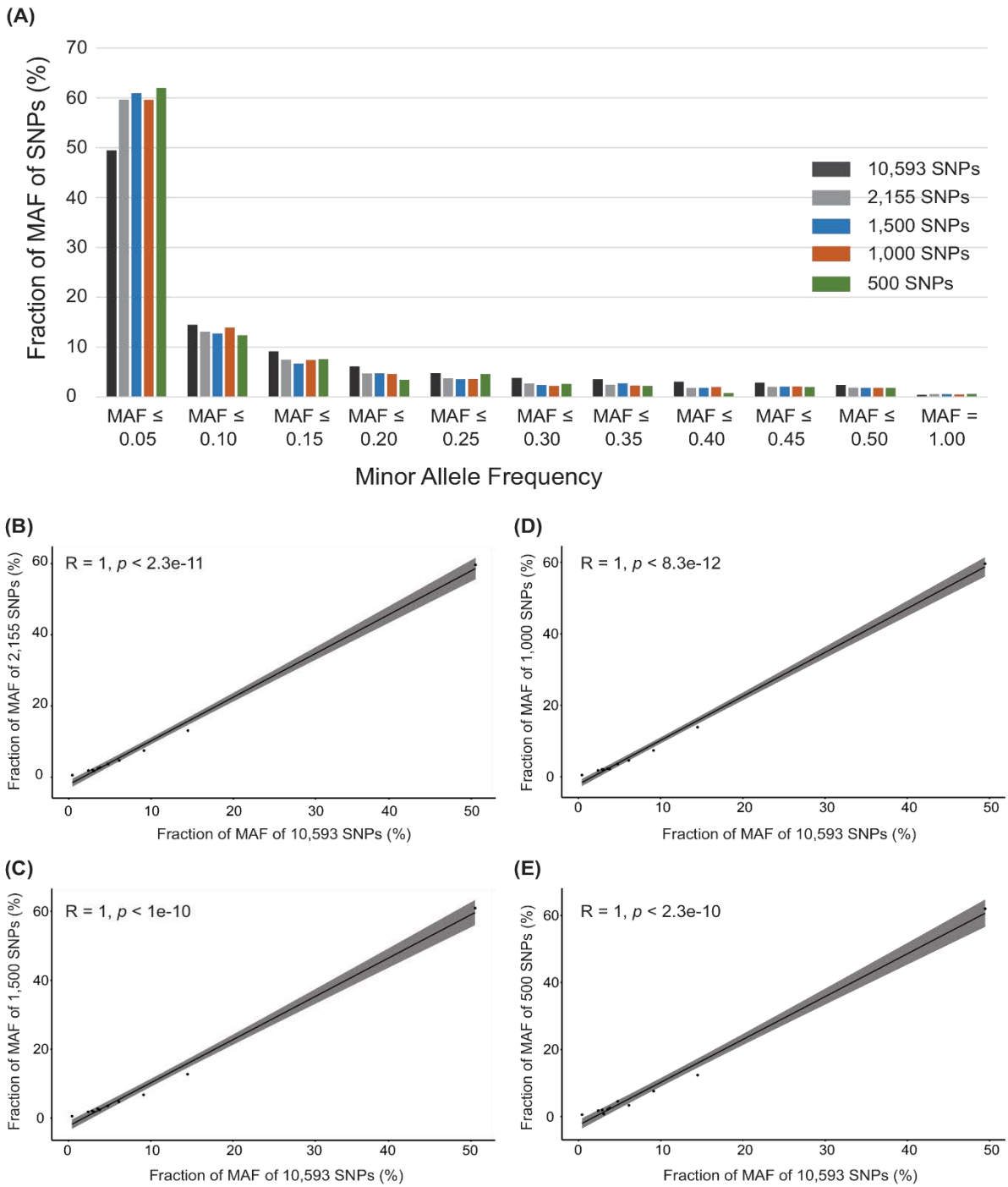


Figure 5.3. (**A**) Comparison of MAF fraction between the entire dataset (*n* = 10,593 SNPs), the full diversity SNP panel (*n* = 2,155), and the three smaller genetic diversity subsets (*n* = 1,500, 1,000, and 500) across sampled Indo-Pacific *Penaeus monodon* individuals (*n* = 520). (**B-E**) Pearson correlations between the MAF of fraction determined for each Indo-Pacific population using entire SNP dataset (*n* = 10,593; Chapter 4) and the full diversity SNP panel (**B**), 1,500 SNP marker subset (**C**), 1,000 SNP marker subset (**D**), and 500 SNP marker subset (**E**).

Mean A_R per population estimated with the entire dataset (n = 10,593; Chapter 4) was significant correlated with the mean A_R per population estimated with the full diversity SNP panel (n = 2,155; R = 1; $p < 0.00001$; Figure 5.4A) and the smallest diversity SNP panel (n = 500; R = 0.99; $p < 0.0001$; Figure 5.4B). Mean H_O per population estimated with the entire dataset (n = 10,593; Chapter 4) was significant correlated with the mean A_R per population estimated with the full diversity SNP panel (n = 2,155; R = 1; $p < 0.00001$; Figure 5.4C) and the smallest diversity SNP subset (n = 500; R = 0.99; $p < 0.0001$; Figure 5.4D). Mean H_E per population estimated with the entire dataset (n = 10,593; Chapter 4) was significant correlated with the mean A_R per population estimated with the full diversity SNP panel (n = 2,155; R = 0.9; $p < 0.00001$; Figure 5.4E) and the smallest diversity SNP subset (n = 500; R = 0.9; $p < 0.0001$; Figure 5.4F). Of note is that mean H_E values of the WIO populations (containing KE and SA) dropped from 0.1 when estimated based on the entire SNP dataset to 0.03 when estimated based on the four diversity SNP subsets (Table 5.1). The lowest accuracy of diversity determination (i.e., highest standard error) was observed for the smallest diversity SNP subset (n = 500; Table 5.1).

Mean F_{IS} values were 0.06, 0.07, 0.06, 0.07, 0.07, 0.07, 0.08, -0.06, 0.01, 0.08, 0.10, 0.09, 0.03, and 0.09 for the FJ, BB, EB, TSV, GC, JBG, TIW, WA, PHI, NTVN, CMVN, INDO, THL, and SLK populations when estimated based on the entire dataset (n = 10,593; Chapter 4) and 0.01, 0.01, 0.01, 0.03, -0.04, 0.02, 0.03, and 0.02 when estimated based on the full diversity SNP panel (n = 2,155) whereas both WIO populations (KE and SA) exhibited substantial reductions in mean F_{IS} values when based on the entire dataset (0.8 and 0.78) compared to the full diversity SNP panel (0.03 and 0.02), respectively (Table 5.1). F_{IS} values estimated based on the three small diversity SNP subsets (n = 1,500, 1,000, and 500) were variable within locations depending on the panel in a comparison with agreeable with the full diversity SNP panel (n = 2,155). Lastly, Mantel correlograms revealed strong correlations between pairwise F_{ST} estimates based on the entire SNP dataset (n = 10,593) and pairwise F_{ST} estimates based on the full diversity SNP panel ($r^2 = 0.9996$, $p < 0.001$; Figure 5.5A) as well as all three smaller diversity subsets consisting of 1,500 SNPs ($r^2 = 0.9995$, $p < 0.001$; Figure 5.5B), 1,000 SNPs ($r^2 = 0.9991$, $p < 0.001$; Figure 5.5C), and 500 SNPs ($r^2 = 0.9985$, $p < 0.001$; Figure 5.5D).

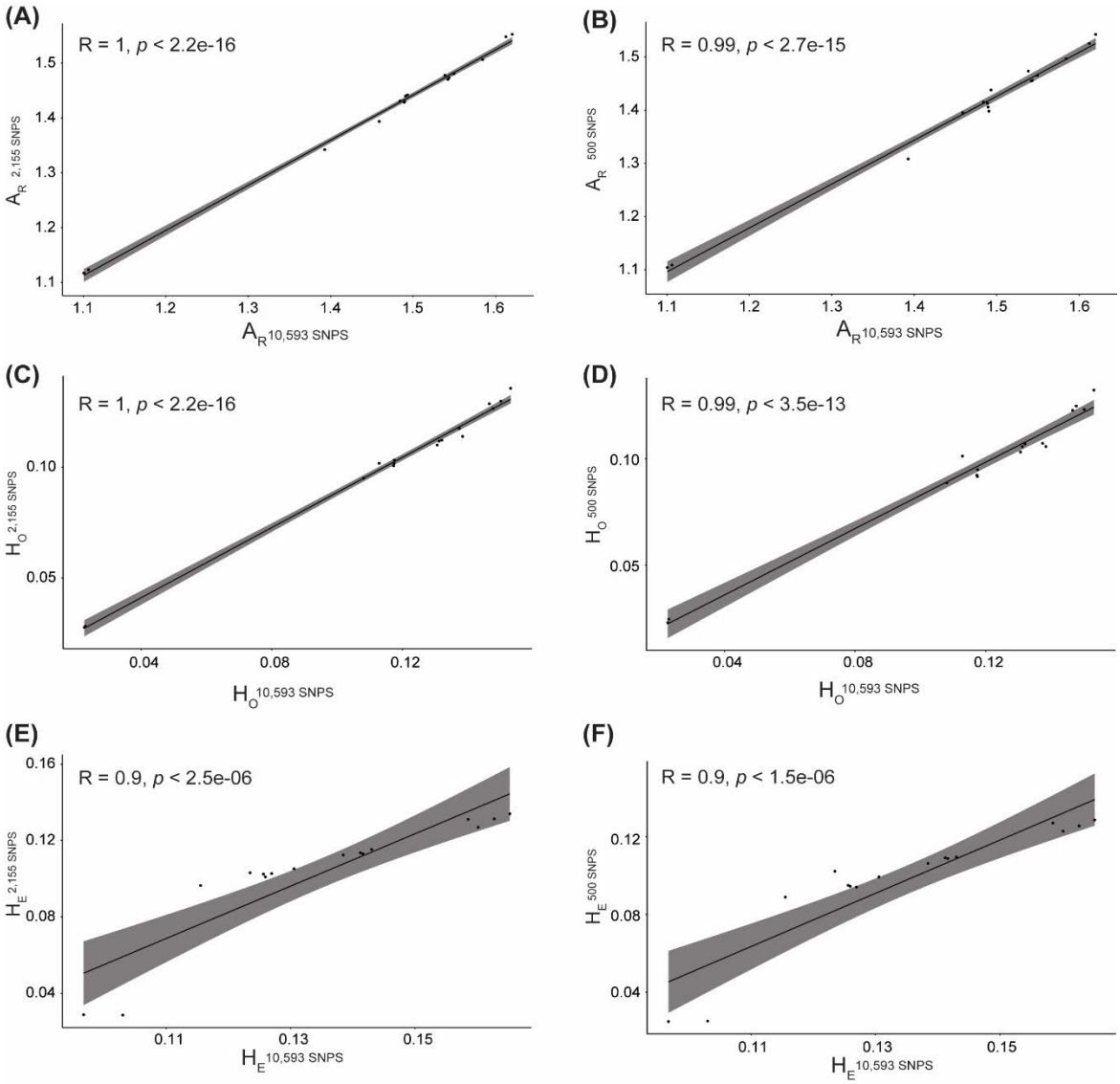


Figure 5.4 Pearson correlations between the entire SNP dataset (*n* = 10,593; Vu, et al., 2021) and the full diversity SNP panel (*n* = 2,155) and the 500 SNP marker subset for mean allelic richness ($A_{R;}$ panels **A** and **B**), mean observed heterozygosity (H_O; panels **C** and **D**), and mean expected heterozygosity (H_E; panels **E** and **F**), respectively.

Table 5.1 Genetic diversity indices (mean ± SE) for the sixteen Indo-Pacific *Penaeus monodon* populations used assessed all five SNP marker panels/subsets. Fiji (FJ), Eastern Australia (BB, EB, and TSV), Northern Australia (GC, JBG, and TIW), Western Australia (WA), (5) Vietnam (CMVN and NTVN), Indonesia (INDO), Thailand (THL), Sri Lanka (SLK), Kenya (KE), and South Africa (SA).

SNP dataset	FJ	BB	EB	TSV	GC	JBG	TIW	WA	PHI	NTVN	CMVN	INDO	THL	SLK	KE	SA
	Allele richness (A_R)															
10593	1.39 ± 0.004	1.49 ± 0.004	1.48 ± 0.004	1.49 ± 0.004	1.54 ± 0.004	1.54 ± 0.004	1.55 ± 0.004	1.46 ± 0.005	1.49 ± 0.004	1.58 ± 0.004	1.61 ± 0.004	1.62 ± 0.004	1.54 ± 0.004	1.49 ± 0.005	1.11 ± 0.003	1.10 ± 0.003
2155	1.34 ± 0.009	1.43 ± 0.009	1.43 ± 0.009	1.43 ± 0.009	1.47 ± 0.009	1.47 ± 0.009	1.48 ± 0.009	1.39 ± 0.009	1.44 ± 0.011	1.50 ± 0.009	1.55 ± 0.009	1.55 ± 0.009	1.47 ± 0.010	1.44 ± 0.010	1.12 ± 0.007	1.11 ± 0.006
1500	1.33 ± 0.011	1.42 ± 0.011	1.42 ± 0.011	1.41 ± 0.011	1.46 ± 0.011	1.45 ± 0.011	1.46 ± 0.011	1.37 ± 0.012	1.43 ± 0.013	1.50 ± 0.011	1.54 ± 0.010	1.55 ± 0.010	1.47 ± 0.012	1.44 ± 0.012	1.13 ± 0.008	1.12 ± 0.008
1000	1.33 ± 0.014	1.42 ± 0.013	1.43 ± 0.014	1.42 ± 0.014	1.46 ± 0.014	1.46 ± 0.013	1.47 ± 0.013	1.38 ± 0.014	1.43 ± 0.016	1.50 ± 0.014	1.54 ± 0.013	1.55 ± 0.013	1.47 ± 0.015	1.43 ± 0.014	1.11 ± 0.010	1.10 ± 0.009
500	1.31 ± 0.019	1.41 ± 0.018	1.42 ± 0.019	1.41 ± 0.020	1.46 ± 0.019	1.46 ± 0.019	1.47 ± 0.018	1.40 ± 0.020	1.40 ± 0.022	1.50 ± 0.020	1.53 ± 0.018	1.54 ± 0.018	1.47 ± 0.020	1.44 ± 0.020	1.11 ± 0.014	1.10 ± 0.013
	Observed heterozygosity (H_O)															
10593	0.11 ± 0.002	0.12 ± 0.002	0.12 ± 0.002	0.12 ± 0.002	0.13 ± 0.002	0.13 ± 0.002	0.13 ± 0.002	0.14 ± 0.002	0.14 ± 0.002	0.15 ± 0.002	0.15 ± 0.002	0.15 ± 0.002	0.15 ± 0.002	0.11 ± 0.002	0.02 ± 0.001	0.02 ± 0.001
2155	0.10 ± 0.003	0.10 ± 0.003	0.10 ± 0.003	0.10 ± 0.003	0.11 ± 0.003	0.11 ± 0.003	0.12 ± 0.003	0.11 ± 0.003	0.11 ± 0.003	0.13 ± 0.003	0.13 ± 0.003	0.13 ± 0.003	0.13 ± 0.003	0.10 ± 0.003	0.03 ± 0.002	0.03 ± 0.002
1500	0.09 ± 0.004	0.10 ± 0.004	0.10 ± 0.004	0.10 ± 0.004	0.11 ± 0.004	0.11 ± 0.004	0.11 ± 0.004	0.10 ± 0.004	0.11 ± 0.004	0.13 ± 0.004	0.13 ± 0.004	0.13 ± 0.004	0.13 ± 0.004	0.10 ± 0.004	0.03 ± 0.002	0.03 ± 0.002
1000	0.10 ± 0.005	0.10 ± 0.005	0.10 ± 0.005	0.10 ± 0.005	0.11 ± 0.005	0.11 ± 0.005	0.11 ± 0.005	0.10 ± 0.005	0.11 ± 0.005	0.13 ± 0.005	0.13 ± 0.005	0.13 ± 0.005	0.13 ± 0.005	0.10 ± 0.005	0.03 ± 0.003	0.03 ± 0.003

500	0.09 ± 0.007	0.09 ± 0.007	0.09 ± 0.007	0.09 ± 0.007	0.11 ± 0.007	0.11 ± 0.007	0.11 ± 0.007	0.10 ± 0.007	0.11 ± 0.007	0.12 ± 0.007	0.13 ± 0.007	0.13 ± 0.007	0.13 ± 0.007	0.10 ± 0.007	0.03 ± 0.004	0.02 ± 0.004
	Expected heterozygosity (H_E)															
10593	0.12 ± 0.002	0.13 ± 0.002	0.13 ± 0.002	0.13 ± 0.002	0.14 ± 0.002	0.14 ± 0.002	0.14 ± 0.002	0.13 ± 0.002	0.14 ± 0.002	0.16 ± 0.002	0.16 ± 0.002	0.17 ± 0.002	0.16 ± 0.002	0.12 ± 0.002	0.10 ± 0.003	0.10 ± 0.003
2155	0.10 ± 0.003	0.10 ± 0.003	0.10 ± 0.003	0.10 ± 0.003	0.11 ± 0.003	0.11 ± 0.003	0.12 ± 0.003	0.11 ± 0.003	0.11 ± 0.003	0.13 ± 0.003	0.13 ± 0.003	0.13 ± 0.003	0.13 ± 0.004	0.10 ± 0.003	0.03 ± 0.002	0.03 ± 0.002
1500	0.09 ± 0.004	0.10 ± 0.004	0.10 ± 0.004	0.10 ± 0.004	0.11 ± 0.004	0.11 ± 0.004	0.11 ± 0.004	0.10 ± 0.004	0.11 ± 0.004	0.13 ± 0.004	0.13 ± 0.004	0.13 ± 0.004	0.13 ± 0.004	0.10 ± 0.004	0.03 ± 0.002	0.03 ± 0.002
1000	0.10 ± 0.005	0.10 ± 0.005	0.10 ± 0.005	0.10 ± 0.005	0.11 ± 0.005	0.11 ± 0.005	0.11 ± 0.005	0.10 ± 0.005	0.11 ± 0.005	0.13 ± 0.005	0.13 ± 0.005	0.13 ± 0.005	0.13 ± 0.005	0.10 ± 0.005	0.03 ± 0.003	0.03 ± 0.003
500	0.09 ± 0.007	0.09 ± 0.006	0.09 ± 0.006	0.09 ± 0.006	0.11 ± 0.007	0.11 ± 0.007	0.11 ± 0.007	0.10 ± 0.007	0.11 ± 0.007	0.12 ± 0.007	0.13 ± 0.007	0.13 ± 0.007	0.13 ± 0.007	0.10 ± 0.007	0.03 ± 0.004	0.02 ± 0.004
	Inbreeding coefficient (F_{IS})															
10593	0.06	0.07	0.06	0.07	0.07	0.07	0.08	-0.06	0.01	0.08	0.10	0.09	0.03	0.09	0.80	0.78
2155	0.015	0.009	-0.007	0.007	0.031	0.010	0.028	-0.081	-0.043	0.006	0.024	0.030	-0.032	0.016	0.028	0.018
1500	0.010	0.008	-0.010	0.003	0.031	0.015	0.024	-0.088	-0.026	0.011	0.020	0.034	-0.032	0.009	-0.010	-0.013
1000	0.012	0.013	-0.001	0.020	0.043	0.014	0.030	-0.083	-0.051	-0.003	0.035	0.029	-0.016	0.008	-0.012	-0.020
500	0.0003	0.023	0.001	0.028	0.053	0.018	0.030	-0.064	-0.010	-0.014	0.025	0.041	-0.040	0.011	0.080	0.010

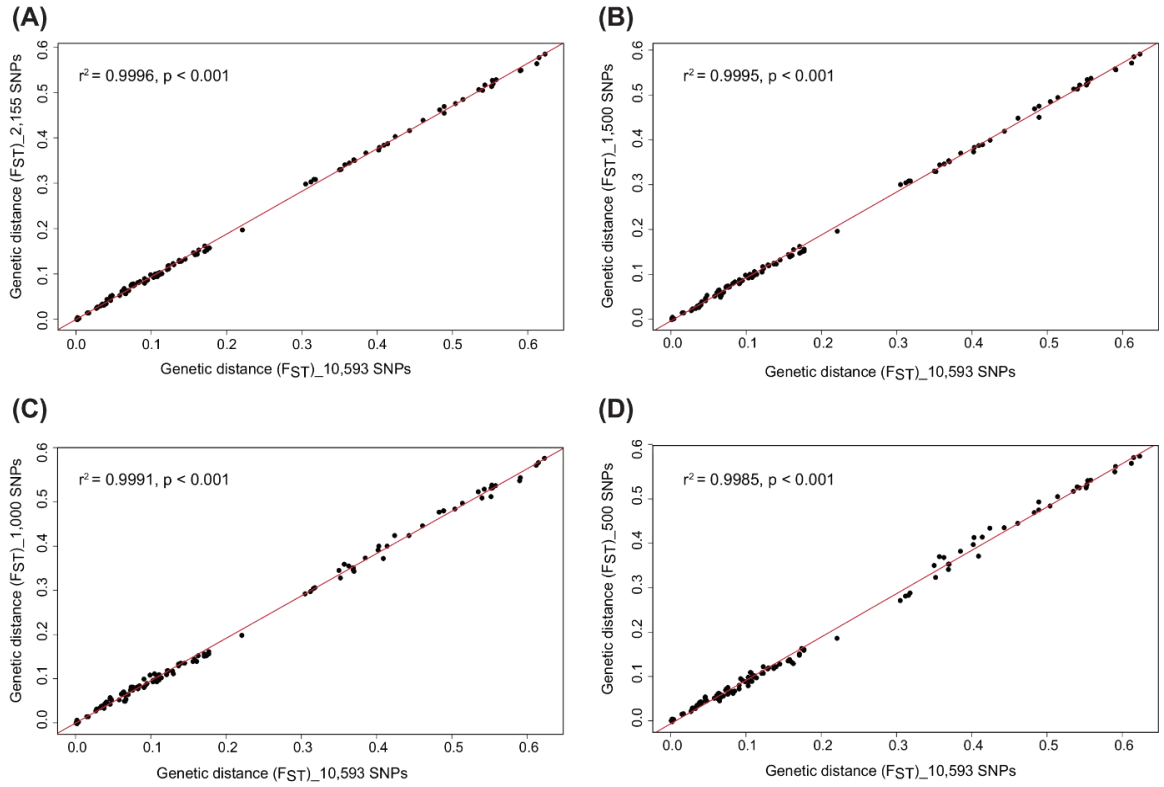


Figure 5.5. Mantel correlograms demonstrating the relationship between pairwise genetic differentiation (F_{ST}) based on the entire dataset dataset (n = 10,593; Chapter 3) and the full diversity SNP panel (n = 2,155; panel **A**) as well as the three smaller genetic diversity subsets consisting of 1,500 SNPs (panel **B**), 1,000 SNPs (panel **C**), and 500 SNPs (panel **D**).

5.3.2 Provenance SNP panel

From the entire dataset (n = 10,593; Chapter 3), three separate processes were used for identifying a provenance SNP panel (Figure 5.2). Firstly, 663 outlier SNPs with pairwise F_{ST} per SNP > 0.1 were identified using two independent population differentiation approaches (*BayeScan* and *PCAdapt*), as described in Chapter 3. Secondly, 1,628 unique SNPs were identified by combining the first four Principal Components (PC1 to PC4; that explained 96.6% variation of the observed population structure) in DAPC analysis, which was further reduced to 1,006 unique SNPs following subsequent filtering that removed all SNPs with pairwise F_{ST} values ≤ 0.1. Thirdly, 1,016 SNPs private allelic richness (A_{RA}) with pairwise F_{ST} values > 0.1 were identified using HP-RARE. Combining the three processes together, a final provenance SNP panel was identified (n = 1,200; Figure 5.2).

Based on the provenance SNP panel both DAPC and *NetView* analyses demonstrated that *P. monodon* collected from the same region could be traced back to the same genetic cluster (Figure 5.6). More specifically, DAPC analysis (with *a priori* population assumptions) demonstrated that all Indo-Pacific *P. monodon* individuals could be successfully traced back to their original cluster (Figure 5.6A-B). Particularly, DAPC analysis based on the provenance SNP panel (n = 1,200) correctly traced all *P. monodon* individuals from 14 of 16 Indo-Pacific populations back to their original region (i.e., genetic cluster). All *P. monodon* individuals collected from WA were incorrectly traced back to the NA region whereas 35.3% of *P. monodon* individuals collected from THL were correctly traced back to the SEA region while the remaining 64.7% were traced back to a new region (i.e., not one of the seven clusters identified in Chapter 3. Likewise, *NetView* analysis (without *a priori* population assumptions) using low, medium, or high kNN values (15, 30, or 50), respectively, demonstrated that *P. monodon* collected from the same location clustered tightly to their predefined populations and *P. monodon* collected from proximal locations could be successfully traced back to their original region (i.e., one of seven clusters) regardless of whether low, medium, or high kNN values (15, 30, or 50) were used (Figure 5.6C, D, and E), respectively.

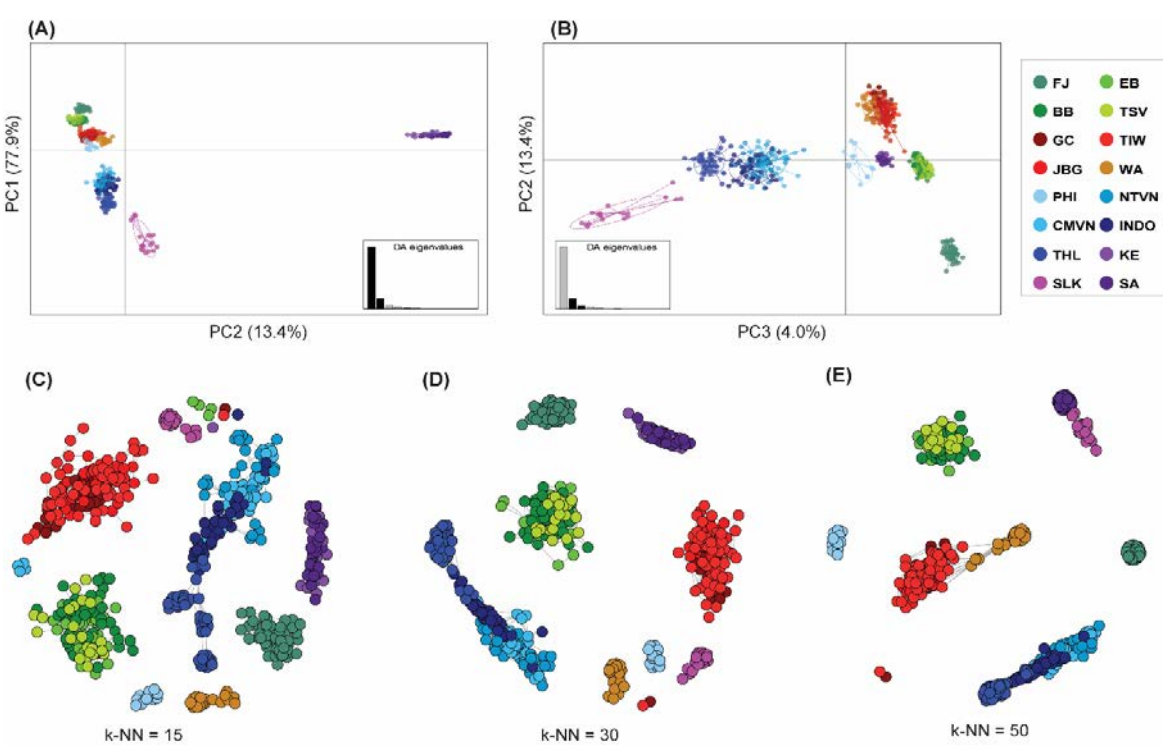


Figure 5.6 Genomic clustering of Indo-Pacific *Penaeus monodon* sample (n = 520) based on the provenance SNP panel (n = 1,200). Discriminant Analysis of Principal Components (DAPC) scatterplots showing PC1 and PC2 (**A**) as well as PC2 and PC3 (**B**). Percentage of variance explained by each PC is indicated in parenthesis on each axis. Population networks

constructed using *NetView* at kNN = 15 (**C**), kNN = 30 (**D**), and kNN = 50 (**E**). Dots represent individuals and colours correspond to sampling origin. FJ: Fiji; Eastern Australia (EA; BB: Bramston Beach, EB: Etty Bay; TSV: Townsville); Northern Australia (NA; GC: Gulf of Carpentaria, JBG: Joseph Bonaparte Gulf, TIW: Tiwi Island); Western Australia (WA); Vietnam (VN; CMVN: Ca Mau, NTVN: Nha Trang); PHI: Philippines; INDO: Indonesia; THL: Thailand; SLK: Sri Lanka; KE: Kenya; SA: South Africa.

All Monte-Carlo tests based on the provenance SNP panel (n = 1,200) indicated a high yet variable accuracy level depending on the proportion of individuals and F_{ST} ranked SNPs used (Figure 5.7). For basin traceability, Indo-Pacific *P. monodon* samples could be traced back to their original basin (PO or IO) with 100% accuracy regardless of whether 50%, 70%, or 90% of individuals were tested or if the top 25%, top 50%, top 75%, or 100% of F_{ST} ranked SNP markers were utilized (Figure 5.7A). For region traceability, individuals from FJ, EA, NA, and WIO could be traced back to their original region with 100% accuracy regardless of whether 50%, 70%, or 90% of individuals were tested or if the top 25%, top 50%, top 75%, or 100% of F_{ST} ranked SNP markers were utilized whereas individuals from WA, SEA, and SLK could be traced back to their original region with 96 – 98%, 97 – 100%, and 93 – 96% accuracy, respectively (Figure 5.7B). For sampling location traceability, individuals from three PO genetic units (FJ, EA, and NA) could be traced back to their original sampling location with 100% accuracy regardless of whether 50%, 70%, or 90% of individuals were tested or if the top 25%, top 50%, top 75%, or 100% of F_{ST} ranked SNP markers were utilized while individuals from the other five PO genetic units (WA, PHI, VN, INDO, and THL) could be traced back to their original sampling location with 95 – 98%, 97 – 100%, 69 – 78%, 78 – 87%, and 97 – 100% accuracy, respectively (Figure 5.7C). Individuals from one IO genetic unit (SLK) could be traced back to their original sampling location with 100% accuracy regardless of whether 50%, 70%, or 90% of individuals were tested or if the top 25%, top 50%, top 75%, or 100% of F_{ST} ranked SNP markers were utilized while individuals from the other two IO genetic units (KE and SA) could be traced back to their original sampling location with 81 – 92% and 90 – 94% accuracy, respectively (Figure 5.7D).

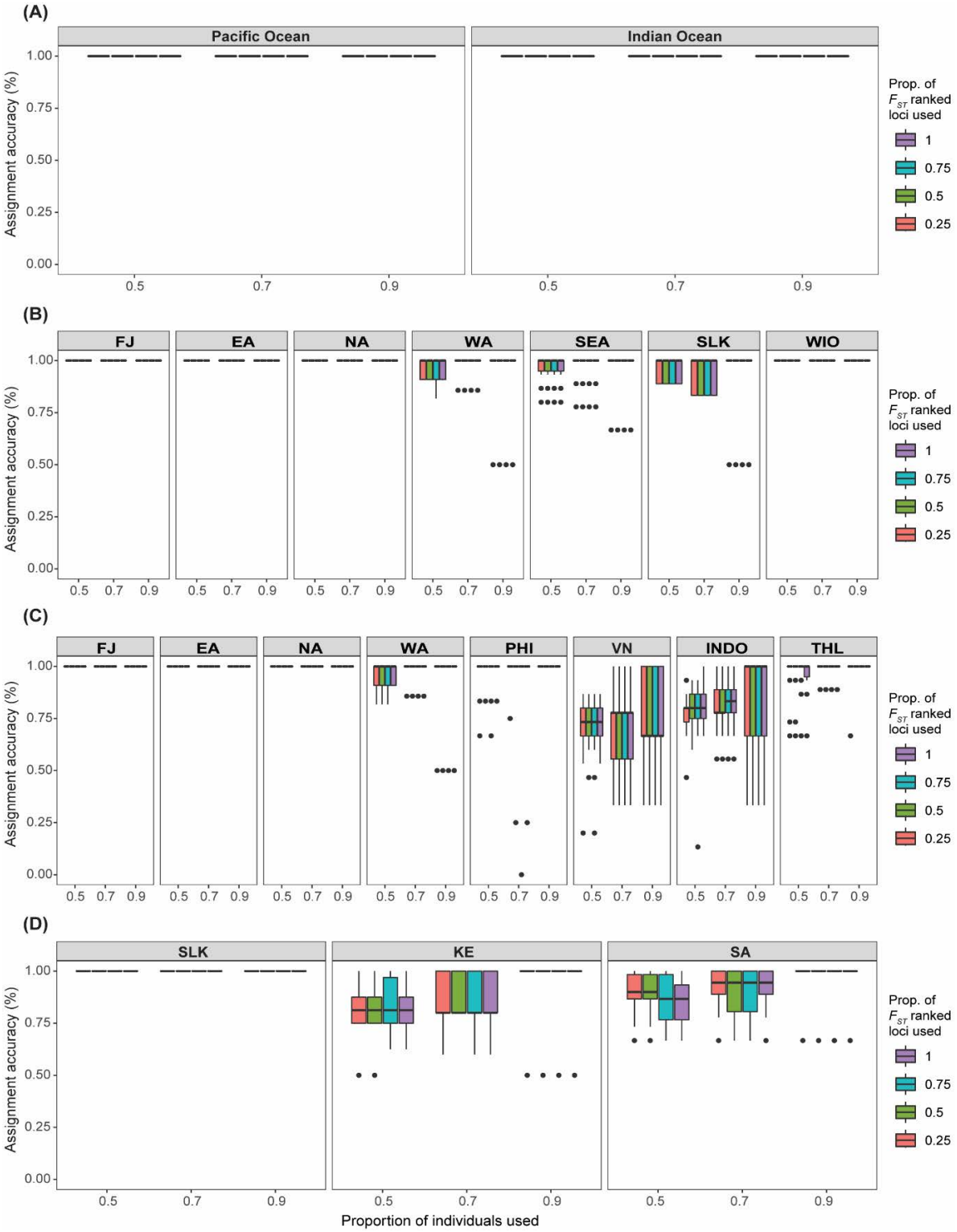


Figure 5.7 Monte Carlo tests based on the provenance SNP panel (n = 1,200) demonstrating the accuracy of tracing Indo-Pacific *Penaeus monodon* individuals (n = 520) back to their original basin (**A**), original region (**B**), original sampling location within Pacific Ocean (**C**), and original sampling location within Indian Ocean (**D**). Three proportions of individuals (50%, 70% and 90%) and F_{ST} ranked SNPs (top 25%, top 50%, top 75%, and 100%) were considered for all tests. FJ: Fiji; Eastern Australia (EA; BB: Bramston Beach, EB: Etty Bay;

TSV: Townsville); Northern Australia (NA; GC: Gulf of Carpentaria, JBG: Joseph Bonaparte Gulf, TIW: Tiwi Island); Western Australia (WA); Vietnam (VN; CMVN: Ca Mau, NTVN: Nha Trang); PHI: Philippines; INDO: Indonesia; THL: Thailand; SLK: Sri Lanka; KE: Kenya; SA: South Africa; SEA: Southeast Asia.

5.3.3 Pedigree SNP panels

Initial *dartQC* quality control filtering of the raw DArTseq^{TM} dataset (n = 131,929; Chapter 3) for each Indo-Pacific *P. monodon* individual (n = 508) removed 127,306 (96.5%), 119,456 (92.3%), and 124,036 (94.0%) SNPs for each individual from the PO, SLK, and WIO genetic groups (n = 4,713 SNPs, 7,110 SNPs, and 7,893 SNPs retained), respectively (Figure 5.2; Table 5.2). Subsequent quality control filtering retained 218, 433, and 233 high-quality SNPs for the PO, SLK, and WIO genetic groups, respectively (Table 5.2). LD and HWE tests removed an additional one SNP from the PO genetic group and three from the WIO genetic group (Table 5.2). The final pedigree SNP panels for the PO, SLK, and WIO groups contained 217, 226, and 220 SNPs, respectively (Figure 5.2; Table 5.2). There was minimal overlap between the SNP content of the three pedigree panels with 25, 3, and 10 of these common between the PO and SLK groups, PO and WIO groups, and SLK and WIO groups, respectively, while one SNP marker was common across all panels.

The PO pedigree SNP panel exhibited a mean MAF that ranged from 0.30 ± 0.11 in the FJ genetic unit to 0.34 ± 0.11 in the EA genetic unit, a mean H_O that ranged from 0.40 ± 0.12 in the INDO genetic unit to 0.46 ± 0.16 in the WA genetic unit, and a mean H_E that ranged from 0.40 ± 0.10 in the WA genetic unit to 0.43 ± 0.07 in the NA genetic unit (Table 5.3). The SLK pedigree SNP panel exhibited a mean MAF, H_O, and H_E of 0.39 ± 0.07, 0.47 ± 0.17, and 0.47 ± 0.03, respectively (Table 5.3). The WIO pedigree SNP panel exhibited a mean MAF, H_O, and H_E of 0.40 ± 0.06, 0.47 ± 0.03, and 0.49 ± 0.15 (Table 5.3).

Table 5.2 The number of SNPs retained for pedigree analysis after sequential filtering applications. The proportion of SNPs retained at each threshold are indicated in parenthesis.

Filtering Steps	**Retained SNP count**								
Raw SNPs	131,929 (100%)								
Average read depth ≥ 7 and replication average ≥ 0.9	129,416 (98.1%)								
Call rate ≥ 98%	PO							SLK	WIO
	6,188 (4.7%)							9,960 (7.6%)	10,638 (8.1%)
Similar sequence clusters ≥ 0.95	4,713 (3.6%)							7,110 (5.4%)	7,893 (6%)
	Polymorphic SNPs								
Genetic groups	PO							SLK	WIO
Genetic units	FJ	EA	NA	WA	VN	INDO	THL	SLK	WIO
Polymorphic SNPs *	1,129	1,472	1,732	1,026	1,955	1,933	1,585	2,668	1,287
Polymorphic SNPs for each genetic unit**	614	601	686	506	852	883	911	433	223
Common polymorphic SNPs	218							433	223
LD filter ($r^2 > 0.2$) and HWE filter ($p \geq 0.0001$)	217							226	220
Final pedigree SNP panels	217 (0.16%)							226 (0.17%)	220 (0.17%)

*Polymorphic SNPs used to calculate exclusion probability (PE) values in PE3 scenario. **Polymorphic SNPs for the Pacific Ocean (PO), Sri Lanka (SLK), and West Indian Ocean (WIO) genetic units with MAF > 0.1, MAF > 0.28, and MAF > 0.28, respectively. Fiji (FJ), Eastern Australia (EA), Northern Australia (NA), Western Australia (WA), Vietnam (VN), Indonesia (INDO), and Thailand (THL).

Table 5.3 Summary statistics (mean ± SD) for the Pacific Ocean (PO), Sri Lanka (SLK), and West Indian Ocean (WIO) pedigree SNP panels.

Genetic groups	Genetic units	*N*	MAF	H_O	H_E
PO	EA	104	0.34 ± 0.11	0.42 ± 0.10	0.42 ± 0.08
	NA	114	0.34 ± 0.10	0.42 ± 0.09	0.43 ± 0.07
	WA	22	0.31 ± 0.12	0.46 ± 0.16	0.40 ± 0.10
	FJ	49	0.30 ± 0.11	0.41 ± 0.11	0.41 ± 0.08
	VN	69	0.33 ± 0.11	0.41 ± 0.12	0.42 ± 0.09
	INDO	38	0.33 ± 0.10	0.40 ± 0.12	0.42 ± 0.08
	THL	46	0.31 ± 0.11	0.42 ± 0.15	0.40 ± 0.09
SLK	SLK	19	0.39 ± 0.07	0.47 ± 0.17	0.47 ± 0.03
WIO	WIO	47	0.40 ± 0.06	0.47 ± 0.03	0.49 ± 0.15

Number of *P. monodon* samples (N); observed heterozygosity (H_O); expected heterozygosity (H_E); minor allele frequency (MAF); Fiji (FJ); Eastern Australia (EA), Northern Australia (NA), Western Australia (WA), Vietnam (VN), Indonesia (INDO), and Thailand (THL).

Probability of exclusion (PE) estimates confirmed the accuracy of the PO, SLK, and WIO pedigree SNP panels across the PE1, PE2, and PE3 test scenarios (Figure 5.8). In all cases, as the number of pedigree SNP markers increased the PE estimates also increased until they reached a plateau (Figure 5.8). For scenarios PE1 and PE2, the minimum number of pedigree SNP markers required to achieve a PE estimate of 99% was 20 for all PO genetic units and 15 for all SLK and WIO genetic units (Figure 5.8). For scenario PE3, the minimum number of polymorphic SNP markers required to achieve a PE estimate of 99% was 35 for all PO, SLK, and WIO genetic units (Figure 5.8).

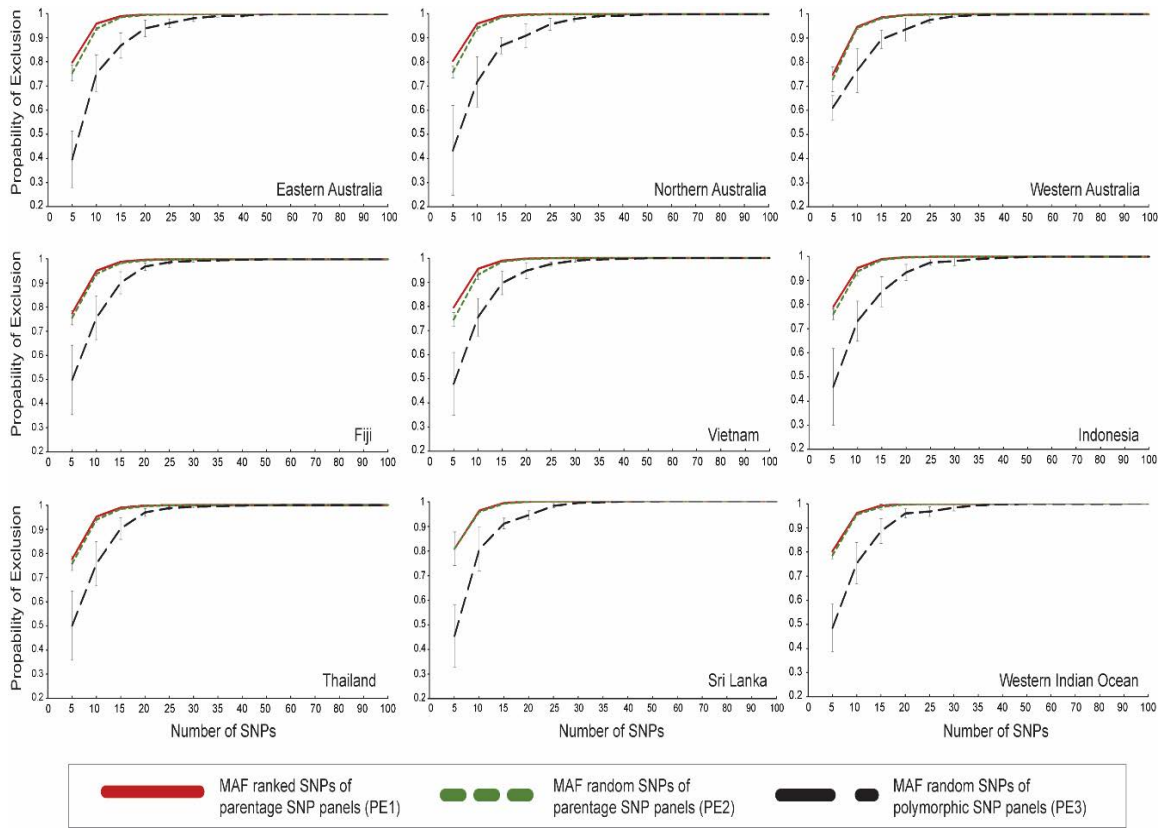


Figure 5.8 Relationship between the number of pedigree SNP markers and exclusion probability (PE) estimates for correctly assigning parent-pair to an individual within each of the nine genetic units (Table 5.3). PE calculations were performed in CERVUS using multiple SNP subsets (n = 14) that consisted of various SNP marker quantities (n = 5 – 100) across 10 replicates for three discrete test scenarios (PE1, PE2, and PE3). PE1: incremental SNP marker subsets were selected high to low MAF-ranking from the PO, SLK, and WIO pedigree SNP panels; PE2: incremental SNP marker subsets were selected randomly from the PO, SLK, and WIO pedigree SNP panels; PE3: incremental SNP marker subsets were selected randomly from the filtered polymorphic datasets (MAF ≤ 0.01 and missing data per SNP ≥ 5%).

Simulations indicated that when parent-pair was known, 35 pedigree SNP markers were required to obtain 100% assignment accuracy (95% confidence) for progenies from all PO genetic units except for WA, which required 40 pedigree SNP markers to obtain 100% accuracy (Figure 5.9). For SLK and WIO genetic units, 30 pedigree SNP markers were required to obtain 100% parent assignment accuracy (95% confidence) when the parent-pair was known (Figure 5.9).

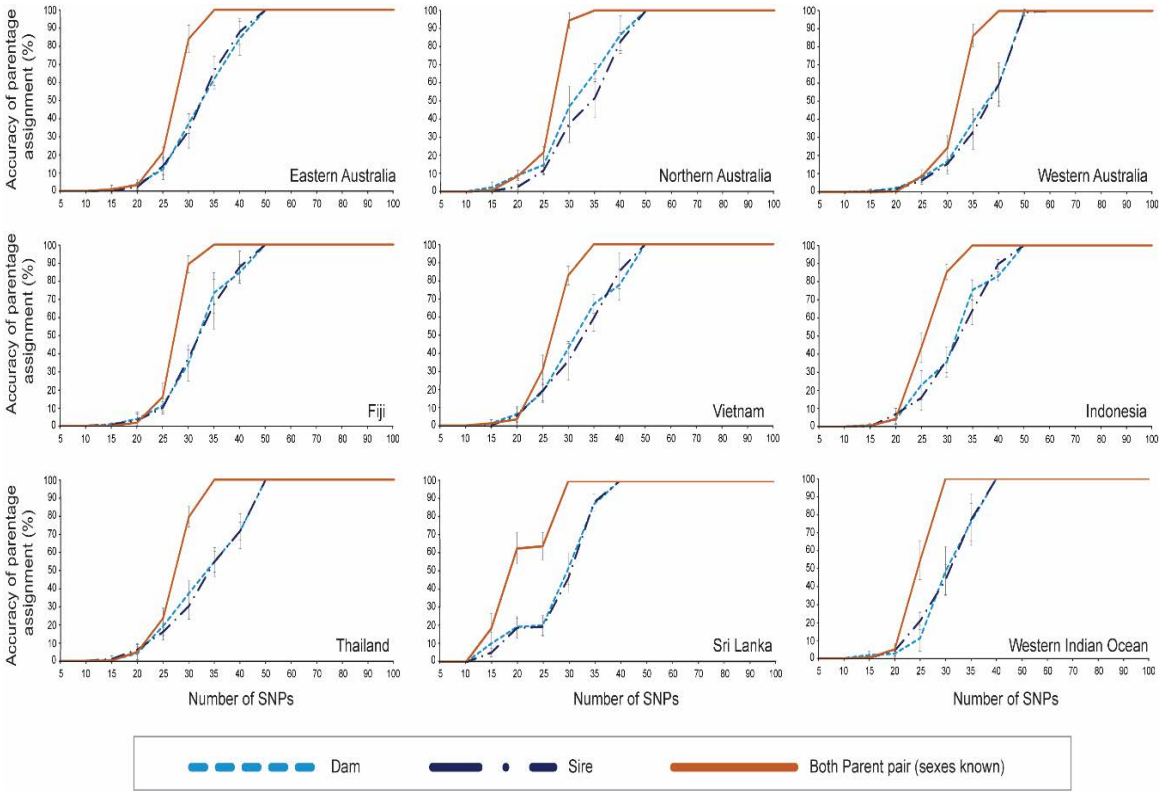


Figure 5.9 Predicted accuracy of Pacific Ocean (PO), Sri Lanka (SLK), and West Indian Ocean (WIO) pedigree SNP panels to correctly assign progeny to parents with 95% confidence using different sized marker subsets. Subsets were randomly selected from the corresponding PO, SLK, or WIO pedigree SNP panels (Table 5.3).

The EA genetic unit was used to evaluate the accuracy of the PO pedigree SNP panel (n = 217) across 10 generations using four subsets (n = 30, 35, 40, and 50 SNP markers; Table 5.4). When 30 SNPs markers were used 78.6%, 75.4%, 63.2%, 53.0%, and 41.2% of F1, F3, F5, F8, and F10 generation progenies could be accurately (i.e., 95% confidence) assigned back to their parent-pairs, respectively (Table 5.4). When 35 SNP markers were used all F1, F3, and F5 generation progenies could be accurately assigned back to their parent-pairs while 99.6% and 94.0% of F8 and F10 generation progenies could be accurately assigned back to their parent-pairs, respectively. When 40 SNP markers were used all 100% F1, F3, F5, and F8 generation progenies could be accurately assigned back to their parent-pairs while 99.2% F10 generation progenies could be accurately assigned back to their parent-pairs. When 50 SNP markers were used all F1, F3, F5, F8, and F10 generation progenies could be accurately assigned back to their parent-pairs (Table 5.4).

Table 5.4 Pedigree determination accuracy (95% confidence) across 10 simulated generations of Eastern Australia progenies using four SNP marker subsets (*n* = 30, 35, 40, and 50) that were randomly selected from the Pacific Ocean (PO) pedigree SNP panel (*n* = 217).

SNP marker subsets	Simulated generations				
	F1	F3	F5	F8	F10
30	78.6 ± 14.7	75.4 ± 19.4	63.2 ± 23.8	53.0 ± 25.1	41.2 ± 16.0
35	100	100	100	99.6 ± 0.5	94.0 ± 7.7
40	100	100	100	100	99.2 ± 1.8
50	100	100	100	100	100

5.3.4 Genomic distribution of SNP panels

The genetic diversity and provenance SNP panels exhibited the chromosome-scale genome assembly of *P. monodon* (Uengwetwanit et al., 2021) matches for 84.5% and 79.3% (*n* = 1,820 and 952, respectively) while only 32.8% and 37.8% (*n* = 706 and 454, respectively) passed further stringency testing and exhibited contig matches distributed across 43 and 43 of 44 chromosomes, respectively (Table 5.5). The PO, SLK, and WIO pedigree SNP panels exhibited genome accessions matches for 82.0%, 82.7%, and 87.3% (*n* = 178, 187, and 192, respectively) while only 27.7%, 47.4%, and 47.3% (*n* = 60, 107, and 104, respectively) passed further stringency and exhibited contig matches distributed across 41, 40, and 41 of 44 chromosomes, respectively. When assessed together, the common SNP markers exhibited genome accessions matches for 83.1% (*n* = 2,853) while only 36.7% (*n* = 1,261) passed further stringency and exhibited contig matches (*n* = 181) distributed across 43 of 44 chromosomes (Appendix 5.2).

Table 5.5 Number of SNPs of the five SNP panels were associated with the chromosome-scale genome assembly of *Penaeus monodon*.

Identified SNP panels	Total SNP markers	Accession matched	Stringency passed*	Chromosome distribution
Diversity SNP panel	2,155 (100%)	1,820 (84.5%)	706 (32.8%)	43/44
Provenance SNP panel	1,200 (100%)	952 (79.3%)	454 (37.8%)	43/44
PO pedigree SNP panel	217 (100%)	178 (82.0%)	60 (27.7%)	41/44
SLK pedigree SNP panel	226 (100%)	187 (82.7%)	107 (47.4%)	40/44

WIO pedigree SNP panel	220 (100%)	192 (87.3%)	104 (47.3%)	41/44
Common SNPs across panels	3,433 (100%)	2,853 (83.1%)	1,261 (36.7%)	43/44

*SNP marker stringency set to pairwise identity > 93% and grade > 80%. Pacific Ocean (PO); Sri Lanka (SLK); West Indian Ocean (WIO).

5.4 Discussion

Using both raw (*n* = 131,929) and high-quality (*n* = 10,593) genome-wide SNP datasets for Indo-Pacific *P. monodon* (Chapter 3) this study identified the following low-density SNP panels for global commercial applications: 1) a single diversity SNP panel (*n* = 2,155), 2) a single provenance SNP panel (*n* = 1,200), and 3) three pedigree SNP panels (*n* = 217, 226, and 220 for PO, SLK, and WIO genetic groups), respectively (Figure 5.2). These low-density SNP panels keep genotyping and analytical costs to a minimum while also providing improvement in SNP genotype data consistency across global studies, assisting with development of a global reference genotypes database, and progressing towards robust global analysis standards establishment.

The identified diversity SNP panel was able to determine the genetic diversity inherent within Indo-Pacific *P. monodon* population to a degree of accuracy that was indistinguishable from genetic diversity analyses based on the entire high-quality SNP dataset (Figures 3 - 5). The small differences in accuracy observed appear to be in agreement with previous studies by Gebrehiwot et al. (2021) and Tsairidou et al. (2020), which reported a similarly small trade-off between increased genotyping cost (i.e., low- versus high-density SNP panels) and decreased accuracy in genetic diversity determination. More specifically, the slightly reduced correlation coefficient for H_E (R = 0.90) could be due to the presence of null alleles within the WIO group, as previously reported in Chapter 3. Overall, the diversity SNP panel exhibited high accuracy as evidenced by the low SE values observed for A_R, H_O, and H_E (Table 5.1). All three smaller diversity subsets (*n* = 1,500 SNPs, 1,000 SNPs, and 500 SNPs) exhibited strong correlation coefficients across all tests (R ≥ 0.89; Figure 5.4); however, as expected, SE values were slightly elevated compared to genetic diversity metrics determined using the entire high-quality dataset (*n* = 10,593). Compared to the entire high-quality SNP dataset, the diversity SNP panel reduced the F_{IS} values observed for the KE and SA populations from 0.80 and 0.78 to 0.03 and 0.02, respectively (Table 5.1), which is consistent with the reduction in F_{IS} values observed for KE and SA populations during similar SNP subset comparisons (Chapter 4). The significantly

elevated F_{IS} values observed for these WIO populations could be due to technical artifacts of the RAD-seq genotyping approaches (i.e., the presence of null alleles; Chapter 4). Based on these comparisons, I showed that the 500 SNP panel is effective for determination of general genetic diversity indices within and among Indo-Pacific *P. monodon* populations when compared to other tested SNP panels (i.e., 1,000; 1,500; 2,155 and 10,953). However, it should be noted that the variance of the diversity estimates increases (albeit very small here) with the smaller SNP panels (Table 5.1) and this should be considered when choosing optimum SNP panel numbers.

The identified provenance SNP panel (n = 1,200; Figure 5.2) was able to accurately trace individuals back to their original basin (PO or IO), genetic cluster (FJ, EA, NA, WA, SEA, SLK and WIO; Chapter 4), and sampling location within PO and IO basins using three independent analyses (Figures 5.6 and 5.7). DAPC analysis (i.e., with *a priori* population assumptions) accurately traced all individuals back to their original region (i.e., genetic cluster) except for individuals from WA and THL populations, which were not entirely assigned back to the WA and SEA clusters, respectively. WA individuals were most likely traced back to the NA cluster given that the provenance SNP panel contained a mix of outlier, structure, and private allele SNPs (section 2.4 and 3.2) as this clustering was observed for the WA population when broad scale analysis was conducted using outlier SNPs (Chapter 4). Tracing the majority of THL individuals back to a novel cluster was unexpected but most likely due to the unique mix of SNPs present within the provenance SNP panel. *NetView* analysis (i.e., without *a priori* population assumptions) accurately grouped all individuals with those collected from the same or nearby sampling locations (Figure 5.6C-E). Monte Carlo tests also demonstrated the ability of the provenance SNP panel to accurately assign individuals back to their original basin, region, and sampling location within PO and IO basins (Figure 5.7). Across all levels there was no observed difference in traceability accuracy between the top 25% (n = 300), top 50% (n = 600), top 75% (n = 900), or 100% (n = 1,200) of F_{ST} ranked markers; however, inclusion of 70 – 90% of individuals tended to result in better traceability at the region and sampling location levels. This observation was not surprising considering the high admixture and low genetic difference between populations within genetic clusters (e.g., SEA cluster contains VN, INDO, and THL while WIO cluster contains KE and SA). Pickrell and Pritchard (2012) and Drinan et al. (2018) reported the success of provenance tests depends on the level of genetic differentiation between putative source populations. The lack of power for accurate provenance determination has also been observed in other species, including American lobster (*Homarus*

americanus) (Benestan et al., 2015), European lobster (*Homarus gammarus*) (Jenkins et al., 2019), and Chilean blue mussels (*Mytilus chilensis*) (Araneda et al., 2016). Taken together, genotyping Indo-Pacific *P. monodon* individuals using the full provenance SNP panel (*n* = 1,200; Figure 5.2) can accurately trace individuals back to their original basin, region, and sampling location; however, genotyping only 50 - 90% of individuals with the top 25% of F_{ST} ranked markers (*n* = 300) can be sufficient for accurate provenance determination depending on the required level of resolution (e.g., basin, region, or sampling location).

The identified PO, SLK, and WIO pedigree SNP panels (*n* = 217, 226, and 220; Figure 5.2), respectively, were able to accurately assign theoretical progenies back to their parent-pair (i.e., both parents known) *in silico* (Figures 8 and 9). Simulations revealed that progenies from PO units required 35 or 40 markers from the PO pedigree SNP panel to reach 100% assignment accuracy when the parent-pair was known, whereas progenies from SLK and WIO units required 30 markers from the SLK and WIO pedigree SNP panels to reach 100% assignment accuracy when only one parent or the parent-pair was known, respectively (Figure 5.9). The higher number of pedigree SNP markers required for progenies from PO units compared to progenies from SLK and WIO units is likely due to the higher genetic divergence amongst PO populations compared to SLK and WIO populations (Chapter 4). The observed ability to accurately assign Indo-Pacific *P. monodon* progenies using 30 – 40 markers from the larger pedigree SNP panel aligns with the lower end of previously reported SNP marker requirements for high accuracy in pedigree determination (e.g., *n* = 50 – 200; (Holman et al., 2017; Nguyen et al., 2014; Perez-Enriquez & Max-Aguilar, 2016; Sellars et al., 2014; Thongda et al., 2018; Zhao et al., 2018)). For example, Holman et al. (2017) achieved 100% accuracy for assigning Atlantic salmon (*Salmo salar L.*) progenies back to their parent-pairs using a low-density SNP panel (*n* = 94) that was identified from a genome-wide Restriction Site Associated DNA Sequencing (RAD-seq) dataset.

Moreover, the power of pedigree determination is known to be affected by fundamental factors including the MAF and PE of individual markers (Liu et al., 2016; Liu et al., 2017; Premachandra et al., 2019). MAF in populations has been shown to be one of the most important criteria for guiding selection of informative pedigree SNP markers in that assignment accuracy increases when MAF increases (Anderson and Garza, 2006; Liu et al., 2016). In this study, mean MAF was > 0.30 for the PO pedigree SNP panel and 0.39 and 0.4 for the SLK and WIO pedigree SNP panels, respectively, which was consistent with MAF values reported for pedigree SNP panels identified in three other species (MAF ≥ 0.3): 1) rainbow trout

(*Oncorhynchus mykiss*) (S. Liu et al., 2016), 2) Florida bass (*Micropterus floridanus*) (Zhao et al., 2018), and 3) diverse breeds of sheep (Heaton et al., 2014). Pedigree determination accuracy also depends on the assignment power of the SNP markers used, which is determined by their PE (Liu et al., 2017). PE was calculated from allele frequencies based on the theory that a parents and progenies must share an allele at every locus (Chakraborty et al., 1988). *In silico* pedigree determination simulation using known parent-pairs revealed that scenarios PE1 and PE2 (high-to-low and random MAF ranking of pedigree SNP panels) achieved 99% accuracy with 20 SNP markers for the PO pedigree SNP panel and 15 SNP markers for the SLK and WIO pedigree SNP panels whereas the number of SNP markers required to achieve 100% assignment accuracy increased to 35 under scenario PE3 (random MAF ranking of larger PO, SLK, and WIO polymorphic datasets), respectively. Also, *in silico* testing of the PO pedigree SNP panel (EA unit only) demonstrated that 35, 40, or 50 randomly selected SNP markers were sufficient to achieve 100% assignment accuracy for simulated F1 - F5, F1 – F8, and F1 – F10 generation progenies, respectively (Table 5.4). The relatively low number of SNP markers required to achieve high pedigree determination accuracy across all *in silico* simulations could be due to the relatively low number of candidate parents used (n = 18) and simulated population size (n = 100) given that breeding programs typically used more candidate parents and generate more progenies and, thus, require more SNP markers to achieve high pedigree determination accuracy (Arbon et al., 2021; May et al., 2020). More specifically, 50 and 75 SNP markers were required to achieve > 95% parent-pair assignment accuracy in Sockeye salmon (*Oncorynchus nerka*) when population sizes were 100 and 500, respectively (May et al., 2020). Additionally, to assess the effect of candidate parents on assignment success in Australian greenlip abalone (*Haliotis laevigata*), Arbon et al. (2021) indicated that 26, 37, and 41 SNPs were required to assign progeny correctly (i.e., 95% confidence) for 16 candidate parents, 2 – 100 candidate parents, and 3 – 200 candidate parents, respectively. Taken together, the identified PO, SLK, and WIO pedigree SNP panels (Figure 5.2) accurately determined pedigree for individuals from the Indo-Pacific *P. monodon* populations within the PO, SLK, and WIO genetic groups, respectively, and, in so doing, effectively provide new cost-effective and robust genetic tools for future selective breeding programs that aim to increase *P. monodon* aquaculture production.

Lastly, the common SNP markers across all five panels (n = 3,433) were associated with 181 contigs from the *P. monodon* genome assembly (Uengwetwanit et al., 2021) that exhibited distribution across 43 of 44 chromosomes (Table 5.5, Appendix 5.2) while individual

SNP panels covered 40 to 43 of 44 chromosomes (Table 5.5). Such broad genome coverage is ideal for designing low density SNP panels as suggested in (Holman et al., 2017), (Reverter et al., 2020), and (Wang et al., 2021). Moreover, the identified common SNP markers provide a low-density panel for cost-effective genotyping of genome-wide SNPs should information about diversity, provenance, and/or pedigree be of interest to future research or commercial applications. While genetic distance between SNP markers was not tested in this study my results regarding overall SNP panel performance supports the recent finding that genetic distance or regions of high recombination have little-to-no impact on SNP panel performance (Tsairidou et al., 2020).

5.5 Conclusion

The five SNP panels identified in this study (Figure 5.2) provide a suite of novel genetic tools for the accurate determination of diversity, provenance, and pedigree within and among all Indo-Pacific *P. monodon* populations. Given the native distribution of *P. monodon* throughout the Indo-Pacific region and the global development of *P. monodon* aquaculture, identification of these five low-density SNP panels permits cost-effective determination of diversity, provenance, and pedigree, which, in turn, advances *P. monodon* aquaculture and fishery management. The diversity and provenance SNP panels (n = 2,155 and 1,200), respectively, provided accurate results that were consistent with the results generated from the entire high-quality SNP dataset for Indo-Pacific *P. monodon* (n = 10,593; Chapter 3). Likewise, *in silico* simulations confirmed that the PO, SLK, and WIO pedigree SNP panels (n = 217, 226, and 220), respectively, can assign *P. monodon* individuals from PO, SLK, and WIO unit populations back to one parent or parent-pairs with 100% accuracy using ≤ 40 markers. The common SNP markers across all five panels (n = 3,433) collectively provide a useful low-density and cost-effective tool for genome-wide genotyping of Indo-Pacific *P. monodon*, which has a diversity of aquaculture and fishery management applications.

Chapter 6. General Discussion

6.1 Significance of research and contribution to the field

The giant black tiger shrimp *Penaeus monodon* is native to and considered one of the most important aquaculture species throughout the Indo-Pacific region. Despite the commercial importance of *P. monodon*, information on levels of predicted gene flow and patterns of genetic connectivity within and among Indo-Pacific populations were not consistently reported across previous studies, as outlined in Chapters 2 and 3. Prior to the research outlined in this thesis, both fine- and broad-scale population genetic structure and connectivity patterns as well as phylogenetic relationships within and among Indo-Pacific *P. monodon* were based on limited resolution DNA markers such as mitochondrial DNA (mtDNA), microsatellites, and/or allozymes (Benzie et al., 2002; Benzie et al., 1992; Brooker et al., 2000; Waqairatu et al., 2012; You et al., 2008). Additionally, only limited information was available regarding the influence of environmental pressures (e.g., temperature) on the genetic structure of geographically discrete populations at both fine- and broad-scales as well as only limited development of commercial resources for genetic diversity, provenance, and pedigree determination. In light of these knowledge gaps the four interdependent studies that constitute this thesis utilized a multi-level approach to generate new population genomics insights about and resources for Indo-Pacific *P. monodon* that aimed to help advance sustainable aquaculture and fishery management practices. More specifically, I: 1) utilized both neutral and outlier SNP datasets to determine the fine-scale population structure, connectivity, and putative local adaptation of Australian *P. monodon* populations (Chapter 2), 2) utilized both neutral and outlier SNP datasets to determine the broad-scale population structure, connectivity, and local adaptation of Indo-Pacific *P. monodon* populations (Chapter 3), 3) evaluated the phylogenomic and phylogeographic relationships among Indo–Pacific *P. monodon* populations and sister taxa using genome-wide SNP and presence-absence variant (PAV) loci datasets (Chapter 4), and 4) identified five low-density SNP panels that permit cost-effective and accurate diversity, provenance, and pedigree determination for all Indo-Pacific *P. monodon* populations (Chapter 5).

Chapter 2 addressed the knowledge gap about Australian *P. monodon* population genomics by generating a high-quality genome-wide single nucleotide polymorphism (SNP)

dataset (n = 10,624 total) for each individual (n = 278 total) collected from geographically disparate populations (n = 7). This study confirmed that both geographical distance and environmental factors interact to shape population structure in this species and revealed the existence of three separate stocks (Eastern Australia, Northern Australia, and Western Australia) across the geographical distribution of Australian *P. monodon*. Based on this dataset, Western Australia *P. monodon* appears to not have undergone a bottleneck (i.e., heterozygosity not reduced) as previously implied in other studies (Benzie, 2000; Benzie et al., 1992), but rather has retained similar heterozygosity to Northern Australia and Eastern Australia populations. The combination of population differentiation and environmental association approaches identified 89 putatively adaptive SNPs that revealed geographically discrete Australian *P. monodon* populations have adapted to region-specific environmental variables (e.g., sea surface temperature). Of these putatively adaptive SNPs, only 11 outlier SNPs matched *P. monodon* transcriptome contigs (Huerlimann et al., 2018) with subsequent protein translation of eight outlier SNPs demonstrating more non-synonymous changes than synonymous changes across all six reading frames. This study provides the most comprehensive and first SNP-based analysis of fine-scale Australian *P. monodon* population genomics to date, which can help aquaculture practices and fishery management strategies through improved stock identification, genetic diversity estimations, local adaptation determination, and selective breeding programs.

Chapter 3 addressed the knowledge gap about Indo-Pacific *P. monodon* population genomics by generating a high-quality genome-wide SNP dataset (n = 10,593 total) for each individual (n = 532 total) collected from geographically discrete populations (n = 16). This study revealed the existence of seven distinct genetic clusters and indicated migration to Southeast Asia (SEA) from the Western Indian Ocean (WIO). The scope of Indo-Pacific Ocean sampling covered most of the known *P. monodon* distribution and included 16 sampled sites across 10 countries (n = 532 individuals total). Both neutral (n = 9,930) and outlier (n = 663) loci datasets revealed a pattern of strong genetic structure of Indo-Pacific *P. monodon* corresponding with broad geographical regions and clear genetic breaks among samples within regions. Neutral loci revealed seven genetically distinct *P. monodon* stocks (Fiji: FJ, Eastern Australia: EA, Northern Australia: NA, Western Australia: WA, Southeast Asia: SEA, Sri Lanka: SLK, and WIO) and the separation of FJ and WIO clusters from all other clusters. 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 (Amaral et al., 2017; Hui et al., 2016; Huyghe & Kochzius, 2017; Vogler et al., 2012). Analyses based on neutral SNPs (n = 9,930) revealed: 1) five migration events that indicated gene flow to SEA from WIO sites, 2) gene flow among *P. monodon* populations in the Pacific region, and 3) partial connectivity among populations native to both oceans. Moreover, the combined environmental association (EA) and population differentiation (PD) approaches used for broad-scale analyses revealed full redundancy analysis (RDA) model support for the role of sea surface temperature in *P. monodon* SNP genotype distribution ($p < 0.001$; $R^2 = 0.058$; adjusted $R^2 = 0.055$; Chapter 3). Greater levels of genetic diversity were observed in SEA, EA, and NA *P. monodon* populations (Chapter 3); however, this observation does not preclude populations situated outside these regions from being potentially valuable as broodstock sources for genetic improvement programs. These genetic diversity insights can be utilized to improve the establishment of new base populations for *P. monodon* selective breeding programs by providing guidance regarding which wild populations to utilize as broodstock sources (i.e., highly genetically diverse) to ensure that broad genetic variation is captured. These findings were further supported by the identification of 26 putatively adaptive SNPs that exhibited significant Pearson correlation ($p < 0.05$) between minor allele frequency and maximum or minimum sea surface temperature (Chapter 3). A population that is potentially under local adaptive pressures may be an important source of private or rare alleles that can enhance population resistance to future environmental change (e.g., naturally or via selective breeding programs), or assist natural migration (Golbuu et al., 2016; Ofori et al., 2017). Selective breeding programs aiming to improve growth rate, pigmentation, pathogen resistance, or other characteristic(s) of *P. monodon* should consider further candidate gene discovery research into the specific functions and phenotype(s) associated with the identified putatively adaptive loci. Overall, this study provides the most comprehensive and first SNP-based analysis of broad-scale Indo-Pacific *P. monodon* population genomics to date, which can help aquaculture practices and fishery management strategies through improved stock identification, genetic connectivity establishment, genetic diversity estimation, local adaptation determination, and selective breeding programs.

Chapter 4 addressed the knowledge gap about Indo-Pacific *P. monodon* phylogeography and phylogenomic relationships with sister taxa by identifying genome-wide SNP and presence-absence variant (PAV) marker subsets (n = 4,496 and n = 7,054), respectively. More specifically, the seven tested Penaeidae species divided into the following three distinct and well-supported clades (BS and PP = 100): 1) *F. merguiensis*, *P. monodon, P.*

semisulcatus, *P. esculentus*, and *L. vannamei*, 2) *P. latisulcatus*, and 3) *P. longistylus*. These results confirmed the monophyletic structure of *P. monodon,* as well as the position of six sister taxa compared to non-SNP genetic marker studies (Hurzaid et al., 2020; Lavery et al., 2004; Ma et al., 2011). More specifically, phylogeographic structures revealed that Indo-Pacific *P. monodon* appears to be divided into two lineages: 1) Indo-Polynesian (containing SLK, SEA, Australia, and FJ) and 2) WIO (containing Kenya and South Africa). The phylogenomic reconstructions collectively support the broad-scale population genetic structure conclusion that the WIO lineage is ancestral to the Indo-Polynesian lineage and that SLK serves as a transition zone between these geographically distinct regions (Chapter 3). This study provides the first SNP- and PAV-based phylogenomic analysis of Indo-Pacific *P. monodon* and sister taxa to date, which can help aquaculture practices and fishery management strategies through improved stock identification, species determination, and genetic connectivity monitoring.

Chapter 5 addressed the commercial need for accurate and cost-effective population genomics analysis tools by identifying five low-density SNP panels from the broad-scale Indo-Pacific SNP dataset: 1) global diversity SNP panel (n = 2,155), 2) global provenance SNP panel (n = 1,200), 3) Pacific Ocean (PO) pedigree SNP panel (n = 217), 4) SLK pedigree SNP panel (n = 226), and 5) WIO pedigree SNP panel (n = 220). Low-density SNP panels were targeted to minimize costs associated with sequencing (Gebrehiwot et al., 2021; Herry et al., 2018; Song & Hu, 2021) and computation (Dominik et al., 2021) as well as potentially provide the following long-term advancements to the Indo-Pacific *P. monodon* aquaculture industry: 1) improved consistency in SNP genotype data across global studies (e.g., direct comparison across common SNP panels), 2) development of a global reference database of wild and farmed population genotypes without the need for additional and expensive genome-wide sequencing/genotyping, and 3) establishment of robust and a globally accepted standard for genetic diversity, provenance, and pedigree analyses. The identified global diversity SNP panel (n = 2,155) and its three smaller subsets (n = 1,500, 1,000, and 500 SNPs) were able to estimate the genetic diversity within and among all Indo-Pacific *P. monodon* populations as accurately as the entire broad-scale SNP dataset (n = 10,593; Chapter 3). Similarly, mean A_R, H_O, and H_E estimated with the full diversity SNP panel (n = 2,155) and three smaller subsets (n = 500 – 1,500 SNPs) were significantly correlated ($R \geq 0.89$; $p < 0.00001$) with the mean A_R, H_O, and H_E per population estimated with the entire high-quality SNP dataset (n = 10,593; Chapter 3). The full diversity SNP panel and all three smaller subsets also reduced the F_{IS} values observed for the KE and SA populations from 0.80 and 0.78 to 0.03 and 0.02, respectively, which closely

reflected the reduction in F_{IS} values observed during high-quality Indo-Pacific *P. monodon* SNP dataset analysis (Chapter 3). Based on these comparisons, the full diversity SNP panel (*n* = 2,155; Figure 2) provides accurate and cost-effective determination of genetic diversity within and among Indo-Pacific *P. monodon* populations while the smaller subsets (*n* = 500 – 1,500 SNPs) could provide more cost-effective data for specialized purposes depending on required sensitivity. The global provenance SNP panel (*n* = 1,200 SNPs) was identified from the entire high-quality Indo-Pacific *P. monodon* dataset (*n* = 10,593; Chapter 3) using three discrete approaches that were then merged (Chapter 5). Both DAPC and *NetView* analyses demonstrated that *P. monodon* collected from the same region could be traced back to their original basin (IO or PO) and genetic cluster (FJ, EA, NA, WA, SEA, SLK and WIO) using the provenance SNP panel while Monte-Carlo tests demonstrated the ability of the provenance SNP panel to accurately assign individuals back to their original basin, region, and sampling location within IO and PO basins; however, genotyping only 50 - 90% of individuals within the top 25% of F_{ST} ranked markers (*n* = 300) can be sufficient for accurate provenance determination depending on the required level of resolution (Chapter 5). The identified region-specific (PO, SLK, and WIO) pedigree SNP panels (*n* = 217, 226, and 220), respectively, were able to accurately assign theoretical progenies back to their parent-pair (i.e., both parents known) *in silico* (Chapter 5). *In silico* simulations revealed that progenies from PO units required 35 or 40 SNP markers from the PO pedigree panel to reach 100% assignment accuracy when the parent-pair was known, whereas progenies from SLK and WIO units required 30 SNP markers from the SLK and WIO pedigree panels to reach 100% assignment accuracy when only one parent or the parent-pair was known, respectively. The relatively low number of SNP markers required to achieve high parentage assignment accuracy across all *in silico* simulations could be due to the relatively low number of candidate parents used (*n* = 18) and simulated population size (*n* = 100). Given that prawn aquaculture selective breeding programs typically use more candidate parents (i.e., broodstock) and generate more progenies than were tested *in silico* (Chapter 5), more SNP markers from the PO, SLK, or WIO pedigree SNP panel could be required to achieve high parentage assignment accuracy for individuals from PO, SLK, or WIO region populations, respectively (Arbon et al., 2021; May et al., 2020). This study provides the first global cost-effective and accurate low-density SNP panels for the Indo-Pacific *P. monodon* industry, which can help aquaculture practices and fishery management strategies through improved diversity, provenance, and pedigree determination as well as facilitation of global data consistency, global database development, and global analysis standards establishment.

Using high-quality genome-wide SNP data, Chapters 2 – 5 provide the most comprehensive evaluation of Australian (fine-scale) and Indo-Pacific (broad-scale) *P. monodon* population genomics to date as well as identify a suite of accurate and cost-effective genetic tools for global industry use. Taken together, these four interdependent studies provide new population genomic insights and resources that contribute towards closing the current knowledge gaps about key aspects of *P. monodon* aquaculture and fishery management practices.

6.2 Future research

6.2.1 Improved origin and demographic history reconstruction for Indo-Pacific black tiger shrimp *P. monodon*

The demographic structure of populations is affected by life history strategies and how these interact with exploitation, climate change, and biotic interaction (Ohlberger et al., 2018). Thus, inference of demographic history can help to understand responses to past environmental changes, making them relevant for management strategies (Barth et al., 2017). For *P. monodon*, the demographic and evolutionary history within and among Indo-Pacific populations and genetic clusters remains unresolved, particularly as it relates to divergence and dispersal events in response to migration since the late Pleistocene. Fine- and broad-scale analyses were performed on Australian and Indo-Pacific *P. monodon* populations (Chapters 2 and 3), respectively, which provided novel and significant advances in understanding the distribution of genetic diversity, population structure, and environmental factors influencing population structure (e.g., existence of seven genetic clusters across all sampled Indo-Pacific *P. monodon* populations); however, the sampled Indo-Pacific populations ($n = 16$) and individuals ($n = 532$) did not include individuals from North Pacific Ocean (NPO) populations (e.g., China, Taiwan, and Japan) or additional IO populations (e.g., Bangladesh, India, and Madagascar). Additionally, the exact divergence time between the ancestral (WIO populations) and Indo-Polynesian lineage (SLK, SEA, Australia, and FJ) was not able to be conclusively determined using phylogenomic analyses (Chapter 4); however, based on the reported lineage divergence timing of other species endemic to and spanning the Indo-Pacific region, the exact divergence time might have occurred during the late Pleistocene (Farhadi et al., 2017; Hui et al., 2016; Huyghe & Kochzius, 2017; Vogler et al., 2012). Some of these limitations have recently been

overcome by the publication of a chromosome-scale genome assembly for *P. monodon* (Uengwetwanit et al., 2021) but this resource was not available while this phylogenomics study was undertaken. In light of these limitations, future studies should aim to incorporate genome-wide SNP sequences for NPO and additional IO populations to more completely reconstruct the origin and demographic history of Indo-Pacific *P. monodon* populations.

6.2.2 Advancement of *P. monodon* selective breeding programs

The giant black tiger shrimp is the second most farmed Penaeid shrimp species globally (FAO, 2020). The limited production of *P. monodon* may be due to the lack of genetically improved stocks, inconsistent seedstock quality, and numerous devastating disease outbreaks (Guppy et al., 2020; S. V. Nguyen et al., 2020). Additionally, *P. monodon* aquaculture development in terms of selective breeding programs and characterisation of commercially important traits remains largely under-developed (Guppy et al., 2020; Norman-López et al., 2015). Chapter 3 provides the global *P. monodon* industry with a better understanding of the distribution of genetic diversity and genetic differentiation within and among Indo-Pacific populations, which can advance the efficiency and success of selective breeding programs through improved selection of genetically diverse broodstocks; however, the high-quality SNP dataset (n = 10,593; Chapter 3) and SNP panels (n = 210 – 2,155; Chapter 5) that can be used for guided broodstock selection have not been linked to specific phenotypic traits of commercial interest. Thus, future research should aim to elucidate which of the high-quality SNPs (or regions of the genome) are directly linked to phenotype(s) of interest for selective breeding programs (e.g., increased fecundity, enhanced growth rate, altered pigmentation, improved disease resistance, specific meat qualities, etc).

At both fine- and broad-scales *P. monodon* population structure exhibited signatures of local adaption to maximum or minimum sea surface temperature (Chapters 2 and 3). Surface temperature maximum and minimum occur during summer and winter, respectively, which is when *P. monodon* broodstock emigrate out of estuaries into foreshores for spawning and, thus, extreme temperature swings could influence individual fitness (e.g., breeding success) given the known effects of temperature on *P. monodon* (Chen & Chen, 1999; Jackson & Wang, 1998; Kurmaly et al., 1989; Vuthiphandchai et al., 2005). Thus, future studies should aim to elucidate if *P. monodon* thermal tolerance can be improved by using the temperature-correlated outlier

SNP markers (i.e., identify a thermal tolerance SNP panel) to select potentially tolerant broodstock from different regions for selective breeding trials.

Additionally, SNP markers that were found to match transcriptome contig annotations suggest putatively adaptive involvement in cellular and metabolic processes, pigmentation, immune response, neuronal and cardiac function, and currently unknown functions (Chapters 2 and 3). As such, future *P. monodon* gene-by-environment studies should consider investigating the potential role of these associated genes in local adaptation to population-specific environmental conditions in order to potentially develop trait-specific selective breeding programs (e.g., broodstock selection or genetic editing).

6.2.3 *In situ* validation of identified low-density SNP panels

Genomic resources available in giant black tiger shrimp research are growing rapidly. Over the past two decades several studies have investigated the genetic architecture of *P. monodon*, including fosmid library end sequencing (Huang et al., 2011), linkage map construction (Baranski et al., 2014), transcriptome (Huerlimann et al., 2018; Pootakham et al., 2020; Uengwetwanit et al., 2018), whole-genome assembly (Van Quyen et al., 2020), and chromosome-scale genome assembly (Uengwetwanit et al., 2021). DArTseq™ was utilized to generate a high-quality genome-wide SNP genotype database (n = 10,539) for each *P. monodon* individual (n = 532 total) collected from 16 discrete Indo-Pacific Ocean locations (Chapter 3). However, considering the larger number of animals that would require SNP genotyping for commercial applications (i.e., selective breeding programs), the use of high-density SNP panels (n > 10,000 loci) may not be as cost-effective as low-density SNP panels (n < 10,000 loci). For example, Duarte et al. (2013) reported that the cost of genotyping pigs with a high-density commercial SNP panel (n = 60,000) was more than twice that of genotyping the same individuals with a low-density commercial SNP panel (n = 9,000). As such, there is a commercial need to identify low-density SNP panels for rapid, robust, and cost-effective use in global diversity, provenance, and/or pedigree determination (Dierens et al., 2014; Guppy et al., 2020; Henshall et al., 2014; Sellars et al., 2014). To address this commercial need for accurate and cost-effective population genomics tools, five low-density SNP panels were identified from the broad-scale Indo-Pacific SNP dataset; however, the identified low-density SNP panels were only validated *in silico* (Chapter 5). Therefore, future studies should aim to validate these five low-density SNP panels *in situ* using the workflows developed for Atlantic

salmon (*Salmo salar* (Holman et al., 2017)) and Florida bass (*Micropterus floridanus*; (Zhao et al., 2018)) as references. These two studies indicated the utility of simple and low-cost genotyping-by-sequencing techniques for SNP discovery and the relatively small number of variable SNPs by using Fluidigm SNPtype SNP genotyping assays (Holman et al., 2017) and Agena MassARRAY technology (Zhao et al., 2018) to develop parentage SNP panels for Atlantic salmon (*S. salar* L.) and Florida bass (*M. floridanus*), respectively.

References

Adamkewicz, S. L., & Harasewych, M. (1996). Systematics and biogeography of the genus *Donax* (Bivalvia: Donacidae) in eastern North America. *American Malacological Bulletin, 13*, 97-104.

Aguirre-Pabón, J., Orozco Berdugo Jr, G., & Narvaez Barandica, J. (2015). Genetic status, source and establishment risk of the giant tiger shrimp (Penaeidae: *Penaeus monodon*), an invasive species in Colombian Caribbean waters. *Acta Biológica Colombiana, 20*(1), 117-127.

Ahti, P. A., Coleman, R. R., DiBattista, J. D., Berumen, M. L., Rocha, L. A., & Bowen, B. W. (2016). Phylogeography of Indo-Pacific reef fishes: sister wrasses Coris gaimard and *C. cuvieri* in the Red Sea, Indian Ocean and Pacific Ocean. *Journal of Biogeography, 43*(6), 1103-1115.

Al-Breiki, R. D., Kjeldsen, S. R., Afzal, H., Al Hinai, M. S., Zenger, K. R., Jerry, D. R., . . . Delghandi, M. (2018). Genome-wide SNP analyses reveal high gene flow and signatures of local adaptation among the scalloped spiny lobster (*Panulirus homarus*) along the Omani coastline. *BMC genomics, 19*(1), 690.

Alam, M. M., & Pálsson, S. (2016). Population structure of the giant tiger shrimp *Penaeus monodon* in Bangladesh based on variation in microsatellites and immune-related genes. *Marine Biology Research, 12*(7), 706-714.

Alam, M. M., Westfall, K. M., & Pálsson, S. (2016). Mitogenomic variation of Bangladesh *Penaeus monodon* (Decapoda, Panaeidae) and reassessment of its phylogeography in the Indo-West Pacific region. *Hydrobiologia, 763*(1), 249-265. doi:10.1007/s10750-015-2381-3

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research, 19*(9), 1655-1664. doi:10.1101/gr.094052.109

Alfaro-Montoya, J., Monge-Ortiz, A. M., Martínez-Fernández, D., & Herrera-Quesada, E. (2015). First record of the nonindigenous *Penaeus monodon* Fabricius, 1798 (Penaeidae) in the Caribbean Sea of Costa Rica, Central America, with observations on selected aspects of its reproductive biology. BioInvasions Record, 4(3), 217-222. doi:http://dx.doi.org/10.3391/bir.2015.4.3.11

Alho, J. S., VÄLIMÄKI, K., & MERILÄ, J. (2010). Rhh: an R extension for estimating multilocus heterozygosity and heterozygosity–heterozygosity correlation. *Molecular Ecology Resources, 10*(4), 720-722. doi:10.1111/j.1755-0998.2010.02830.x

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics, 11*(10), 697-709.

Allendorf, F. W., & Luikart, G. (2007). Conservation and the genetics of populations. *mammalia, 2007*, 189-197.

Amaral, A. R., Smith, B. D., Mansur, R. M., Brownell, R. L., & Rosenbaum, H. C. (2017). Oceanographic drivers of population differentiation in Indo-Pacific bottlenose (*Tursiops aduncus*) and humpback (*Sousa spp*.) dolphins of the northern Bay of Bengal. *Conservation Genetics, 18*(2), 371-381.

Anderson, E. (2010). Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Molecular Ecology Resources, 10*(4), 701-710.

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. *Ecology and Evolution, 9*(18), 10354-10364. doi:10.1002/ece3.5554

Andrew, S. C., Jensen, H., Hagen, I. J., Lundregan, S., & Griffith, S. C. (2018). Signatures of genetic adaptation to extremely varied Australian environments in introduced European house sparrows. *Molecular Ecology, 27*(22), 4542-4555.

Andrews, K. R., & Luikart, G. (2014). Recent novel approaches for population genomics data analysis. *Molecular Ecology, 23*(7), 1661-1667.

Araneda, C., Larraín, M. A., Hecht, B., & Narum, S. (2016). Adaptive genetic variation distinguishes Chilean blue mussels (*Mytilus chilensis*) from different marine environments. *Ecology and Evolution, 6*(11), 3632-3644.

Arbon, P. M., Silva, C. N., Jones, D. B., Jaccoud, D., Gervis, M., Jerry, D. R., & Strugnell, J. M. (2021). Development and validation of a SNP-based genotyping tool for pedigree establishment in Australian greenlip abalone *Haliotis laevigata* Donovan, 1808. *Aquaculture Reports, 20*, 100746.

Arnold, S. J., Coman, G. J., & Emerenciano, M. (2013). Constraints on seedstock production in eighth generation domesticated *Penaeus monodon* broodstock. *Aquaculture, 410*, 95-100.

Askari, G., Shabani, A., & Kolangi Miandare, H. (2013). Application of molecular markers in fisheries and aquaculture. *J Animal Sci, 2*, 82-88.

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., & De Clerck, O. (2018). Bio-ORACLE v2. 0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography, 27*(3), 277-284.

Avise, J. C. (2000). *Phylogeography: the history and formation of species*: Harvard university press.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology, 13*(4), 729-744.

Baranski, M., Gopikrishna, G., Robinson, N. A., Katneni, V. K., Shekhar, M. S., Shanmugakarthik, J., . . . Kent, M. (2014). The development of a high density linkage map for black tiger shrimp (*Penaeus monodon*) based on cSNPs. *PloS one, 9*(1), e85413.

Barth, J. M., Damerau, M., Matschiner, M., Jentoft, S., & Hanel, R. (2017). Genomic differentiation and demographic histories of Atlantic and Indo-Pacific yellowfin tuna (*Thunnus albacares*) populations. *Genome biology and evolution, 9*(4), 1084-1098.

Batista, P. D., Janes, J. K., Boone, C. K., Murray, B. W., & Sperling, F. A. (2016). Adaptive and neutral markers both show continent-wide population structure of mountain pine beetle (*Dendroctonus ponderosae*). *Ecology and Evolution, 6*(17), 6292-6300. doi:10.1002/ece3.2367

Beaumont, M. A., & Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology, 13*(4), 969-980.

Benestan, L. (2019). Population genomics applied to fishery management and conservation. In *Population genomics: Marine organisms* (pp. 399-421): Springer.

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology, 24*(13), 3299-3315. doi:10.1111/mec.13245

Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., . . . Bernatchez, L. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Molecular Ecology, 25*(20), 5073-5092. doi:10.1111/mec.13811

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

Benzie, J. (2000). Population genetic structure in penaeid prawns. *Aquaculture Research, 31*(1), 95-119. doi:DOI 10.1046/j.1365-2109.2000.00412.x

Benzie, J., Ballment, E., Forbes, A., Demetriades, N., Sugama, K., & Moria, S. (2002). Mitochondrial DNA variation in Indo-Pacific populations of the giant tiger prawn, *Penaeus monodon*. *Molecular Ecology, 11*(12), 2553-2569. doi:10.1046/j.1365-294X.2002.01638.x

Benzie, J., Ballment, E., & Frusher, S. (1993). Genetic structure of *Penaeus monodon* in Australia: concordant results from mtDNA and allozymes. *Aquaculture, 111*(1-4), 89-93.

Benzie, J., Frusher, S., & Ballment, E. (1992). Geographical variation ion allozyme frequencies of populations of *Penaeus monodon* (Crustacea: Decapoda) in Australia. *Marine and Freshwater Research, 43*(4), 715-725.

Bernal Rodríguez, C., Garcia, 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., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M., Beacham, T. D., . . . Naish, K. A. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends in ecology & evolution, 32*(9), 665-680.

Bernatchez, S., Xuereb, A., Laporte, M., Benestan, L., Steeves, R., Laflamme, M., . . . Mallet, M. A. (2018). Seascape genomics of eastern oyster (*Crassostrea virginica*) along the Atlantic coast of Canada. *Evolutionary Applications*.

Bernatchez, S., Xuereb, A., Laporte, M., Benestan, L., Steeves, R., Laflamme, M., . . . Mallet, M. A. (2019). Seascape genomics of eastern oyster (*Crassostrea virginica*) along the Atlantic coast of Canada. *Evolutionary Applications, 12*(3), 587-609. doi:10.1111/eva.12741

Bester-van der Merwe, A., Maduna, S., Hull, K., Bell, J., Rossouw, C., & Wintner, S. (2019). Evidence for multiple paternity and confirmation of an Indo-Pacific origin of blacktip shark *Carcharhinus limbatus* occurring in South Africa. *African Journal of Marine Science, 41*(3), 281-289. doi:10.2989/1814232x.2019.1653991

Bowen, B. W., Gaither, M. R., DiBattista, J. D., Iacchei, M., Andrews, K. R., Grant, W. S., . . . Briggs, J. C. (2016). Comparative phylogeography of the ocean planet. *Proceedings of the National Academy of Sciences, 113*(29), 7962-7969.

Bradbury, I. R., Hubert, S., Higgins, B., Bowman, S., Borza, T., Paterson, I. G., . . . Hardie, D. (2013). Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evolutionary Applications, 6*(3), 450-461.

Briggs, M. (2005). *Introductions and movement of two penaeid shrimp species in Asia and the Pacific*: Food & Agriculture Org.

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. *Marine Biology, 136*(1), 149-157.

Bruneaux, M., Johnston, S. E., Herczeg, G., Merilä, J., Primmer, C. R., & Vasemägi, A. (2013). Molecular evolutionary and population genomic analysis of the nine-spined stickleback using a modified restriction-site-associated DNA tag approach. *Molecular Ecology, 22*(3), 565-582.

Buckworth, R., Newman, S., Ovenden, J., Lester, R., & McPherson, G. (1998). The stock structure of northern and western Australian Spanish Mackerel. Final Report. *Fisheries Research and Development Corporation Project, 159*, 225.

Budd, A. M., Hinton, T. M., Tonks, M., Cheers, S., & Wade, N. M. (2017). Rapid expansion of pigmentation genes in penaeid shrimp with absolute preservation of function. *Journal of Experimental Biology, 220*(22), 4109-4118. doi:10.1242/jeb.164988

Calo-Mata, P., Pascoal, A., Fernández-No, I., Böhme, K., Gallardo, J. M., & Barros-Velázquez, J. (2009). Evaluation of a novel 16S rRNA/tRNAVal mitochondrial marker for the identification and phylogenetic analysis of shrimp species belonging to the superfamily Penaeoidea. *Analytical biochemistry, 391*(2), 127-134.

Candy, J. R., Campbell, N. R., Grinnell, M. H., Beacham, T. D., Larson, W. A., & Narum, S. R. (2015). Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. *Molecular Ecology Resources, 15*(6), 1421-1434.

Casey, J., Jardim, E., & Martinsohn, J. T. (2016). The role of genetics in fisheries management under the EU common fisheries policy. *Journal of Fish Biology*, *89*(6), 2755-2767.

Chakraborty, R., Meagher, T., & Smouse, P. (1988). Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. *Genetics, 118*(3), 527-536.

Chan, T.-Y., Tong, J., Tam, Y. K., & Chu, K. H. (2008). Phylogenetic relationships among the genera of the Penaeidae (Crustacea: Decapoda) revealed by mitochondrial 16S rRNA gene sequences. *Zootaxa, 1694*(3), 38.

Chan, T. (1998). Shrimps and prawns. In 'The living marine resources of the western central Pacific, Vol. 2. *FAO Species identification guide for fishery purposes.(Eds KE Carpenter and VH Niem)(FAO: Rome)*.

Chen, H.-Y., & Chen, Y.-L. (1999). Temperature preferendum of postlarval black tiger shrimp (*Penaeus monodon*). *Marine and Freshwater Research, 50*(1), 67-70.

Chen, K. Y., Marschall, E. A., Sovic, M. G., Fries, A. C., Gibbs, H. L., & Ludsin, S. A. (2018). *assign POP*: An R package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. *Methods in Ecology and Evolution, 9*(2), 439-446. doi:https://doi.org/10.1111/2041-210X.12897.

Chenoweth, S. F., Hughes, J. M., Keenan, C. P., & Lavery, S. (1998). Concordance between dispersal and mitochondrial gene flow: isolation by distance in a tropical teleost, *Lates calcarifer* (Australian barramundi). *Heredity, 80*(2), 187-197.

Chessel, D., Dufour, A. B., & Thioulouse, J. (2004). The ade4 package-I-One-table methods. R news, 4(1), 5-10.

Chivas, A. R., García, A., van der Kaars, S., Couapel, M. J., Holt, S., Reeves, J. M., . . . Banerjee, D. (2001). Sea-level and environmental changes since the last interglacial in the Gulf of Carpentaria, Australia: an overview. *Quaternary International, 83*, 19-46.

Choy, S. (1982). The biology of littoral penaeid prawns in Fiji waters with particular reference to *Penaeus (Melicertus) canaliculatus* Pérez-Farfante. *The University of the South Pacific, Suva, Fiji, 161*.

Coman, G., Arnold, S., Peixoto, S., Crocos, P., Coman, F., & Preston, N. (2006). Reproductive performance of reciprocally crossed wild-caught and tank-reared *Penaeus monodon* broodstock. *Aquaculture, 252*(2-4), 372-384.

Cowen, R. K., Paris, C. B., & Srinivasan, A. (2006). Scaling of connectivity in marine populations. *Science, 311*(5760), 522-527.

Cowen, R. K., & Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Annual review of marine science, 1*, 443-466.

Crandall, E. D., Riginos, C., Bird, C. E., Liggins, L., Treml, E., Beger, M., . . . DiBattista, J. D. (2019). The molecular biogeography of the Indo-Pacific: Testing hypotheses with multispecies genetic patterns. *Global Ecology and Biogeography, 28*(7), 943-960. doi:10.1111/geb.12905

D'ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S., Poncet, C., . . . Charles, M. (2019). Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genetics Selection Evolution, 51*(1), 1-15.

Dabney, A., Storey, J. D., & Warnes, G. (2010). qvalue: Q-value estimation for false discovery rate control. *R package version, 1*(0). Retrieved from http://github.com/jdstorey/qvalue

Dall, W., Hill, B., Rothlisberg, P., & Sharples, D. (1990). The biology of the Penaeidae. *The biology of the Penaeidae., 27*.

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.

Dao, H. T., Smith-Keune, C., Wolanski, E., Jones, C. M., & Jerry, D. R. (2015). Oceanographic currents and local ecological knowledge indicate, and genetics does not refute, a contemporary pattern of larval dispersal for the ornate spiny lobster, *Panulirus ornatus* in the South-East Asian Archipelago. *PloS one, 10(5)*, e0124568. doi:10.1371/journal.pone.0124568.

Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., & Blaxter, M. L. (2013). Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology, 22*(11), 3151-3164. doi:10.1111/mec.12084

De Grave,S. and Fransen, C. (2011). Carideorum catalogus: the recent species of the dendrobranchiate, stenopodidean, procarididean and caridean shrimps (Crustacea: Decapoda).Zool. Meded., 85 (2011), pp. 195-589.

De Meeûs, T. (2017). Revisiting F IS, F ST, Wahlund effects, and null alleles. *Journal of Heredity, 109*(4), 446-456.

De Mita, S., Thuillet, A. C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: analysis of eight methods

and their effectiveness for outbreeding and selfing populations. *Molecular Ecology, 22*(5), 1383-1399.

de Souza Junior, E. M., Júnior, J. G., do Nascimento Araújo, P. V., Alencar, C. E. R. D., & de Morais Freire, F. A. (2015). Second report of the occurrence of giant tiger prawn, *Penaeus monodon* Fabricus, 1798 (Crustacea: Decapoda), in Rio Grande Do Norte State, Northest Brazil. *Arquivos de Ciências do Mar, 48*(2), 116-120.

DESA, U. (2019). World Population Prospects 2019. United Nations. Department of Economic and Social Affairs. *World Population Prospects 2019*.

Dierens, L., Henshall, J., & Sellars, M. (2014). An industry friendly, inexpensive DNA extraction method for Penaeid shrimp that is compatible with Sequenom® iPLEX Platinum SNP pedigree genotyping platforms. *Aquaculture, 433*, 102-104.

Do, C., Waples, R. S., Peel, D., Macbeth, G., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources, 14*(1), 209-214.

Dominik, S., Duff, C., Byrne, A., Daetwyler, H., & Reverter, A. (2021). Ultra-small SNP panels to uniquely identify individuals in thousands of samples. *Animal Production Science*, 61, 1796-1800. doi:https://doi.org/10.1071/AN21123.

Dray, S., & Dufour, A.-B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software, 22*(4), 1-20. doi:10.18637/jss.v022.i04

Drinan, D. P., Gruenthal, K. M., Canino, M. F., Lowry, D., Fisher, M. C., & Hauser, L. (2018). Population assignment and local adaptation along an isolation-by-distance gradient in Pacific cod (*Gadus macrocephalus*). *Evolutionary Applications, 11*(8), 1448-1464.

Duarte, J. L. G., Bates, R. O., Ernst, C. W., Raney, N. E., Cantet, R. J., & Steibel, J. P. (2013). Genotype imputation accuracy in a F2 pig population using high density and low density SNP panels. *BMC genetics, 14*(1), 1-13.

Dudgeon, C. L., Blower, D., Broderick, D., Giles, J., Holmes, B. J., Kashiwagi, T., . . . Ovenden, J. (2012). A review of the application of molecular genetics for fisheries management and conservation of sharks and rays. *Journal of Fish Biology, 80*(5), 1789-1843.

Edwards, S. V., Shultz, A. J., & Campbell-Staton, S. C. (2015). Next-generation sequencing and the expanding domain of phylogeography. *Journal of Vertebrate Biology, 64*(3), 187-206.

El Sheikha, A. F., & Xu, J. (2017). Traceability as a key of seafood safety: reassessment and possible applications. *Reviews in fisheries science & aquaculture, 25*(2), 158-170.

Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics, 1*, 117693430500100003. doi:10.1177/117693430500100003

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources, 10*(3), 564-567. doi:10.1111/j.1755-0998.2010.02847.x

Fan, S., Wang, F., Xie, Z., Zhao, C., Wang, P., Yan, L., . . . Qiu, L. (2021). Molecular characterization and functional analysis of scavenger receptor class B from black tiger shrimp (*Penaeus monodon*). *Electronic Journal of Biotechnology, 51*, 40-49.

FAO. (2005-2021). Cultured Aquatic Species Information Programme. *Penaeus monodon*. *FAO Fisheries and Aquaculture Division [online], 2021*(17/10/21).

FAO. (2020). *The State of World Fisheries and Aquaculture 2020*. Rome, Italy: Food and Agriculture Organization.

FAO, IFAD, UNICEF, WFP, & WHO. (2021). *The State of Food Security and Nutrition in the World (SOFI). Transforming food systems for food security, improved nutrition and affordable healthy diets for all*. Rome, Italy: FAO.

Farhadi, A., Jeffs, A. G., Farahmand, H., Rejiniemon, T. S., Smith, G., & Lavery, S. D. (2017). Mechanisms of peripheral phylogeographic divergence in the indo-Pacific: lessons from the spiny lobster *Panulirus homarus*. *BMC Evolutionary Biology, 17*(1), 195.

Ferchaud, A.-L., Pedersen, S. H., Bekkevold, D., Jian, J., Niu, Y., & Hansen, M. M. (2014). A low-density SNP array for analyzing differential selection in freshwater and marine populations of threespine stickleback (*Gasterosteus aculeatus*). *BMC genomics, 15*(1), 1-11.

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. *Journal of Sea Research, 76*, 146-153. doi:10.1016/j.seares.2012.08.004

Flanagan, S. P., & Jones, A. G. (2019). The future of parentage analysis: From microsatellites to SNPs and beyond. *Molecular Ecology, 28*(3), 544-567.

Foll, M. (2012). Bayescan v2. 1 user manual. *Ecology, 20*, 1450-1462. Retrieved from http://cmpg.unibe.ch/software/BayeScan/files/BayeScan2.1_manual.pdf

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. doi:10.1534/genetics.108.092221

Foote, A., Simma, D., Khatkar, M., Raadsma, H., Guppy, J., Coman, G., . . . Wade, N. (2019). Considerations for maintaining family diversity in commercially mass-spawned penaeid shrimp: a case study on *Penaeus monodon*. *Frontiers in genetics, 10*, 1127.

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. *African Journal of Marine Science, 21*, 271-277. doi:10.2989/025776199784125881

Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology, 25*(1), 104-120.

Frankham, R., Ballou, J., & Briscoe, D. (2002). Introduction to Conservation Genetics Cambridge University Press. *Cambridge, UK, 617*.

Frichot, E., & François, O. (2015). LEA: an R package for landscape and ecological association studies. *Methods in Ecology and Evolution, 6*(8), 925-929.

Fu, Y.-B., Cheng, B., & Peterson, G. W. (2014). Genetic diversity analysis of yellow mustard (*Sinapis alba L.*) germplasm based on genotyping by sequencing. *Genetic Resources and Crop Evolution, 61*(3), 579-594.

Funge-Smith, S., & Briggs, M. (2003). The introduction of *Penaeus vannamei* and *P. stylirostris* into the Asia-Pacific region. *International Mechanisms for the Control and Responsible Use of Alien Species in Aquatic Ecosystems, Jinghong, Xishuangbanna, People's Republic of China*, 26-29.

Funk, W. C., Lovich, R. E., Hohenlohe, P. A., Hofman, C. A., Morrison, S. A., Sillett, T. S., . . . Day, M. D. (2016). Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). *Molecular Ecology, 25*(10), 2176-2194.

Gagnaire, P. A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., . . . Bierne, N. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications, 8*(8), 769-786.

Gaither, M. R., Jones, S. A., Kelley, C., Newman, S. J., Sorenson, L., & Bowen, B. W. (2011). High connectivity in the deepwater snapper *Pristipomoides filamentosus* (Lutjanidae) across the Indo-Pacific with isolation of the Hawaiian Archipelago. *PloS one, 6*(12). doi:10.1371/journal.pone.0028913

Gaither, M. R., Toonen, R. J., Robertson, D. R., Planes, S., & Bowen, B. W. (2010). Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira and Lutjanus fulvus*). *Journal of Biogeography,* 37, 133-147. doi:10.1111/j.1365-2699.2009.02188.x

Galloway, R. W. (1981). Late cainozoic environments in Australia. *Ecological biogeography of Australia*, 53-80.

Gandon, S., & Michalakis, Y. (2002). Local adaptation, evolutionary potential and host–parasite coevolution: interactions between migration, mutation, population size and generation time. *Journal of evolutionary biology, 15*(3), 451-462.

GE, H. L. S. (2007). *Illustra AutoSeq G-50 and AutoSeq 96 dye terminator removal*. Retrieved from Illustra AutoSeq G-50 documents: https://gels.yilimart.com/Assets/Images/doc/file/17085501_Datafile_01.pdf

Gebrehiwot, N. Z., Strucken, E. M., Marshall, K., Aliloo, H., & Gibson, J. P. (2021). SNP panels for the estimation of dairy breed proportion and parentage assignment in African crossbred dairy cattle. *Genetics Selection Evolution, 53*(1), 1-18.

Gjedrem, T. (2012). Genetic improvement for the development of efficient global aquaculture: a personal opinion review. *Aquaculture, 344*, 12-22. doi:10.1016/j.aquaculture.2012.03.003

Gjedrem, T., & Robinson, N. (2014). Advances by selective breeding for aquatic species: a review. *Agricultural Sciences, 5*(12), 1152.

Golbuu, Y., Gouezo, M., Kurihara, H., Rehm, L., & Wolanski, E. (2016). Long-term isolation and local adaptation in Palau's Nikko Bay help corals thrive in acidic waters. *Coral Reefs, 35*(3), 909-918.

Gomes, I., Prinz, M., Pereira, R., Meyers, C., Mikulasovich, R. S., Amorim, A., . . . Gusmão, L. (2007). Genetic analysis of three US population groups using an X-chromosomal STR decaplex. *International journal of legal medicine, 121*(3), 198-203.

Gopurenko, D., & Hughes, J. M. (2002). Regional patterns of genetic structure among Australian populations of the mud crab, *Scylla serrata* (Crustacea: Decapoda): evidence from mitochondrial DNA. *Marine and Freshwater Research, 53*(5), 849-857.

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software, 22*(7), 1-19.

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular ecology notes, 5*(1), 184-186.

Grant, W. S., Jasper, J., Bekkevold, D., & Adkison, M. (2017). Responsible genetic approach to stock restoration, sea ranching and stock enhancement of marine fishes and invertebrates. *Reviews in Fish Biology and Fisheries, 27*(3), 615-649. doi:10.1007/s11160-017-9489-7

Gribble, N., Atfield, J., Dredge, M., White, D., & Kistle, S. (2003). *Sustainable Penaeus monodon (*Black Tiger prawn*) Populations for Broodstock Supply* the Department of Primary Industries, Queensland: Brisbane : Queensland Dept. of Primary Industries.

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2017). dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, *18*(3), 691-699. doi:https://doi-org.elibrary.jcu.edu.au/10.1111/1755-0998.12745.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology, 59*(3), 307-321.

Guppy, J. L., Jones, D. B., Kjeldsen, S. R., Le Port, A., Khatkar, M. S., Wade, N. M., . . . Jerry, D. R. (2020). Development and validation of a RAD-Seq target-capture based genotyping assay for routine application in advanced black tiger shrimp (*Penaeus monodon*) breeding programs. *BMC genomics, 21*(1), 1-16.

Hale, M. L., Burg, T. M., & Steeves, T. E. (2012). Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. *PloS one, 7*(9). doi:https://doi.org/10.1371/journal.pone.0045170

Hänfling, B., Edwards, F., & Gherardi, F. (2011). Invasive alien Crustacea: dispersal, establishment, impact and control. *BioControl, 56*(4), 573-595.

Harrisson, K. A., Amish, S. J., Pavlova, A., Narum, S. R., Telonis-Scott, M., Rourke, M. L., . . . Ingram, B. A. (2017). Signatures of polygenic adaptation associated with climate across the range of a threatened fish species with high genetic connectivity. *Molecular Ecology, 26*(22), 6253-6269.

Harrisson, K. A., Pavlova, A., Telonis-Scott, M., & Sunnucks, P. (2014). Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications, 7*(9), 1008-1025.

Hartl, D. L., Clark, A. G., & Clark, A. G. (1997). *Principles of population genetics* (Vol. 116): Sinauer associates Sunderland, MA.

Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic biology, 65*(5), 910-924.

Heaton, M. P., Leymaster, K. A., Kalbfleisch, T. S., Kijas, J. W., Clarke, S. M., McEwan, J., . . . Simpson, B. (2014). SNPs for parentage testing and traceability in globally diverse breeds of sheep. *PloS one, 9*(4), e94851.

Hecht, B. C., Matala, A. P., Hess, J. E., & Narum, S. R. (2015). Environmental adaptation in Chinook salmon (Oncorhynchus tshawytscha) throughout their North American range. *Molecular Ecology, 24*(22), 5573-5595.

Hemmer-Hansen, J., Nielsen, E. E., GRØNKJÆR, P., & Loeschcke, V. (2007). Evolutionary mechanisms shaping the genetic population structure of marine fishes; lessons from the European flounder (*Platichthys flesus* L.). *Molecular Ecology, 16*(15), 3104-3118. doi:10.1111/j.1365-294X.2007.03367.x

Hendricks, S., Anderson, E. C., Antao, T., Bernatchez, L., Forester, B. R., Garner, B., . . . Koop, B. (2018). Recent advances in conservation and population genomics data analysis. *Evolutionary Applications, 11*(8), 1197-1211.

Hendry, A. P., & Taylor, E. B. (2004). How much of the variation in adaptive divergence can be explained by gene flow? An evaluation using lake-stream stickleback pairs. *Evolution, 58*(10), 2319-2331.

Henshall, J. M., Dierens, L., & Sellars, M. J. (2014). Quantitative analysis of low-density SNP data for parentage assignment and estimation of family contributions to pooled samples. *Genetics Selection Evolution, 46*(1), 1-17.

Herry, F., Hérault, F., Druet, D. P., Varenne, A., Burlot, T., Le Roy, P., & Allais, S. (2018). Design of low density SNP chips for genotype imputation in layer chicken. *BMC genetics, 19*(1), 1-14.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular biology and evolution, 35*(2), 518-522.

Hodel, R. G., Chen, S., Payton, A. C., McDaniel, S. F., Soltis, P., & Soltis, D. E. (2017). Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. *Scientific reports, 7*(1), 1-14.

Holman, L. E., Onoufriou, A., Hillestad, B., & Johnston, I. A. (2017). A workflow used to design low density SNP panels for parentage assignment and traceability in aquaculture species and its validation in Atlantic salmon. *Aquaculture, 476*, 59-64.

Horne, J. B., van Herwerden, L., Choat, J. H., & Robertson, D. R. (2008). High population connectivity across the Indo-Pacific: congruent lack of phylogeographic structure in three reef fish congeners. *Molecular phylogenetics and evolution, 49*(2), 629-638. doi:10.1016/j.ympev.2008.08.023

Hu, D., Ansari, D., Zhou, Q., Sasor, A., Hilmersson, K. S., Bauden, M., . . . Andersson, R. (2018). Calcium-activated chloride channel regulator 1 as a prognostic biomarker in pancreatic ductal adenocarcinoma. *BMC cancer, 18*(1), 1096.

Hu, D., Pan, L., Zhao, Q., & Ren, Q. (2015). Transcriptomic response to low salinity stress in gills of the Pacific white shrimp, *Litopenaeus vannamei*. *Marine genomics, 24*, 297-304.

Hualkasin, W., Sirimontaporn, P., Chotigeat, W., Querci, J., & Phongdara, A. (2003). Molecular phylogenetic analysis of white prawns species and the existence of two clades in *Penaeus merguiensis*. *Journal of Experimental Marine Biology and Ecology, 296*(1), 1-11.

Huang, S.-W., Lin, Y.-Y., You, E.-M., Liu, T.-T., Shu, H.-Y., Wu, K.-M., . . . Ma, G.-C. (2011). Fosmid library end sequencing reveals a rarely known genome structure of marine shrimp *Penaeus monodon*. *BMC genomics, 12*(1), 1-19.

Huerlimann, R., Wade, N. M., Gordon, L., Montenegro, J. D., Goodall, J., McWilliam, S., . . . Donovan, D. (2018). De novo assembly, characterization, functional annotation and expression patterns of the black tiger shrimp (*Penaeus monodon*) transcriptome. *Scientific reports, 8*(1), 1-14. doi:10.1038/s41598-018-31148-4

Hui, M., Kraemer, W. E., Seidel, C., Nuryanto, A., Joshi, A., & Kochzius, M. (2016). Comparative genetic population structure of three endangered giant clams (Cardiidae: *Tridacna* species) throughout the Indo-West Pacific: implications for divergence,

connectivity and conservation. *Journal of Molluscan Studies, 82*(3), 403-414. doi:10.1093/mollus/eyw001

Hurzaid, A., Chan, T. Y., Mohd Nor, S. A., Muchlisin, Z. A., & Chen, W. J. (2020). Molecular phylogeny and diversity of penaeid shrimps (Crustacea: Decapoda) from South-East Asian waters. *Zoologica Scripta, 49*(5), 596-613.

Hutchison, J., Manica, A., Swetnam, R., Balmford, A., & Spalding, M. (2014). Predicting global patterns in mangrove forest biomass. *Conservation Letters, 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. *Marine Ecology, 38*(1), e12357. doi:10.1111/maec.12357

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). doi:0.1371/journal.pone.0193825

Jackson, C., & Wang, Y. G. (1998). Modelling growth rate of *Penaeus monodon* Fabricius in intensively managed ponds: effects of temperature, pond age and stocking density. *Aquaculture Research, 29*(1), 27-36.

Jeffery, N. W., Bradbury, I. R., Stanley, R. R., Wringe, B. F., Van Wyngaarden, M., Lowen, J. B., . . . DiBacco, C. (2018). Genomewide evidence of environmentally mediated secondary contact of European green crab *(Carcinus maenas*) lineages in eastern North America. *Evolutionary Applications, 11*(6), 869-882. doi:10.1111/eva.12601.

Jeffery, N. W., DiBacco, C., Van Wyngaarden, M., Hamilton, L. C., Stanley, R. R., Bernier, R., . . . Nadukkalam Ravindran, P. (2017). RAD sequencing reveals genomewide divergence between independent invasions of the European green crab (*Carcinus maenas*) in the Northwest Atlantic. *Ecology and Evolution, 7*(8), 2513-2524.

Jeffries, D. L., Copp, G. H., Lawson Handley, L., Olsén, K. H., Sayer, C. D., & Hänfling, B. (2016). Comparing RAD seq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*, L. *Molecular Ecology, 25*(13), 2997-3018.

Jenkins, T. L., Ellis, C. D., & Stevens, J. R. (2018). SNP discovery in European lobster (*Homarus gammarus*) using RAD sequencing. *Conservation Genetics Resources*, 1-5.

Jenkins, T. L., Ellis, C. D., Triantafyllidis, A., & Stevens, J. R. (2019). Single nucleotide polymorphisms reveal a genetic cline across the north-east Atlantic and enable powerful population assignment in the European lobster. *Evolutionary Applications, 12*(10), 1881-1899.

Jensen, J. D., Foll, M., & Bernatchez, L. (2016). The past, present and future of genomic scans for selection. *Molecular Ecology, 25*(1), 1-4.

Jerry, D. R., Evans, B. S., Kenway, M., & Wilson, K. (2006). Development of a microsatellite DNA parentage marker suite for black tiger shrimp *Penaeus monodon*. *Aquaculture, 255*(1-4), 542-547.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics, 24*(11), 1403-1405. doi:10.1093/bioinformatics/btn129

Jump, A. S., Hunt, J. M., Marti'nez-izquierdo, J. a., & Penuelas, J. (2006). Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in Fagus sylvatica. *Molecular Ecology, 15*(11), 3469-3480.

Kaka, R. M., Jung'a, J. O., Badamana, M., Ruwa, R. K., & Karisa, H. C. (2019). Morphometric length-weight relationships of wild penaeid shrimps in Malindi-Ungwana Bay: Implications to aquaculture development in Kenya. *The Egyptian Journal of Aquatic Research, 45*(2), 167-173.

Kalinowski, S. T. (2004). Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conservation Genetics, 5*(4), 539-543.

Kalinowski, S. T. (2005). hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular ecology notes, 5*(1), 187-189. doi:10.1111/j.1471-8286.2004.00845.x

Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology, 16*(5), 1099-1106.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods, 14*(6), 587-589.

Kassambara, A. (2018). ggpubr:'ggplot2'Based Publication Ready Plots. R package version 0.2. *Retrieved September*, 12, 2020.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Duran, C. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*(12), 1647-1649. doi:10.1093/bioinformatics/bts199

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution, 4*(8), 782-788. doi:10.1111/2041-210x.12067

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., . . . Hopper, C. (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. *Data Production and Analysis in Population Genomics: Methods and Protocols*, 67-89.

Kjeldsen, S. R., Raadsma, H. W., Leigh, K. A., Tobey, J. R., Phalen, D., Krockenberger, A., . . . Zenger, K. R. (2019). Genomic comparisons reveal biogeographic and anthropogenic impacts in the koala (*Phascolarctos cinereus*): a dietary-specialist species distributed across heterogeneous environments. *Heredity, 122*(5), 525-544.

Kongkeo, H. (2005–2020). Cultured Aquatic Species Information Programme: *Penaeus monodon*. Rome: Food & Agriculture Organization. https://www.fao.org/fishery/docs/CDrom/aquaculture/I1129m/file/en/en_gianttigerprawn.htm.

Kool, J. T., Moilanen, A., & Treml, E. A. (2013). Population connectivity: recent advances and new perspectives. *Landscape Ecology, 28*(2), 165-185.

Kriaridou, C., Tsairidou, S., Houston, R. D., & Robledo, D. (2020). Genomic prediction using low density marker panels in aquaculture: performance across species, traits, and genotyping platforms. *Frontiers in genetics, 11*, 124.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution, 35*(6), 1547-1549. doi:10.1093/molbev/msy096

Kurmaly, K., Yule, A., & Jones, D. (1989). Effects of body size and temperature on the metabolic rate of *Penaeus monodon*. *Marine Biology, 103*(1), 25-30.

Lal, M. M., Southgate, P. C., Jerry, D. R., Bosserelle, C., & Zenger, K. R. (2016). A parallel population genomic and hydrodynamic approach to fishery management of highly-dispersive marine invertebrates: the case of the Fijian black-lip pearl oyster *Pinctada margaritifera*. *PloS one, 11*(8), e0161390.

Lal, M. M., Southgate, P. C., Jerry, D. R., Bosserelle, C., & Zenger, K. R. (2017). Swept away: ocean currents and seascape features influence genetic structure across the 18,000 Km Indo-Pacific distribution of a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. *BMC genomics, 18*(1), 66. doi:10.1186/s12864-016-3410-y

Lal, M. M., Southgate, P. C., Jerry, D. R., & Zenger, K. R. (2018). Genome-wide comparisons reveal evidence for a species complex in the black-lip pearl oyster *Pinctada margaritifera* (Bivalvia: Pteriidae). *Scientific reports, 8*(1), 191. doi:10.1038/s41598-017-18602-5

Lavery, S., Chan, T., Tam, Y., & Chu, K. (2004). Phylogenetic relationships and evolutionary history of the shrimp genus Penaeus sl derived from mitochondrial DNA. *Molecular phylogenetics and evolution, 31*(1), 39-49.

Lawrie, D. S., Messer, P. W., Hershberg, R., & Petrov, D. A. (2013). Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genetics, 9*(5), e1003527.

Leaché, A. D., & Oaks, J. R. (2017). The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics, 48*, 69-84.

Leal, M. C., Pimentel, T., Ricardo, F., Rosa, R., & Calado, R. (2015). Seafood traceability: current needs, available tools, and biotechnological challenges for origin certification. *Trends in biotechnology, 33*(6), 331-336.

Lebeuf-Taylor, E., McCloskey, N., Bailey, S. F., Hinz, A., & Kassen, R. (2019). The distribution of fitness effects among synonymous mutations in a gene under directional selection. *eLife, 8*, e45952. doi:10.7554/elife.45952

LeBlanc, G. A. (2007). Crustacean endocrine toxicology: a review. *Ecotoxicology, 16*(1), 61-81.

Lee, S. Y. (2004). Relationship between mangrove abundance and tropical prawn production: a re-evaluation. *Marine Biology, 145*(5), 943-949.

Lemopoulos, A., Prokkola, J. M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., . . . Vainikka, A. (2019). Comparing RADseq and microsatellites for estimating genetic diversity and relatedness—Implications for brown trout conservation. *Ecology and Evolution, 9*(4), 2106-2120.

Lichstein, J. W. (2007). Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecology, 188*(2), 117-131.

Lind, C. E., Agyakwah, S. K., Attipoe, F. Y., Nugent, C., Crooijmans, R. P., & Toguyeni, A. (2019). Genetic diversity of Nile tilapia (*Oreochromis niloticus*) throughout West Africa. *Scientific reports, 9*(1), 1-12.

Lischer, H. E., Excoffier, L., & Heckel, G. (2014). Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of Microtus voles. *Molecular biology and evolution, 31*(4), 817-831.

Liu, S.-Y. V., Chang, F.-T., Borsa, P., Chen, W.-J., & Dai, C.-F. (2014). Phylogeography of the humbug damselfish, *Dascyllus aruanus* (Linnaeus, 1758): evidence of Indo-Pacific

vicariance and genetic differentiation of peripheral populations. *Biological Journal of the Linnean Society, 113*(4), 931-942.

Liu, S., Palti, Y., Gao, G., & Rexroad III, C. E. (2016). Development and validation of a SNP panel for parentage assignment in rainbow trout. *Aquaculture, 452*, 178-182.

Liu, T., Li, Q., Kong, L., & Yu, H. (2017). Comparison of microsatellites and SNPs for pedigree analysis in the Pacific oyster *Crassostrea gigas*. *Aquaculture international, 25*(4), 1507-1519.

Liu, Z., & Cordes, J. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture, 238*(1), 1-37.

Liu, Z. J. (2011). Genomic Variations and Marker Technologies for Genome-Based Selection. In Z. J. Liu (Ed.), *Next Generation Sequencing and Whole Genome Selection in Aquaculture* (Vol. 99, pp. 3-19): Blackwell Publishing Ltd.

Lorenzen, K., Beveridge, M., & Mangel, M. (2012). Cultured fish: integrative biology and management of domestication and interactions with wild fish. *Biological Reviews, 87*(3), 639-660. doi:10.1111/j.1469-185X.2011.00215.x

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology, 24*(5), 1031-1046. doi:10.1111/mec.13100

Lough, J. (2008). Shifting climate zones for Australia's tropical marine ecosystems. *Geophysical Research Letters, 35*(14).

Loughnan, S. R., Smith-Keune, C., Beheregaray, L. B., Robinson, N. A., & Jerry, D. R. (2019). Population genetic structure of barramundi (*Lates calcarifer*) across the natural distribution range in Australia informs fishery management and aquaculture practices. *Marine and Freshwater Research*, -. doi:https://doi.org/10.1071/MF18330

Ludt, W. B., & Rocha, L. A. (2015). Shifting seas: The impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. *Journal of Biogeography, 42*(1), 25-38.

Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources, 17*(1), 67-77. doi:10.1111/1755-0998.12592

Ma, K., Chan, T.-Y., & Chu, K. (2009). Phylogeny of penaeoid shrimps (Decapoda: Penaeoidea) inferred from nuclear protein-coding genes. *Molecular phylogenetics and evolution, 53*(1), 45-55.

Ma, K. Y., Chan, T. Y., & Chu, K. H. (2011). Refuting the six-genus classification of Penaeus sl (Dendrobranchiata, Penaeidae): a combined analysis of mitochondrial and nuclear genes. *Zoologica Scripta, 40*(5), 498-508.

Massault, C., Jones, D., Zenger, K., Strugnell, J., Barnard, R., & Jerry, D. (2021). A SNP parentage assignment panel for the silver lipped pearl oyster (*Pinctada maxima*). *Aquaculture Reports, 20*, 100687.

Mastrochirico-Filho, V. A., Del Pazo, F., Hata, M. E., Villanova, G. V., Foresti, F., Vera, M., . . . Hashimoto, D. T. (2019). Assessing genetic diversity for a pre-breeding program in *Piaractus mesopotamicus* by SNPs and SSRs. *Genes, 10*(9), 668.

Matolweni, L. O., Balkwill, K., & McLellan, T. (2000). Genetic diversity and gene flow in the morphologically variable, rare endemics *Begonia dregei* and *Begonia homonyma* (Begoniaceae). *American Journal of Botany, 87*(3), 431-439. doi:10.2307/2656639.

Marsden, G., Richardson, N., Mather, P., & Knibb, W. (2013). Reproductive behavioural differences between wild-caught and pond-reared *Penaeus monodon* prawn broodstock. *Aquaculture*, 402, 141-145.

May, S. A., McKinney, G. J., Hilborn, R., Hauser, L., & Naish, K. A. (2020). Power of a dual-use SNP panel for pedigree reconstruction and population assignment. *Ecology and Evolution, 10*(17), 9522-9531.

Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G. R., . . . Geffen, A. J. (2014). Outlier SNP markers reveal fine-scale genetic structuring across E uropean hake populations (*Merluccius merluccius)*. *Molecular Ecology, 23*(1), 118-135. doi:10.1111/mec.12568

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution, 37*(5), 1530-1534.

Montgomery, J. C., Jeffs, A., Simpson, S. D., Meekan, M., & Tindle, C. (2006). Sound as an orientation cue for the pelagic larvae of reef fishes and decapod crustaceans. *Advances in marine biology, 51*, 143-196.

Moore, J. S., Bourret, V., Dionne, M., Bradbury, I., O'Reilly, P., Kent, M., . . . Bernatchez, L. (2014). Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. *Molecular Ecology, 23*(23), 5680-5697. doi:10.1111/mec.12972

Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in ecology & evolution, 19*(4), 208-216. doi:10.1016/j.tree.2004.01.009

Morse, P., Kjeldsen, S. R., Meekan, M. G., Mccormick, M. I., Finn, J. K., Huffard, C. L., & Zenger, K. R. (2018). Genome-wide comparisons reveal a clinal species pattern within a holobenthic octopod—the Australian Southern blue-ringed octopus, *Hapalochlaena maculosa* (Cephalopoda: Octopodidae). *Ecology and Evolution, 8*(4), 2253-2267.

Moss, S. M., Moss, D. R., Arce, S. M., Lightner, D. V., & Lotz, J. M. (2012). The role of selective breeding and biosecurity in the prevention of disease in penaeid shrimp aquaculture. *Journal of Invertebrate Pathology, 110*(2), 247-250.

Motoh, H. (1985). *Biology and ecology of Penaeus monodon.* Paper presented at the Proceedings of the First International Conference on the Culture of Penaeid Prawns/Shrimps, 4-7 December 1984, Iloilo City, Philippines.

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., . . . Labhane, N. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment, 32*(2), 261-285.

Naylor, R. L., Hardy, R. W., Buschmann, A. H., Bush, S. R., Cao, L., Klinger, D. H., . . . Troell, M. (2021). A 20-year retrospective review of global aquaculture. *Nature, 591*(7851), 551-563.

Nazari, S., & Pourkazmi, M. (2021). Isolation and characterization of SNP markers of rainbow trout (*Onchorhynchus mykiss* Walbaum, 1792) from transcriptomic sequences. *Molecular biology reports, 48*(1), 989-995.

Nei, M. (1972). Genetic distance between populations. *The American Naturalist, 106*(949), 283-292.

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

Neuditschko, M., Khatkar, M. S., & Raadsma, H. W. (2012). NetView: a high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. *PloS one, 7*(10), e48375. https://github.com/esteinig/netview/blob/master/R/installNetView.R

Nguyen, S. V., Nguyen, T. L., Nguyen, V. H., Tran, V. N., Vu, N. T., & Nguyen, N. H. (2020). Genotype by environment interaction for survival and harvest body weight between recirculating tank system and pond culture in *Penaeus monodon*. *Aquaculture, 525*, 735278.

Nguyen, T. T., Hayes, B. J., & Ingram, B. A. (2014). Genetic parameters and response to selection in blue mussel (*Mytilus galloprovincialis*) using a SNP-based pedigree. *Aquaculture, 420*, 295-301.

Nielsen, E. E., Hemmer-Hansen, J., Larsen, P. F., & Bekkevold, D. (2009). Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology, 18*(15), 3128-3150.

Niu, S., Yang, L., Zuo, H., Zheng, J., Weng, S., He, J., & Xu, X. (2018). A chitinase from pacific white shrimp *Litopenaeus vannamei* involved in immune regulation. *Developmental & Comparative Immunology, 85*, 161-169. doi:10.1016/j.dci.2018.04.013

Norman-López, A., Sellars, M. J., Pascoe, S., Coman, G. J., Murphy, B., Moore, N., & Preston, N. (2016). Productivity benefits of selectively breeding Black Tiger shrimp (*Penaeus monodon*) in Australia. *Aquaculture Research, 47*(10), 3287-3296.

Nowland, S. J., Silva, C. N., Southgate, P. C., & Strugnell, J. M. (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. doi:10.1186/s12864-019-6052-z

Nyström, E. E., Birchenough, G. M., van der Post, S., Arike, L., Gruber, A. D., Hansson, G. C., & Johansson, M. E. (2018). Calcium-activated Chloride Channel Regulator 1 (CLCA1) Controls Mucus Expansion in Colon by Proteolytic Activity. *EBioMedicine*.

Ofori, B. Y., Beaumont, L. J., & Stow, A. J. (2017). Cunningham's skinks show low genetic connectivity and signatures of divergent selection across its distribution. *Ecology and Evolution, 7*(1), 48-57.

Ogden, R. (2011). Unlocking the potential of genomic technologies for wildlife forensics. *Molecular Ecology Resources, 11*, 109-116.

Ohlberger, J., Ward, E. J., Schindler, D. E., & Lewis, B. (2018). Demographic changes in Chinook salmon across the Northeast Pacific Ocean. *Fish and Fisheries, 19*(3), 533-546.

Oksanen, J., Blanchet, F., Kindt, R., Legendere, P., Minchin, P., & O'Hara, R. (2015). Vegan: Community Ecology Package. R Package Vegan, Version 2.2-1. In.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., . . . Wagner, H. (2018). vegan: Community Ecology Package. R package version 2.5-2. Retrieved from https://cran.r-project.org/web/packages/vegan/index.html

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., . . . Wagner, H. (2013). Package 'vegan'. *Community ecology package, version, 2*(9).

Oleksiak, M. F. (2019). Adaptation Without Boundaries: Population Genomics in Marine Systems. In O. P. Rajora (Ed.), *Population Genomics: Concepts, Approaches and Applications* (pp. 587-612). Cham: Springer International Publishing.

Otwoma, L. M., & 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).

Ovenden, J. R., Berry, O., Welch, D. J., Buckworth, R. C., & Dichmont, C. M. (2015). Ocean's eleven: a critical evaluation of the role of population, evolutionary and molecular genetics in the management of wild fisheries. *Fish and Fisheries, 16*(1), 125-159.

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.

Park, L. K., & Moran, P. (1995). Developments in molecular genetic techniques in fisheries. In *Molecular genetics in fisheries* (pp. 1-28): Springer.

Parmley, J. L., Chamary, J., & Hurst, L. D. (2005). Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. *Molecular biology and evolution, 23*(2), 301-309.

Pascual, M., Rives, B., Schunter, C., & Macpherson, E. (2017). Impact of life history traits on gene flow: A multispecies systematic review across oceanographic barriers in the Mediterranean Sea. *PloS one, 12*(5), e0176419.

Pavlidis, P., Jensen, J. D., Stephan, W., & Stamatakis, A. (2012). A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. *Molecular biology and evolution, 29*(10), 3237-3248.

Pearce, A., & Feng, M. (2007). Observations of warming on the Western Australian continental shelf. *Marine and Freshwater Research, 58*(10), 914-920.

Pembleton, L. W., Cogan, N. O., & Forster, J. W. (2013). StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources, 13*(5), 946-952. doi:10.1111/1755-0998.12129

Perez-Enriquez, R., & Max-Aguilar, A. (2016). Pedigree traceability in whiteleg shrimp (*Litopenaeus vannamei*) using genetic markers: A comparison between microsatellites and SNPs. *Ciencias marinas, 42*(4), 227-235.

Phillips, B. F., Melville-Smith, R., Kay, M. C., & Vega-Velázquez, A. (2013). Panulirus species. *Lobsters: Biology, Management, Aquaculture and Fisheries, Second Edition*, 289-325.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics, 8*(11), e1002967. doi:10.1371/journal.pgen.1002967

Pirog, A., Latreille, A. C., Madelaine, C., Gélin, P., Frouin, P., & Magalon, H. (2019). High clonal propagation and low population connectivity in the holothurian *Stichopus chloronotus* from the Indo-Pacific. *Marine Biology, 166*(5), 63. doi:10.1007/s00227-019-3512-6

Pitcher, C., Turnbull, C., Atfield, J., Griffin, D., Dennis, D., & Skewes, T. (2005). Biology, larval transport modelling and commercial logbook data analysis to support management of the NE Queensland rock lobster *Panulirus ornatus* fishery. FRDC Project, 2002/008. CSIRO Marine and Atmospheric Research, Cleveland, Qld., p. 144.

Plön, S., Thakur, V., Parr, L., & Lavery, S. D. (2019). Phylogeography of the dugong (*Dugong dugon*) based on historical samples identifies vulnerable Indian Ocean populations. *PloS one, 14*(9), e0219350.

Pootakham, W., Uengwetwanit, T., Sonthirod, C., Sittikankaew, K., & Karoonuthaisiri, N. (2020). A novel full-Length transcriptome resource for black tiger shrimp (*Penaeus monodon*) developed using isoform sequencing (Iso-Seq). *Frontiers in Marine Science, 7*, 172.

Premachandra, H., Nguyen, N. H., & Knibb, W. (2019). Effectiveness of SNPs for parentage and sibship assessment in polygamous yellowtail kingfish *Seriola lalandi*. *Aquaculture, 499*, 24-31.

Ptacek, M. B., Sarver, S. K., Childress, M. J., & Herrnkind, W. F. (2001). Molecular phylogeny of the spiny lobster genus *Panulirus* (Decapoda: Palinuridae). *Marine and Freshwater Research, 52*(8), 1037-1047.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*(3), 559-575. doi:10.1086/519795

Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution, 4*(22), 4399-4428.

Quan, J., Zhuang, Z., Deng, J., Dai, J., & Zhang, Y.-p. (2004). Phylogenetic relationships of 12 Penaeoidea shrimp species deduced from mitochondrial DNA sequences. *Biochemical Genetics, 42*(9-10), 331-345.

Queiroga, H., & Blanton, J. (2004). Interactions between behaviour and physical forcing in the control of horizontal transport of decapod crustacean larvae. *Advances in marine biology, 47*, 107-214. doi:10.1016/s0065-2881(04)47002-3

Rambaut, A. (2018). FigTree v1.4.4. Retrieved from http://tree.bio.ed.ac.uk/software/figtree/

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic biology, 67*(5), 901. doi:10.1093/sysbio/syy032

Randall, J. E. (1998). Zoogeography of shore fishes of the Indo-Pacific region. *Zoological Studies, 37*(4), 227-268. Retrieved from http://zoolstud.sinica.edu.tw/Journals/37.4/227.pdf

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature, 461*(7263), 489-494.

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology, 24*(17), 4348-4370.

Reverter, A., Hudson, N. J., McWilliam, S., Alexandre, P. A., Li, Y., Barlow, R., . . . Dominik, S. (2020). A low-density SNP genotyping panel for the accurate prediction of cattle breeds. *Journal of Animal Science, 98*(11), skaa337.

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., & Houston, R. (2018). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in Aquaculture, 10*(3), 670-682.

Rocha, L., Craig, M., & Bowen, B. (2007). Phylogeography and the conservation of coral reef fishes. *Coral Reefs, 26*(3), 501-512. doi:10.1007/s00338-007-0280-4

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., . . . Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology, 61*(3), 539-542.

Rothlisberg, P. C. (1982). Vertical migration and its effect on dispersal of penaeid shrimp larvae in the Gulf of Carpentaria, Australia.

Saint-Pé, K., Leitwein, M., Tissot, L., Poulet, N., Guinand, B., Berrebi, P., . . . Blanchet, S. (2019). Development of a large SNPs resource and a low-density SNP array for brown trout (*Salmo trutta*) population genetics. *BMC genomics, 20*(1), 1-13.

Sandoval-Castillo, J., Robinson, N. A., Hart, A. M., Strain, L. W., & Beheregaray, L. B. (2018). Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevigata*), along a longitudinal environmental gradient. *Molecular Ecology, 27*(7), 1603-1620. doi:10.1111/mec.14526

Sandoval, L. A., Leal-Florez, J., Taborda, A., & Vásquez, J. G. (2014). Spatial distribution and abundance of the giant tiger prawn, *Penaeus monodon* (Fabricius, 1798), in the Gulf of Urabá (Caribbean), Colombia, South America. *BioInvasions Record, 3*(3).

Sanford, E., & Kelly, M. W. (2011). Local adaptation in marine invertebrates. *Annual review of marine science, 3*, 509-535.

Sansaloni, C. P., Petroli, C. D., Carling, J., Hudson, C. J., Steane, D. A., Myburg, A. A., . . . Kilian, A. (2010). A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in Eucalyptus. *Plant Methods, 6*(1), 16.

Schott, F. A., Xie, S. P., & McCreary Jr, J. P. (2009). Indian Ocean circulation and climate variability. *Reviews of Geophysics, 47*(1). doi:10.1029/2007rg000245

Segovia, N. I., Gallardo-Escárate, C., Poulin, E., & Haye, P. A. (2017). Lineage divergence, local adaptation across a biogeographic break, and artificial transport, shape the genetic structure in the ascidian *Pyura chilensis*. *Scientific reports, 7*, 44559. doi:10.1038/srep44559

Selkoe, K., & Toonen, R. J. (2011). Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series, 436*, 291-305.

Sellars, M. J., Dierens, L., McWilliam, S., Little, B., Murphy, B., Coman, G. J., . . . Henshall, J. (2014). Comparison of microsatellite and SNP DNA markers for pedigree assignment in Black Tiger shrimp, *Penaeus monodon*. *Aquaculture Research, 45*(3), 417-426.

Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution, 68*(1), 1-15.

Sheaves, M., Johnston, R., Connolly, R. M., & Baker, R. (2012). Importance of estuarine mangroves to juvenile banana prawns. *Estuarine, Coastal and Shelf Science, 114*, 208-219.

Silva, C. N., & Gardner, J. P. (2015). Identifying environmental factors associated with the genetic structure of the New Zealand scallop: linking seascape genetics and ecophysiological tolerance. *ICES Journal of Marine Science, 73*(7), 1925-1934.

Silva, I. C., Mesquita, N., & Paula, J. (2010). Genetic and morphological differentiation of the mangrove crab *Perisesarma guttatum* (Brachyura: Sesarmidae) along an East African latitudinal gradient. *Biological Journal of the Linnean Society, 99*(1), 28-46.

Simmonds, S. E., Chou, V., Cheng, S. H., Rachmawati, R., Calumpong, H. P., Mahardika, G. N., & Barber, P. H. (2018). Evidence of host-associated divergence from coral-eating

snails (genus *Coralliophila*) in the Coral Triangle. *Coral Reefs, 37*(2), 355-371. doi:10.1007/s00338-018-1661-6

Sk, N. U., 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. *International Journal of Aquatic Biology, 3*(6), 372. doi:10.22034/ijab.v3i6.2

Sloss, C. R., Nothdurft, L., Hua, Q., O'Connor, S. G., Moss, P. T., Rosendahl, D., . . . Sternes, A. (2018). Holocene sea-level change and coastal landscape evolution in the southern Gulf of Carpentaria, Australia. *The Holocene, 28*(9), 1411-1430.

Song, H., & Hu, H. (2021). Strategies to improve the accuracy and reduce costs of genomic prediction in aquaculture species. *Evolutionary Applications, n/a*(n/a). doi:https://doi.org/10.1111/eva.13262

Steinig, E. J., Neuditschko, M., Khatkar, M. S., Raadsma, H. W., & Zenger, K. R. (2016). netview p: a network visualization tool to unravel complex population structure using genome-wide SNPs. *Molecular Ecology Resources, 16*(1), 216-227.

Stoffel, M. A., Esser, M., Kardos, M., Humble, E., Nichols, H., David, P., & Hoffman, J. I. (2016). inbreedR: an R package for the analysis of inbreeding based on genetic markers. *Methods in Ecology and Evolution, 7*(11), 1331-1339. doi:10.1111/2041-210x.12588

Sukhavachana, S., Tongyoo, P., Luengnaruemitchai, A., & Poompuang, S. (2021). Optimizing genomic prediction using low-density marker panels for streptococcosis resistance in red tilapia (*Oreochromis spp*.). *Animal genetics*.

Sweet, D. L. (2020). Package 'nonpar': A Collection of Nonparametric Hypothesis Tests. Retrieved from https://CRAN.R-project.org/package=nonpar

Symonds, J. E., Clarke, S. M., King, N., Walker, S. P., Blanchard, B., Sutherland, D., . . . Buxton, P. (2019). Developing successful breeding programs for New Zealand aquaculture: a perspective on progress and future genomic opportunities. *Frontiers in genetics, 10*, 27.

Tanya, C., & Kumar, R. (2010). Molecular markers and their applications in fisheries and aquaculture. *Advances in Bioscience and Biotechnology, 2010*.

Teske, P. R., Papadopoulos, I., Newman, B. K., Dworschak, P. C., McQuaid, C. D., & Barker, N. P. (2008). Oceanic dispersal barriers, adaptation and larval retention: an interdisciplinary assessment of potential factors maintaining a phylogeographic break between sister lineages of an African prawn. *BMC Evolutionary Biology, 8*(1), 1-14.

Thongda, W., Zhao, H., Zhang, D., Jescovitch, L. N., Liu, M., Guo, X., . . . Peatman, E. (2018). Development of SNP panels as a new tool to assess the genetic diversity, population structure, and parentage analysis of the eastern oyster (*Crassostrea virginica*). *Marine Biotechnology, 20*(3), 385-395.

Thornber, K., Verner-Jeffreys, D., Hinchliffe, S., Rahman, M. M., Bass, D., & Tyler, C. R. (2020). Evaluating antimicrobial resistance in the global shrimp industry. *Reviews in Aquaculture, 12*(2), 966-986.

Tigano, A., Shultz, A. J., Edwards, S. V., Robertson, G. J., & Friesen, V. L. (2017). Outlier analyses to test for local adaptation to breeding grounds in a migratory arctic seabird. *Ecology and Evolution, 7*(7), 2370-2381.

Treml, E. A., Halpin, P. N., Urban, D. L., & Pratson, L. F. (2008). Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. *Landscape Ecology, 23*(1), 19-36.

Tsairidou, S., Hamilton, A., Robledo, D., Bron, J. E., & Houston, R. D. (2020). Optimizing low-cost genotyping and imputation strategies for genomic selection in Atlantic salmon. *G3: Genes, Genomes, Genetics, 10*(2), 581-590.

Tsoi, K., Chan, T., & Chu, K. (2007). Molecular population structure of the kuruma shrimp *Penaeus japonicus* species complex in western Pacific. *Marine Biology, 150*(6), 1345-1364. doi:10.1007/s00227-006-0426-x

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., & De Clerck, O. (2012). Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography, 21*(2), 272-281.

Uengwetwanit, T., Ponza, P., Sangsrakru, D., Wichadakul, D., Ingsriswang, S., Leelatanawit, R., . . . Karoonuthaisiri, N. (2018). Transcriptome-based discovery of pathways and genes related to reproduction of the black tiger shrimp (*Penaeus monodon*). *Marine genomics, 37*, 69-73.

Uengwetwanit, T., Pootakham, W., Nookaew, I., Sonthirod, C., Angthong, P., Sittikankaew, K., . . . Jenjaroenpun, P. (2021). A chromosome-level assembly of the black tiger shrimp (*Penaeus monodon*) genome facilitates the identification of growth-associated genes. *Molecular Ecology Resources*.

Valenzuela-Quiñonez, F. (2016). How fisheries management can benefit from genomics? *Briefings in functional genomics*, elw006.

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. *Marine Biotechnology, 6*(5), 475-484. doi:10.1007/s10126-004-3138-6

Van Cise, A. M., Baird, R. W., Baker, C. S., Cerchio, S., Claridge, D., Fielding, R., . . . Mignucci-Giannoni, A. A. (2019). Oceanographic barriers, divergence, and admixture: Phylogeography and taxonomy of two putative subspecies of short-finned pilot whale. *Molecular Ecology, 28*(11), 2886-2902.

Van Quyen, D., Gan, H. M., Lee, Y. P., Nguyen, D. D., Nguyen, T. H., Tran, X. T., . . . Austin, C. M. (2020). Improved genomic resources for the black tiger prawn (*Penaeus monodon*). *Marine genomics, 52*, 100751.

Van Wyngaarden, M., Snelgrove, P. V., DiBacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Zhan, L., . . . Bradbury, I. R. (2018). Oceanographic variation influences spatial genomic structure in the sea scallop, *Placopecten magellanicus*. *Ecology and Evolution, 8*(5), 2824-2841.

Vandeputte, M., & Haffray, P. (2014). Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Frontiers in genetics, 5*, 432.

Vela-Avitúa, S., Montaldo, H. H., Márquez Valdelamar, L., Campos Montes, G. R., & Castillo Juárez, H. (2013). Decline of genetic variability in a captive population of Pacific white shrimp Penaeus (*Litopenaeus vannamei*) using microsatellite and pedigree information. *Electronic Journal of Biotechnology, 16*(4), 9-9.

Vogler, C., Benzie, J., Barber, P. H., Erdmann, M. V., Sheppard, C., Tenggardjaja, K., . . . Wörheide, G. (2012). Phylogeography of the crown-of-thorns starfish in the Indian Ocean. *PloS one, 7*(8), e43499. doi:10.1371/journal.pone.0043499

Voloch, C. M., Freire, P. R., & Russo, C. A. (2005). Molecular phylogeny of penaeid shrimps inferred from two mitochondrial markers. *Genetics and Molecular Research, 4*(4), 668-674.

Voloch, C. M., Freire, P. R., & Russo, C. A. (2009). Molecular phylogeny and divergence time estimates of Penaeid Shrimp Lineages (Decapoda: Penaeidae). *Zootaxa, 2107*(1), 41-52.

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

Vuthiphandchai, V., Pengpun, B., & Nimrat, S. (2005). Effects of cryoprotectant toxicity and temperature sensitivity on the embryos of black tiger shrimp (*Penaeus monodon*). *Aquaculture, 246*(1-4), 275-284.

Wade, N. M., Anderson, M., Sellars, M. J., Tume, R. K., Preston, N. P., & Glencross, B. D. (2012). Mechanisms of colour adaptation in the prawn *Penaeus monodon*. *Journal of Experimental Biology, 215*(2), 343-350.

Wade, N. M., Budd, A., Irvin, S., & Glencross, B. D. (2015). The combined effects of diet, environment and genetics on pigmentation in the giant tiger prawn, *Penaeus monodon*. *Aquaculture, 449*, 78-86.

Wade, N. M., Paulo, C., Goodall, J., Fischer, M., Poole, S., & Glencross, B. D. (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, B. J., Arlyza, I. S., & Karl, S. A. (2018). Population genetics of the collector urchin, *Tripneustes gratilla*, in the Indonesian archipelago. *Marine Ecology, 39*(6), e12530. doi:10.1111/maec.12530

Wakida-Kusunoki, A. T., De Anda-Fuentes, D., & López-Téllez, N. A. (2016). Presence of giant tiger shrimp *Penaeus monodon* (Fabricius, 1798) in eastern Peninsula of Yucatan coast, Mexico. *Latin American Journal of Aquatic Research, 44*(1), 155-158.

Wang, Y., Lv, H., Xiang, X., Yang, A., Feng, Q., Dai, P., . . . Zhang, X. (2021). Construction of a SNP Fingerprinting Database and Population Genetic Analysis of Cigar Tobacco Germplasm Resources in China. *Frontiers in plant science, 12*.

Wang, Z. Y., Tsoi, K. H., & Chu, K. H. (2004). Applications of AFLP technology in genetic and phylogenetic analysis of penaeid shrimp. *Biochemical Systematics and Ecology, 32*(4), 399-407.

Waples, R. S. (1998). Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity, 89*(5), 438-450.

Waples, R. S. (2018). Null Alleles and F IS× F ST Correlations. *Journal of Heredity, 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.

Waqairatu, S. S., Dierens, L., Cowley, J. A., Dixon, T. J., Johnson, K. N., Barnes, A. C., & Li, Y. (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. *Ecology and Evolution, 2*(8), 2057-2071. doi:10.1002/ece3.316

Ward, R. D., Ovenden, J. R., Meadows, J. R., Grewe, P. M., & Lehnert, S. A. (2006). Population genetic structure of the brown tiger prawn, *Penaeus esculentus*, in tropical northern Australia. *Marine Biology, 148*(3), 599-607.

Wei, X., Liu, X., Yang, J., Fang, J., Qiao, H., Zhang, Y., & Yang, J. (2012). Two C-type lectins from shrimp *Litopenaeus vannamei* that might be involved in immune response against bacteria and virus. *Fish & shellfish immunology, 32*(1), 132-140.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution, 38*(6), 1358-1370. doi:10.2307/2408641

Wenne, R. (2018). Single nucleotide polymorphism markers with applications in aquaculture and assessment of its impact on natural populations. *Aquatic Living Resources, 31*, 2.

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of F_{ST}. *The American Naturalist, 186*(S1), S24-S36.

Wickham, H., Francois, R., Henry, L., & Müller, K. (2021). dplyr: A grammar of data manipulation (Version 1.0.7). https://cran.r-project.org/. Retrieved from https://dplyr.tidyverse.org, https://github.com/tidyverse/dplyr.

Williams, S., & Benzie, J. (1998). Evidence of a biogeographic break between populations of a high dispersal starfish: congruent regions within the Indo-West Pacific defined by color morphs, mtDNA, and allozyme data. *Evolution, 52*(1), 87-99.

Williams, S. E., & Hoffman, E. A. (2009). Minimizing genetic adaptation in captive breeding programs: a review. *Biological Conservation, 142*(11), 2388-2400. doi:10.1016/j.biocon.2009.05.034.

Wong, N., Tooman, L., Sewell, M., & Lavery, S. (2016). The population genetics and origin of invasion of the invasive Asian paddle crab, *Charybdis japonica* (A. Milne-Edwards, 1861)(Brachyura: Portunidae) in north-eastern New Zealand. *Marine Biology*, *163*(6), 1.

Woodings, L. N., Murphy, N. P., Doyle, S. R., Hall, N. E., Robinson, A. J., Liggins, G. W., . . . Strugnell, J. M. (2018). Outlier SNPs detect weak regional structure against a background of genetic homogeneity in the Eastern Rock Lobster, *Sagmariasus verreauxi*. *Marine Biology, 165*(12), 185. doi:10.1007/s00227-018-3443-7

Wright, S. (1949). The genetical structure of populations. *Annals of eugenics, 15*(1), 323-354. doi:10.1136/bmj.2.4669.36.

Xu, J., Feng, J., Peng, W., Liu, X., Feng, J., & Xu, P. (2017). Development and evaluation of a high-throughput single nucleotide polymorphism multiplex assay for assigning pedigrees in common carp. *Aquaculture Research, 48*(4), 1866-1876.

Xuereb, A., Kimber, C. M., Curtis, J. M., Bernatchez, L., & Fortin, M. J. (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. *Molecular Ecology*.

You, E. M., Chiu, T. S., Liu, K. F., Tassanakajon, A., Klinbunga, S., Triwitayakorn, K., . . . Yu, H. T. (2008). Microsatellite and mitochondrial haplotype diversity reveals population differentiation in the tiger shrimp (*Penaeus monodon*) in the Indo-Pacific region. *Animal genetics, 39*(3), 267-277. doi:10.1111/j.1365-2052.2008.01724.x.

Yue, G., & Wang, L. (2017). Current status of genome sequencing and its applications in aquaculture. *Aquaculture, 468*, 337-347.

Zarraonaindia, I., Iriondo, M., Albaina, A., Pardo, M. A., Manzano, C., Grant, W. S., . . . Estonba, A. (2012). Multiple SNP markers reveal fine-scale population and deep phylogeographic structure in European anchovy (*Engraulis encrasicolus* L.).

Zenger, K. R., Khatkar, M. S., Jones, D. B., Khalilisamani, N., Jerry, D. R., & Raadsma, H. W. (2019). Genomic selection in aquaculture: application, limitations and opportunities with special reference to marine shrimp and pearl oysters. *Frontiers in genetics, 9*, 693.

Zhang, S., Chen, Y., An, H., Liu, H., Li, J., Pang, C., . . . Zhan, Y. (2014). A novel biophysical model on calcium and voltage dual dependent gating of calcium-activated chloride channel. *Journal of theoretical biology, 355*, 229-235.

Zhao, H., Li, C., Hargrove, J. S., Bowen, B. R., Thongda, W., Zhang, D., . . . Peatman, E. (2018). SNP marker panels for parentage assignment and traceability in the Florida bass (*Micropterus floridanus*). *Aquaculture, 485*, 30-38.

Zhou, F., Zhou, K., Huang, J., Yang, Q., Jiang, S., Qiu, L., . . . Jiang, S. (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 immunology, 72*, 117-123. doi:10.1016/j.fsi.2017.10.051.

Zink, I. C., Jackson, T. L., & Browder, J. A. (2018). A note on the occurrence of non-native tiger prawn (*Penaeus monodon* Fabricius, 1798) in Biscayne Bay, FL, USA and review of South Florida sighting and species identification. *BioInvasions Record, 7*(3).

Zitari-Chatti, R., Chatti, N., Fulgione, D., Caiazza, I., Aprea, G., Elouaer, A., . . . Capriglione, T. (2009). Mitochondrial DNA variation in the caramote prawn *Penaeus (Melicertus) kerathurus* across a transition zone in the Mediterranean Sea. *Genetica, 136*(3), 439-447.

Appendices

Chapter 2.

Appendix 2.1 Marine environmental variables (23) based on latitude and longitude of seven sampled sites.

State	Northern Territory			North Queensland			Western Australia
Sampling Region	Gulf of Carpentaria	Joseph Bonaparte Gulf	Tiwi Island	Etty Bay	Bramston Beach	Townsville	Nickol Bay
Code	GC	JBG	TIW	EB	BB	TSV	NKB
Time period and grain	2000 - 2014	2000 - 2014	2000 - 2014	2000 - 2014	2000 - 2014	2000 - 2014	2000 - 2014
Latitude	-14.997655	-14.681809	-11.86683	-17.561792	-17.353008	-19.082107	-20.69701
Longitude	138.780787	128.371358	131.83147	146.10895	146.05272	146.817537	116.8632
Surface Salinity Min (Env_1)	33.05	30.01	29.96	34.05	34.05	33.73	32.94
Surface Temperature Max (Env_2)	30.94	31.29	30.92	29.54	29.54	29.32	31.48
Surface Temperature Min (Env_3)	23.44	21.82	24.02	22.54	22.54	21.92	21.25
Surface Current velocity Mean (Env_4)	0.14	0.03	0.05	0.05	0.05	0.16	0.31
Surface Phytoplankton Mean (Env_5)	1.28	1.45	1.52	1.44	1.44	1.67	1.39
Benthic Temperature Mean (Env_6)	0.14	0.02	0.02	0.05	0.05	0.04	0.05
Benthic Current velocity Mean (Env_7)	0.03	0.03	0.03	0.03	0.03	0.02	0.05
Surface Salinity Max	35.89	35.55	35.54	35.50	35.50	35.80	35.62
Surface Salinity Mean	34.66	33.79	33.13	35.05	35.05	35.19	34.55
Surface Dissolved molecular oxygen Max	227.83	220.27	217.21	227.29	227.29	236.35	227.83
Surface Dissolved molecular oxygen Min	187.51	189.57	190.44	193.65	193.65	196.70	193.00
Surface Current velocity Max	0.06	0.18	0.22	0.40	0.40	0.22	0.15
Surface Current velocity Min	0.04	0.16	0.17	0.37	0.37	0.35	0.04
Surface Phytoplankton Min	1.01	1.10	1.24	1.44	1.44	1.20	1.11
Benthic Salinity Max	35.78	36.09	34.76	35.36	35.36	35.44	35.27
Benthic Salinity Mean	34.41	35.34	34.39	35.02	35.02	35.08	34.92
Benthic Salinity Min	33.03	34.54	33.79	34.48	34.48	34.50	34.42
Benthic Temperature Max	0.51	0.11	0.13	0.33	0.33	0.22	0.13
Benthic Dissolved molecular oxygen Max	211.28	219.13	218.71	219.39	219.39	233.37	224.18
Benthic Dissolved molecular oxygen Mean	201.67	201.81	201.89	207.09	207.09	210.38	207.47
Benthic Current velocity Max	0.21	0.11	0.14	0.32	0.32	0.23	0.13
Benthic Phytoplankton Mean	0.98	1.12	1.13	1.21	1.21	1.28	1.31
Benthic Phytoplankton Min	0.72	0.90	0.91	0.94	0.94	0.97	1.04

Appendix 2.2 Characterization of high-quality BLAST matches obtained in comparison with *P. monod*on sequencing SNP against *P. monod*on transcriptome

SNP (#)	Contig matched	% Pairwise Identity	Bit-Score	Contig Description	E Value	Hit end	Hit start	Query coverage (%)
PM1958_A	contig11442.1	100	128.5	calcium-activated chloride channel regulator 2-like	2.01E-29	2,039	1,971	100
	contig9994.1	100	128.5	Calcium-activated chloride channel regulator	2.01E-29	688	756	100
	contig3578.1	100	128.5	epithelial chloride channel -like	2.01E-29	3,761	3,693	100
	contig3237.1	100	128.5	epithelial chloride channel -like	2.01E-29	3,970	3,902	100
PM1958_B	contig11442.1	98.6	123.0	calcium-activated chloride channel regulator 2-like	9.33E-28	2,039	1,971	100
	contig9994.1	98.6	123.0	Calcium-activated chloride channel regulator	9.33E-28	688	756	100
	contig3578.1	98.6	123.0	epithelial chloride channel -like	9.33E-28	3,761	3,693	100
	contig3237.1	98.6	123.0	epithelial chloride channel -like	9.33E-28	3,970	3,902	100
PM6488_A	contig196702.1	97.1	58.4	---NA---	2.67E-08	283	316	49.28
	contig184051.1	97	56.5	---NA---	9.60E-08	160	192	47.83
	contig169503.1	100	62.1	---NA---	2.06E-09	293	261	47.83
	contig160248.1	100	62.1	---NA---	2.06E-09	112	144	47.83
	contig146699.1	97.1	58.4	---NA---	2.67E-08	259	292	49.28
	contig144611.1	97	56.5	---NA---	9.60E-08	402	370	47.83
	contig141018.1	97.1	58.4	---NA---	2.67E-08	118	85	49.28
	contig121911.1	97	56.5	---NA---	9.60E-08	160	192	47.83
	contig110726.1	97.1	58.4	---NA---	2.67E-08	7	40	49.28
	contig102320.1	97	56.5	---NA---	9.60E-08	203	235	47.83
	contig100878.1	97	56.5	---NA---	9.60E-08	451	483	47.83
	contig89079.1	100	60.2	---NA---	7.42E-09	510	479	46.38
	contig88989.1	97.1	58.4	---NA---	2.67E-08	152	119	49.28
	contig83954.1	97	56.5	---NA---	9.60E-08	479	511	47.83
	contig63011.1	97	56.5	---NA---	9.60E-08	535	503	47.83

Appendix 2.2 (cont.)

SNP (#)	Contig matched	% Pairwise Identity	Bit-Score	Contig Description	E Value	Hit end	Hit start	Query coverage (%)
PM6488_A	contig54766.1	100	60.2	---NA---	7.42E-09	813	782	46.38
	contig47238.2	97	56.5	---NA---	9.60E-08	419	451	47.83
	contig47238.1	97	56.5	---NA---	9.60E-08	419	451	47.83
	contig45025.1	100	54.7	---NA---	3.45E-07	484	456	42.03
	contig43670.1	100	60.2	---NA---	7.42E-09	495	464	46.38
	contig37913.1	100	60.2	---NA---	7.42E-09	183	152	46.38
	contig33852.1	100	60.2	---NA---	7.42E-09	783	752	46.38
	contig32241.1	100	54.7	---NA---	3.45E-07	484	456	42.03
	contig21545.1	97	56.5	---NA---	9.60E-08	1,731	1,763	47.83
	contig3381.1	97	56.5	AF445312_1unknown	9.60E-08	698	730	47.83
	contig1214.2	97	56.5	RNA-directed DNA polymerase from mobile element jockey-like	9.60E-08	159	191	47.83
	contig1214.1	97	56.5	RNA-directed DNA polymerase from mobile element jockey-like	9.60E-08	159	191	47.83
PM6488_B	contig221983.1	96.9	54.7	---NA---	3.45E-07	15	46	46.38
	contig215803.1	97	56.5	---NA---	9.60E-08	100	68	47.83
	contig211217.1	100	63.9	---NA---	5.74E-10	208	175	49.28
	contig209362.1	100	60.2	---NA---	7.42E-09	228	197	46.38
	contig196252.1	97	56.5	---NA---	9.60E-08	94	62	47.83
	contig191456.1	100	60.2	---NA---	7.42E-09	158	189	46.38
	contig190456.1	97.1	58.4	---NA---	2.67E-08	215	248	49.28
	contig183806.1	97	56.5	---NA---	9.60E-08	258	290	47.83
	contig182250.1	100	52.8	---NA---	1.24E-06	183	156	40.58
	contig174356.1	97	56.5	---NA---	9.60E-08	327	295	47.83
	contig173878.1	97	56.5	---NA---	9.60E-08	247	215	47.83
	contig170096.1	100	58.4	---NA---	2.67E-08	77	47	44.93

Appendix 2.2 (cont.)

SNP (#)	Contig matched	% Pairwise Identity	Bit-Score	Contig Description	E Value	Hit end	Hit start	Query coverage (%)
PM6488_B	contig168958.1	100	63.9	---NA---	5.74E-10	146	179	49.28
	contig165232.1	97	56.5	---NA---	9.60E-08	311	279	47.83
	contig160882.1	97.1	58.4	---NA---	2.67E-08	251	218	49.28
	contig160442.1	96.8	52.8	---NA---	1.24E-06	84	114	44.93
	contig155157.1	96.9	54.7	---NA---	3.45E-07	183	214	46.38
	contig153702.1	100	54.7	---NA---	3.45E-07	154	182	42.03
	contig151497.1	100	60.2	---NA---	7.42E-09	144	113	46.38
	contig150717.1	100	54.7	---NA---	3.45E-07	161	133	42.03
	contig146738.1	100	60.2	---NA---	7.42E-09	43	12	46.38
	contig145055.1	96.8	52.8	---NA---	1.24E-06	233	203	44.93
	contig144894.1	100	54.7	---NA---	3.45E-07	291	263	42.03
	contig144413.1	100	60.2	---NA---	7.42E-09	165	196	46.38
	contig142507.1	96.9	54.7	---NA---	3.45E-07	94	125	46.38
	contig141960.1	97	56.5	---NA---	9.60E-08	333	365	47.83
	contig141460.1	100	60.2	---NA---	7.42E-09	252	283	46.38
	contig137288.1	100	62.1	---NA---	2.06E-09	272	304	47.83
	contig136468.1	97	56.5	---NA---	9.60E-08	395	427	47.83
	contig134383.1	97	56.5	---NA---	9.60E-08	306	274	47.83
	contig133091.1	96.8	52.8	---NA---	1.24E-06	163	193	44.93
	contig131883.1	100	56.5	---NA---	9.60E-08	278	249	43.48
	contig131661.1	97	56.5	---NA---	9.60E-08	370	338	47.83
	contig130490.1	100	58.4	---NA---	2.67E-08	356	386	44.93
	contig129639.1	100	58.4	---NA---	2.67E-08	373	343	44.93
	contig127989.1	100	54.7	---NA---	3.45E-07	218	190	42.03

Appendix 2.2 (cont.)

SNP (#)	Contig matched	% Pairwise Identity	Bit-Score	Contig Description	E Value	Hit end	Hit start	Query coverage (%)
	contig126961.1	100	63.9	---NA---	5.74E-10	153	186	49.28
	contig123773.1	100	60.2	---NA---	7.42E-09	222	253	46.38
	contig120030.1	97	56.5	---NA---	9.60E-08	335	303	47.83
	contig119207.1	100	60.2	---NA---	7.42E-09	285	316	46.38
	contig114246.1	100	62.1	---NA---	2.06E-09	233	265	47.83
	contig110605.1	100	60.2	---NA---	7.42E-09	152	121	46.38
	contig110569.1	100	62.1	---NA---	2.06E-09	368	400	47.83
	contig108609.1	97	56.5	---NA---	9.60E-08	459	427	47.83
	contig108525.2	100	60.2	---NA---	7.42E-09	266	297	46.38
	contig108525.1	100	60.2	---NA---	7.42E-09	267	298	46.38
	contig108273.1	97	56.5	---NA---	9.60E-08	60	28	47.83
	contig106547.1	97	56.5	---NA---	9.60E-08	285	317	47.83
PM6488_B	contig106521.1	100	54.7	---NA---	3.45E-07	239	211	42.03
	contig102310.1	100	60.2	---NA---	7.42E-09	526	495	46.38
	contig102032.1	96.9	54.7	---NA---	3.45E-07	159	128	46.38
	contig98550.1	100	56.5	---NA---	9.60E-08	324	295	43.48
	contig95902.1	97	56.5	---NA---	9.60E-08	117	85	47.83
	contig91117.1	100	60.2	---NA---	7.42E-09	307	276	46.38
	contig90378.1	100	54.7	---NA---	3.45E-07	46	74	42.03
	contig88653.1	100	60.2	---NA---	7.42E-09	489	520	46.38
	contig88268.1	100	60.2	---NA---	7.42E-09	64	33	46.38
	contig85300.1	96.9	54.7	---NA---	3.45E-07	331	300	46.38
	contig85217.1	100	58.4	---NA---	2.67E-08	378	408	44.93
	contig84904.1	100	60.2	---NA---	7.42E-09	335	304	46.38

Appendix 2.2 (cont.)

SNP (#)	Contig matched	% Pairwise Identity	Bit-Score	Contig Description	E Value	Hit end	Hit start	Query coverage (%)
PM6488_B	contig84045.1	100	60.2	---NA---	7.42E-09	275	306	46.38
	contig83472.1	97	56.5	---NA---	9.60E-08	211	179	47.83
	contig79424.1	96.9	54.7	---NA---	3.45E-07	392	423	46.38
	contig79102.1	100	60.2	---NA---	7.42E-09	282	313	46.38
	contig75907.1	100	56.5	---NA---	9.60E-08	657	686	43.48
	contig75105.1	100	52.8	---NA---	1.24E-06	698	725	40.58
	contig75049.1	97	56.5	---NA---	9.60E-08	131	99	47.83
	contig74027.1	97.1	58.4	---NA---	2.67E-08	433	466	49.28
	contig70385.1	96.9	54.7	---NA---	3.45E-07	537	506	46.38
	contig69377.1	97	56.5	---NA---	9.60E-08	693	725	47.83
	contig69263.1	100	54.7	---NA---	3.45E-07	491	463	42.03
	contig67491.1	96.9	54.7	---NA---	3.45E-07	51	82	46.38
	contig65678.1	100	54.7	---NA---	3.45E-07	665	637	42.03
	contig63233.1	100	62.1	---NA---	2.06E-09	744	712	47.83
	contig60751.1	100	60.2	---NA---	7.42E-09	167	136	46.38
	contig59941.1	100	60.2	---NA---	7.42E-09	84	53	46.38
	contig58971.1	97.1	58.4	---NA---	2.67E-08	124	91	49.28
	contig56349.1	100	60.2	---NA---	7.42E-09	765	734	46.38
	contig55153.1	97	56.5	---NA---	9.60E-08	263	231	47.83
	contig54969.1	97	56.5	---NA---	9.60E-08	726	758	47.83
	contig54325.1	100	54.7	---NA---	3.45E-07	674	646	42.03
	contig51877.1	100	52.8	---NA---	1.24E-06	396	369	40.58
	contig49724.1	100	60.2	---NA---	7.42E-09	468	437	46.38
	contig47238.2	97	56.5	---NA---	9.60E-08	713	745	47.83

Appendix 2.2 (cont.)

SNP (#)	Contig matched	% Pairwise Identity	Bit-Score	Contig Description	E Value	Hit end	Hit start	Query coverage (%)
PM6488_B	contig47238.1	97	56.5	---NA---	9.60E-08	852	884	47.83
	contig39463.2	100	54.7	---NA---	3.45E-07	747	719	42.03
	contig34890.1	100	54.7	---NA---	3.45E-07	161	133	42.03
	contig33357.1	100	62.1	---NA---	2.06E-09	1,360	1,392	47.83
	contig29020.1	100	54.7	---NA---	3.45E-07	1,314	1,342	42.03
	contig23832.1	96.8	52.8	---NA---	1.24E-06	1,479	1,509	44.93
	contig21081.1	96.9	54.7	---NA---	3.45E-07	837	806	46.38
	contig19487.1	100	56.5	---NA---	9.60E-08	1,328	1,299	43.48
	contig15653.1	100	56.5	---NA---	9.60E-08	88	59	43.48
	contig14585.1	100	54.7	---NA---	3.45E-07	1,202	1,174	42.03
	contig13069.1	100	54.7	---NA---	3.45E-07	684	656	42.03
	contig11529.1	100	60.2	---NA---	7.42E-09	428	459	46.38
	contig9666.1	96.9	54.7	calcium-activated potassium channel slowpoke isoform X7	3.45E-07	3	34	46.38
	contig8956.1	100	54.7	---NA---	3.45E-07	370	342	42.03
PM1771_A	contig169701.1	95.9	78.7	RNA-directed DNA polymerase from mobile element jockey	2.05E-14	69	21	69.57
	contig162921.1	94.1	76.8	---NA---	7.37E-14	51	1	72.46
	contig155224.1	97.1	58.4	---NA---	2.67E-08	103	69	49.28
	contig123561.1	100	54.7	---NA---	3.45E-07	32	4	42.03
	contig72709.1	96.1	82.4	RNA-directed DNA polymerase from mobile element jockey	1.58E-15	249	299	72.46
	contig41955.1	96.1	82.4	RNA-directed DNA polymerase from mobile element jockey-like	1.58E-15	463	513	72.46
	contig32221.1	87.8	80.5	RNA-directed DNA polymerase from mobile element jockey	5.70E-15	1,490	1,417	95.65
	contig29186.1	96.1	82.4	RNA-directed DNA polymerase from mobile element jockey-like	1.58E-15	1,066	1,116	72.46
	contig22935.1	100	54.7	rna-directed dna polymerase from mobile element	3.45E-07	906	878	42.03
	contig17548.1	96.1	82.4	RNA-directed DNA polymerase from mobile element jockey-like	1.58E-15	1,314	1,364	72.46

Appendix 2.2 (cont.)

SNP (#)	Contig matched	% Pairwise Identity	Bit-Score	Contig Description	E Value	Hit end	Hit start	Query coverage (%)
PM1771_A	contig8125.3	94.1	76.8	RNA-directed DNA polymerase from mobile element jockey-like	7.37E-14	355	405	72.46
	contig8125.2	96.1	82.4	RNA-directed DNA polymerase from mobile element jockey	1.58E-15	1,722	1,772	72.46
	contig8125.1	96.1	82.4	rna-directed dna polymerase from mobile element jockey	1.58E-15	2,316	2,366	72.46
	contig8082.1	94.1	76.8	lian-aa1 retrotransposon	7.37E-14	2,428	2,478	72.46
	contig4273.1	87.1	73.1	RNA-directed DNA polymerase from mobile element jockey-like	9.53E-13	2,488	2,419	89.86
	contig3120.1	96.1	82.4	rna-directed dna polymerase from mobile element jockey	1.58E-15	1,491	1,441	72.46
PM1771_B	contig169701.1	95.9	78.7	RNA-directed DNA polymerase from mobile element jockey	2.05E-14	69	21	69.57
	contig162921.1	94.1	76.8	---NA---	7.37E-14	51	1	72.46
	contig155224.1	97.1	58.4	---NA---	2.67E-08	103	69	49.28
	contig123561.1	100	54.7	---NA---	3.45E-07	32	4	42.03
	contig72709.1	96.1	82.4	RNA-directed DNA polymerase from mobile element jockey	1.58E-15	249	299	72.46
	contig41955.1	96.1	82.4	RNA-directed DNA polymerase from mobile element jockey-like	1.58E-15	463	513	72.46
	contig32221.1	95.9	78.7	RNA-directed DNA polymerase from mobile element jockey	2.05E-14	1,490	1,442	69.57
	contig29186.1	96.1	82.4	RNA-directed DNA polymerase from mobile element jockey-like	1.58E-15	1,066	1,116	72.46
	contig22935.1	100	54.7	rna-directed dna polymerase from mobile element	3.45E-07	906	878	42.03
	contig17548.1	96.1	82.4	RNA-directed DNA polymerase from mobile element jockey-like	1.58E-15	1,314	1,364	72.46
	contig8125.3	94.1	76.8	RNA-directed DNA polymerase from mobile element jockey-like	7.37E-14	355	405	72.46
	contig8125.2	96.1	82.4	RNA-directed DNA polymerase from mobile element jockey	1.58E-15	1,722	1,772	72.46
	contig8125.1	96.1	82.4	rna-directed dna polymerase from mobile element jockey	1.58E-15	2,316	2,366	72.46
	contig8082.1	86.8	78.7	lian-aa1 retrotransposon	2.05E-14	2,428	2,503	98.55
	contig4273.1	95.6	71.3	RNA-directed DNA polymerase from mobile element jockey-like	3.43E-12	2,488	2,444	63.77
	contig3120.1	96.1	82.4	rna-directed dna polymerase from mobile element jockey	1.58E-15	1,491	1,441	72.46

Appendix 2.2 (cont.)

SNP (#)	Contig matched	% Pairwise Identity	Bit-Score	Contig Description	E Value	Hit end	Hit start	Query coverage (%)
PM4714_A	contig5062.3	100	121.2	PREDICTED: uncharacterized protein LOC108744461 isoform X4	3.36E-27	348	412	94.2
	contig5062.2	100	121.2	Uncharacterized protein APZ42_034504	3.36E-27	335	399	94.2
	contig5062.1	100	121.2	Uncharacterized protein APZ42_034504	3.36E-27	1,169	1,233	94.2
PM4714_A	contig5062.3	98.5	115.6	PREDICTED: uncharacterized protein LOC108744461 isoform X4	1.56E-25	348	412	94.2
	contig5062.2	98.5	115.6	Uncharacterized protein APZ42_034504	1.56E-25	335	399	94.2
	contig5062.1	98.5	115.6	Uncharacterized protein APZ42_034504	1.56E-25	1,169	1,233	94.2
PM3856_A	contig9977.1	100	54.7	kinase pkn|GO:0008484 GO:0008152	3.45E-07	857	885	42.03
	contig2078.1	100	54.7	kinase pkn|GO:0008484 GO:0008152	3.45E-07	1,459	1,487	42.03
PM3856_B	contig14798.1	100	54.7	kinase pkn|GO:0008484 GO:0008152	3.45E-07	2,223	2,195	42.03
PM10591_A	NO HIT							0
PM10591_B	contig113267.1	100	62.1	---NA---	2.06E-09	156	188	47.83
PM6057_A	contig170249.1	91.5	82.4	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	1.58E-15	124	66	85.51
PM6057_B	contig212013.1	91.4	80.5	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	59	2	84.06
	contig210575.1	88.1	71.3	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	3.43E-12	87	29	85.51
	contig195976.1	89.8	76.8	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	7.37E-14	78	20	85.51
	contig191278.1	94.9	62.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.06E-09	338	300	56.52
	contig151251.1	91.1	76.8	RNA-directed DNA polymerase from mobile element jockey| GO:0003964 GO:0006278	7.37E-14	389	334	81.16
	contig108767.1	89.8	76.8	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	7.37E-14	177	119	85.51
	contig100612.1	91.4	80.5	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	354	297	84.06
	contig99594.1	91.1	76.8	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	7.37E-14	352	407	81.16
	contig20642.1	89.7	75.0	endonuclease-reverse transcriptase	2.65E-13	766	709	84.06
	contig15540.1	92.9	82.4	endonuclease-reverse transcriptase	1.58E-15	1,723	1,778	81.16
	contig2742.1	90.9	58.4	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.67E-08	2,057	2,015	63.77

Appendix 2.2 (cont.)

SNP (#)	Contig matched	% Pairwise Identity	Bit-Score	Contig Description	E Value	Hit end	Hit start	Query coverage (%)
PM2065_A	contig160882.1	93.5	67.6	---NA---	4.44E-11	156	112	66.67
PM2065_B	NO HIT							0
PM5846_A	contig69233.1	91.9	87.9	---NA---	3.40E-17	774	713	89.86
PM5846_B	contig69233.1	90.3	82.4	---NA---	1.58E-15	774	713	89.86
PM10621_A	NO HIT							0
PM10621_B	contig83159.1	95.2	67.6	RNA-directed DNA polymerase from mobile element| GO:0003824 GO:0005488 GO:0044260 GO:0090304	4.44E-11	1	42	60.87
	contig53029.1	95.9	80.5	RNA-directed DNA polymerase from mobile element	5.70E-15	1,011	963	71.01
	contig27951.1	93.9	75.0	RNA-directed DNA polymerase from mobile element jockey-like| GO:0003824 GO:0005488 GO:0044260 GO:0090304	2.65E-13	558	510	71.01
	contig17322.1	93.9	75.0	RNA-directed DNA polymerase from mobile element	2.65E-13	235	283	71.01
PM1296_A	contig46386.1	97.3	63.9	venom protease-like|GO:0008233	5.74E-10	287	251	53.62
	contig40687.1	97.3	63.9	venom protease-like|GO:0008233	5.74E-10	270	234	53.62
	contig21273.1	97.3	63.9	venom protease-like|GO:0008233	5.74E-10	275	239	53.62
	contig17797.1	97.3	63.9	venom protease-like|GO:0008233	5.74E-10	326	290	53.62
	contig6796.1	97.3	63.9	prophenoloxidase activating factor 1| GO:0004252 GO:0006508	5.74E-10	3,195	3,231	53.62
PM1296_B	contig38125.1	97.3	63.9	venom protease-like|GO:0008233	5.74E-10	359	323	53.62

Two alleles generated by outlier SNP denoted as "_A" and "_B"; Shading denotes outlier SNP with only one allele that matched *P. monodon* transcriptome.

Appendix 2.3 Translation of eight outlier SNP with transcriptome matches to protein sequences using all three reading frames.

SNP ID	Sequence (5' - 3')	Translation Frame 1 (5' - 3')	Translation Frame 2 (5' - 3')
PM1958_A	TGCAGACGCGTACGCCGCCTACCCGTACCCCTACCACCCAGGCTACAGCCACGAGGAGCTGCCCCCATA	CRRVRRLPVPLPPRLQPRGAAPI	**ADAYAAYPYPYHPGYSHEELPP**
PM1958_B	TGCAGACGCGTACGCCGCCTACCCATACCCCTACCACCCAGGCTACAGCCACGAGGAGCTGCCCCCATA	CRRVRRLPIPLPPRLQPRGAAPI	**ADAYAAYPYPYHPGYSHEELPP**
PM6488_A	TGCAGGTACCAGTCACCCCAGGATGCTGTGCATGATCCGAGATCGGAAGAGCGGTTCAGCAGGAATGCC	**CRYQSPQDAVHDPRSEERFSRNA**	AGTSHPRMLCMIRDRKSGSAGM
PM6488_B	TGCAGGTACCAGTCGCCCCAGGATGCTGTGCATGATCCGAGATCGGAAGAGCGGTTCAGCAGGAATGCC	**CRYQSPQDAVHDPRSEERFSRNA**	AGTSRPRMLCMIRDRKSGSAGM
PM1771_A	TGCAGCACACGACCGAGTGGGGGGACTTCAATGCACACCACCCCATCCTGGCTCCCTACAGGGCACCGA	CSTRPSGGTSMHTTPSWLPTGHR	AAHDRVGGLQCTPPHPGSLQGT
PM1771_B	TGCAGCACACGACTGAGTGGGGGGACTTCAATGCACACCACCCCATCCTGGCTCCCTACAGGGCACCGA	CSTRLSGGTSMHTTPSWLPTGHR	AAHD*VGGLQCTPPHPGSLQGT
PM4714_A	TGCAGTACATAAGGCACAGCGAGCTGGACGGGCACTCGCGGCCGCACTTGTAGCACACGAAGCCGAGAT	CST*GTASWTGTRGRTCSTRSRD	AVHKAQRAGRALAAALVAHEAE
PM4714_B	TGCAGTACATCAGGCACAGCGAGCTGGACGGGCACTCGCGGCCGCACTTGTAGCACACGAAGCCGAGAT	CSTSGTASWTGTRGRTCSTRSRD	AVHQAQRAGRALAAALVAHEAE
PM3856_A	TGCAGCTGGGAATCCTGCTGTGTCTGCCGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATC	CSWESCCVCRDRKSGSAGMPRPI	AAGNPAVSAEIGRAVQQECRDR
PM3856_B	TGCAGCTGTGAATCCTGCTGTGTCTGCCGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATC	CSCESCCVCRDRKSGSAGMPRPI	AAVNPAVSAEIGRAVQQECRDR
PM6057_A	TGCAGCAACTACTGTGACATTACACTGCTCAGTATACCAGACAAGATTTTCACCCACATCCGTGATAAC	CSNYCDITLLSIPDKIFTHIRDN	AATTVTLHCSVYQTRFSPTSVI
PM6057_B	TGCAGCAACTACTGTGGCATTACACTGCTCAGTATACCAGACAAGATTTTCACCCACATCCGTGATAAC	CSNYCGITLLSIPDKIFTHIRDN	AATTVALHCSVYQTRFSPTSVI
PM5846_A	TGCAGCTTCAGTCCAGGCCACAAATTTCTGTTGCTTTGTCCAATGAGATTAATTACAGCGTTCGGCATG	CSFSPGHKFLLLCPMRLITAFGM	AASVQATNFCCFVQ*D*LQRSA
PM5846_B	TGCAGATTCAGTCCAGGCCACAAATTTCTGTTGCTTTGTCCAATGAGATTAATTACAGCGTTCGGCATG	CRFSPGHKFLLLCPMRLITAFGM	ADSVQATNFCCFVQ*D*LQRSA
PM10621_A	TGCAGGATATCAAGCAAGGGACGGCGCCCCCGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG	CRISSKGRRPRDRKSGSAGMPRP	**AGYQARDGAPEIGRAVQQECRD**
PM10621_B	TGCAGGATATCAAGCAAGGGACGGCGCTCCCGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG	CRISSKGRRSRDRKSGSAGMPRP	**AGYQARDGAPEIGRAVQQECRD**

Appendix 2.3 (cont.)

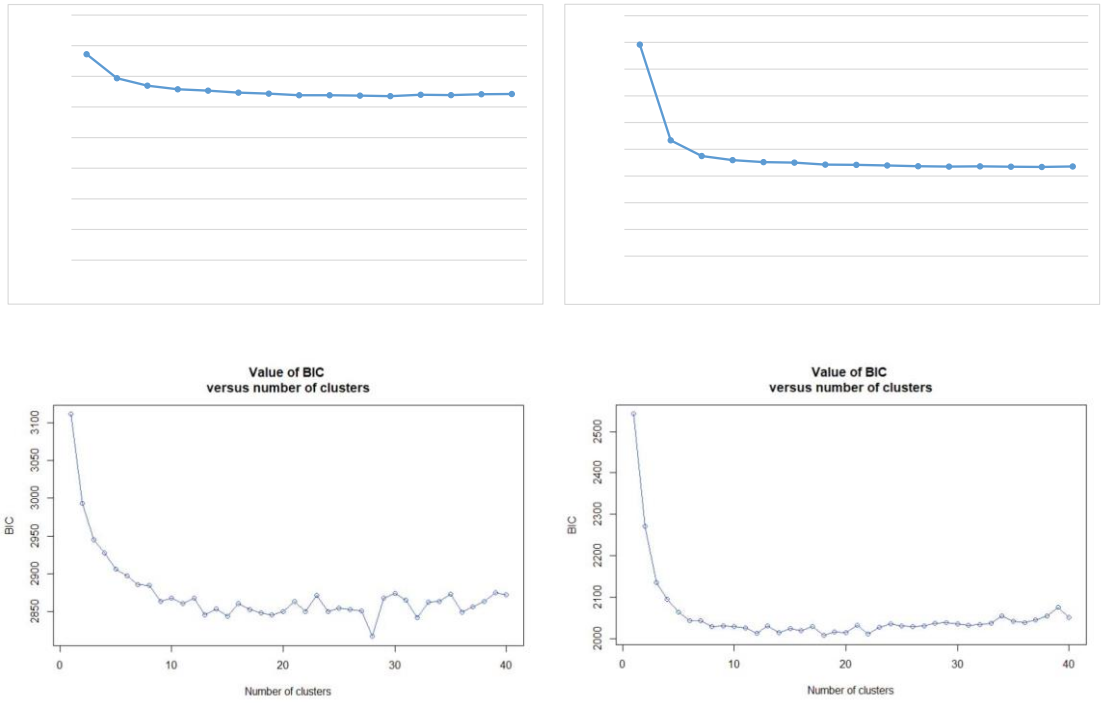
SNP ID	Translation Frame 3 (5' - 3')	Translation Frame 1 (3' - 5')	Translation Frame 2 (3' - 5')	Translation Frame 3 (3' - 5')
PM1958_A	QTRTPPTRTPTTQATATRSCPH	**YGGSSSWL*PGW*GYG*AAYASA**	MGAAPRGCSLGGRGTGRRRTRL	WGQLLVAVAWVVGVRVGGVRVC
PM1958_B	QTRTPPTHTPTTQATATRSCPH	**YGGSSSWL*PGW*GYG*AAYASA**	MGAAPRGCSLGGRGMGRRRTRL	WGQLLVAVAWVVGVWVGGVRVC
PM6488_A	QVPVTPGCCA*SEIGRAVQQEC	GIPAEPLFRSRIMHSILG*LVPA	**AFLLNRSSDLGSCTASWGDWYL**	HSC*TALPISDHAQHPGVTGTC
PM6488_B	QVPVAPGCCA*SEIGRAVQQEC	GIPAEPLFRSRIMHSILGRLVPA	**AFLLNRSSDLGSCTASWGDWYL**	HSC*TALPISDHAQHPGATGTC
PM1771_A	**QHTTEWGDFNAHHPILAPYRAP**	SVPCREPGWGGVH*SPPTRSCAA	RCPVGSQDGVVCIEVPPLGRVL	**GAL*GARMGWCALKSPHSVVCC**
PM1771_B	**QHTTEWGDFNAHHPILAPYRAP**	SVPCREPGWGGVH*SPPTQSCAA	RCPVGSQDGVVCIEVPPLSRVL	**GAL*GARMGWCALKSPHSVVCC**
PM4714_A	**QYIRHSELDGHSRPHL*HTKPR**	ISASCATSAAASARPARCALCTA	SRLRVLQVRPRVPVQLAVPYVL	**LGFVCYKCGRECPSSSLCLMYC**
PM4714_B	**QYIRHSELDGHSRPHL*HTKPR**	ISASCATSAAASARPARCA*CTA	SRLRVLQVRPRVPVQLAVPDVL	**LGFVCYKCGRECPSSSLCLMYC**
PM3856_A	QLGILLCLPRSEERFSRNAETD	DRSRHSC*TALPISADTAGFPAA	IGLGIPAEPLFRSRQTQQDSQL	SVSAFLLNRSSDLGRHSRIPSC
PM3856_B	QL*ILLCLPRSEERFSRNAETD	DRSRHSC*TALPISADTAGFTAA	IGLGIPAEPLFRSRQTQQDSQL	SVSAFLLNRSSDLGRHSRIHSC
PM6057_A	QQLL*HYTAQYTRQDFHPHP**	VITDVGENLVWYTEQCNVTVVAA	LSRMWVKILSGILSSVMSQ*LL	**YHGCG*KSCLVY*AV*CHSSCC**
PM6057_B	QQLLWHYTAQYTRQDFHPHP**	VITDVGENLVWYTEQCNATVVAA	LSRMWVKILSGILSSVMPQ*LL	**YHGCG*KSCLVY*AV*CHSSCC**
PM5846_A	QLQSRPQISVALSNEINYSVRH	HAERCN*SHWTKQQKFVAWTEAA	**MPNAVINLIGQSNRNLWPGLKL**	CRTL*LISLDKATEICGLD*SC
PM5846_B	QIQSRPQISVALSNEINYSVRH	HAERCN*SHWTKQQKFVAWTESA	**MPNAVINLIGQSNRNLWPGLNL**	CRTL*LISLDKATEICGLD*IC
PM10621_A	QDIKQGTAPPRSEERFSRNAET	**RSRHSC*TALPISGAPSLA*YPA**	GLGIPAEPLFRSRGRRPLLDIL	VSAFLLNRSSDLGGAVPCLISC
PM10621_B	QDIKQGTALPRSEERFSRNAET	**RSRHSC*TALPISGAPSLA*YPA**	GLGIPAEPLFRSRERRPLLDIL	VSAFLLNRSSDLGSAVPCLISC

Two alleles generated by each outlier SNP denoted as "A" and "_B"; Red characters indicate SNP locations within each nucleotide or translated protein sequence; Protein translation frames in which SNP caused synonymous mutations are bolded; Asterisks (*) denote stop codons.

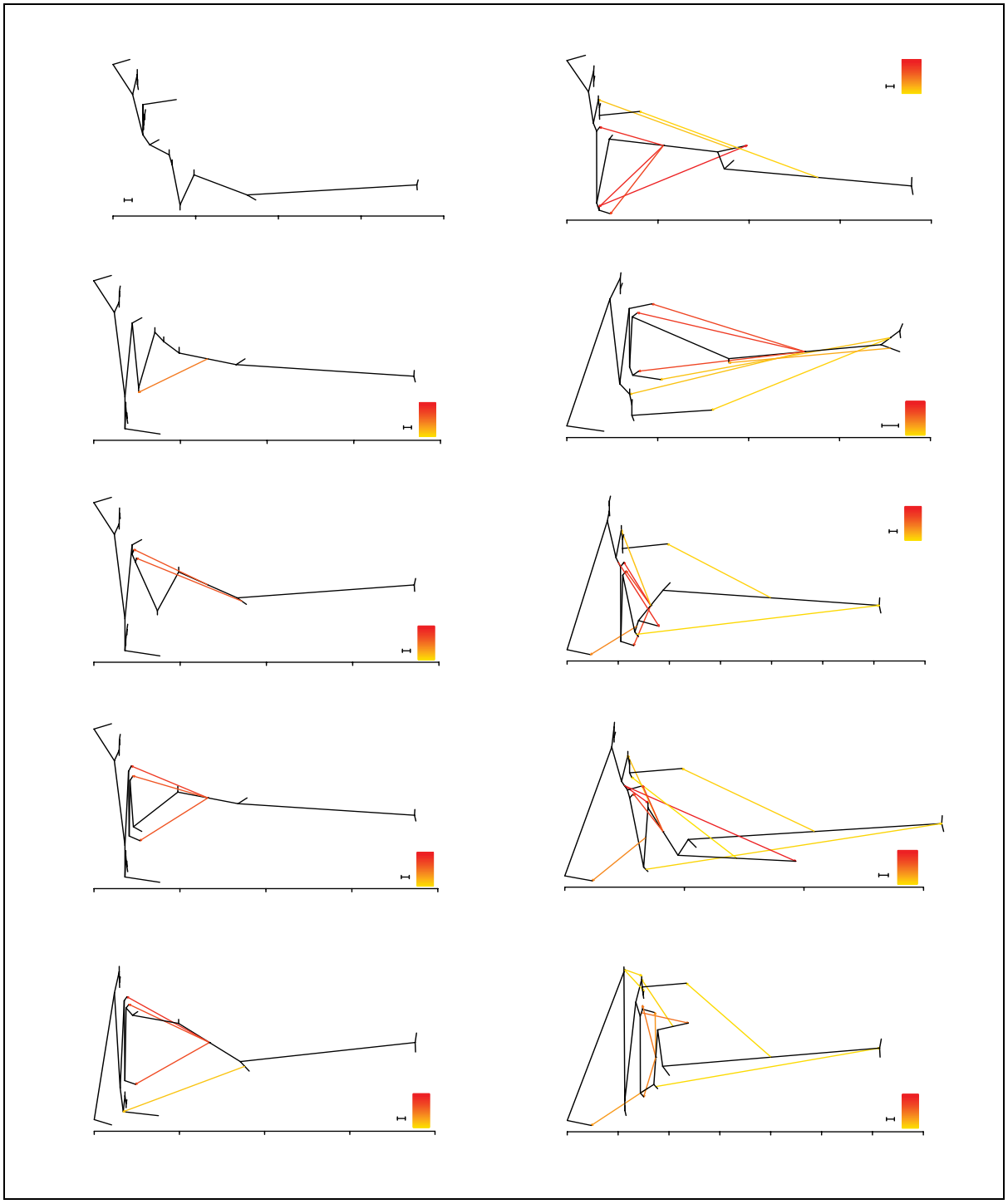
Chapter 3.

Appendix 3.1 Sea surface temperature maximum (SST_max) and minimum (SST_min) relating to each population sample site were selected from the Bio-ORACLE database using monthly averages of climate data from 2000 – 2014.

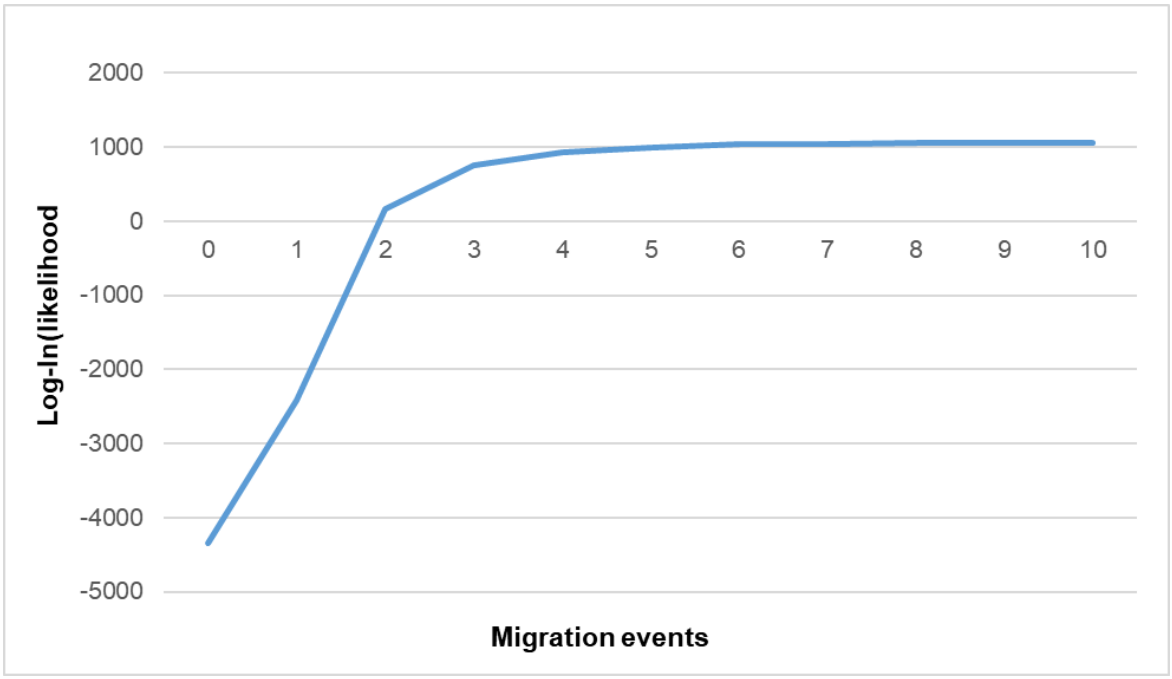
	Populations	Latitude	Longitude	SST_max	SST_min
Oceania	BB	-17.56179	146.108953	29.57	22.18
	EB	-17.35300	146.052726	29.59	22.04
	TSV	-19.08210	146.817537	29.58	21.06
	GC	-14.99765	138.780787	30.94	23.56
	JBG	-14.68180	128.371358	31.39	21.61
	TIW	-11.86683	131.83147	30.79	24.04
	WA	-20.69701	116.863229	31.48	21.25
	FJ	-17.95706	178.608177	29.47	24.78
Southeast Asia (SEA)	PHI	10.546042	122.447387	30.46	26.39
	CMVN	8.5676	105.0526	31.10	25.36
	NTVN	12.2608	109.2383	29.95	23.20
	INDO	-7.702932	108.706541	29.92	24.60
	THL	11.693318	97.657631	31.02	26.95
		11.937713	101.465872		
Northwest Indian Ocean (NWIO)	SLK	6.434407	79.974896	30.20	26.96
Western Indian Ocean (WIO	KE	-3.583314	40.0142664	29.43	25.10
	SA	-28.97839	31.836385	26.36	20.29



Appendix 3.2 Graph of the number of clusters formed by DAPC and ADMIXTURE. The top plot shows the cross-validation error across multiple runs picking K values of 7 and 6 as the optimum number of clusters for neutral (**A**) and outlier loci (**C**) datasets. Inference of the number of clusters in the DAPC performed on neutral (**B**) and outlier loci (**D**) datasets; K values of 7 and 6 (the lowest BIC value) represented the best summary of neutral and outlier loci, respectively.



Appendix 3.3 Population relationship for tree models with ten migration events (m = 0 – 10) inferred by Treemix for neutral SNP dataset. The drift parameter represents the degree to which genetic drift has occurred between each population. BB: Bramston Beach; EB: Etty Bay; TSV: Townsville; GC: Gulf of Carpentaria, JBG: Joseph Bonaparte Gulf; TIW: Tiwi Island, WA: Western Australia; CMVN: Ca Mau, Vietnam; NTVN: Nha Trang, Vietnam; INDO: Indonesia; THL: Thailand; PHI: Philippines; SLK: Sri Lanka; KE: Kenya; SA: South Africa.



Appendix 3.4 The fractions of variance in relatedness between populations explained by phylogenetic models with 0 to 10 migration events.

Appendix 3.5 List of all population triples (X; A, B) with negative values of f3 (Z; A, B) in three-way admixture test among 16 *P. monodon* populations. BB: Bramston Beach; EB: Etty Bay; TSV: Townsville; GC: Gulf of Carpentaria; JBG: Joseph Bonaparte Gulf; TIW: Tiwi Island; NKB: Nickol Bay; CMVN: Ca Mau, Vietnam; NTVN: Nha Trang, Vietnam; FJ: Fiji; INDO: Java, Indonesia; THL: Thailand; PHI: Philippines; SLK: Sri Lanka; KE: Kenya; SA: South Africa.

	Source 1	Source 2	Target	The *f3*-statistics	The standard error	The Z-score
Sixteen *f3*-statistics were significantly positive (Z-score < -20)	BB	SLK	CMVN	-0.00627252	0.000238837	-26.2628
	EB	SLK	CMVN	-0.00637368	0.000247282	-25.7749
	FJ	SLK	CMVN	-0.00616953	0.000290652	-21.2265
	GC	SLK	CMVN	-0.00493383	0.000195865	-25.1899
	JBG	SLK	CMVN	-0.00499505	0.00019597	-25.4888
	PHI	SLK	CMVN	-0.00497183	0.000220372	-22.561
	SLK	TIW	CMVN	-0.00499591	0.000179561	-27.8229
	SLK	TSV	CMVN	-0.00635146	0.0002497	-25.4363
	BB	SLK	INDO	-0.00706338	0.00023785	-29.6967
	EB	SLK	INDO	-0.00713797	0.000243458	-29.3191
	FJ	SLK	INDO	-0.00709535	0.000283549	-25.0234
	GC	SLK	INDO	-0.00592796	0.000196708	-30.1358
	JBG	SLK	INDO	-0.00605581	0.000192236	-31.502
	PHI	SLK	INDO	-0.00573417	0.000222006	-25.8289
	SLK	TIW	INDO	-0.00596621	0.000184414	-32.3523
	SLK	TSV	INDO	-0.00715243	0.000248002	-28.8402
***f3*-statistics were significantly positive (Z-score < 0)**	EB	NKB	BB	-8.50E-05	9.03E-05	-0.941279
	EB	SLK	BB	-2.71E-05	0.000137665	-0.196756
	EB	JBG	BB	-8.55E-06	5.60E-05	-0.152851
	KE	TSV	BB	-0.000124222	0.000214318	-0.579617
	NKB	TSV	BB	-9.11E-06	0.000102147	-0.0891749
	SA	TSV	BB	-8.46E-05	0.000214548	-0.394153
	SLK	TSV	BB	-1.78E-05	0.000155131	-0.114651
	BB	SA	CMVN	-0.00619892	0.000343701	-18.0358
	BB	KE	CMVN	-0.00616245	0.000345254	-17.849
	BB	THL	CMVN	-0.00104389	0.0001283	-8.13631
	BB	INDO	CMVN	-0.00026855	8.22E-05	-3.26803
	EB	SA	CMVN	-0.0062005	0.000359592	-17.2432
	EB	KE	CMVN	-0.00620207	0.000362096	-17.1283
	EB	THL	CMVN	-0.00111778	0.000134065	-8.3376
	EB	INDO	CMVN	-0.000295118	8.46E-05	-3.48807
	FJ	SA	CMVN	-0.00591394	0.000457134	-12.937
	FJ	KE	CMVN	-0.00592354	0.000458042	-12.9323

Appendix 3.5 (cont.)

	Source 1	Source 2	Target	The *f3*-statistics	The standard error	The Z-score
f3-statistics were significantly positive (Z-score < 0)	FJ	THL	CMVN	-0.00104896	0.00016757	-6.25986
	FJ	INDO	CMVN	-0.00013359	0.000105612	-1.26491
	FJ	INDO	CMVN	-0.00013359	0.000105612	-1.26491
	GC	KE	CMVN	-0.00490591	0.000298093	-16.4576
	GC	SA	CMVN	-0.00487082	0.000296389	-16.4339
	GC	THL	CMVN	-0.00066567	0.000111541	-5.96796
	GC	INDO	CMVN	-6.53E-05	7.14E-05	-0.91395
	INDO	PHI	CMVN	-0.000297063	8.10E-05	-3.66553
	INDO	TSV	CMVN	-0.000258441	8.62E-05	-2.99782
	INDO	TIW	CMVN	-8.91E-05	6.71E-05	-1.3284
	JBG	SA	CMVN	-0.00484985	0.000301851	-16.0671
	JBG	KE	CMVN	-0.00486272	0.000303709	-16.0111
	JBG	THL	CMVN	-0.000678441	0.000113735	-5.96511
	KE	TSV	CMVN	-0.00634783	0.00036265	-17.504
	KE	TIW	CMVN	-0.00481594	0.000285106	-16.8917
	KE	PHI	CMVN	-0.00434812	0.000345528	-12.584
	KE	NKB	CMVN	-0.00357229	0.000441941	-8.08317
	NKB	SLK	CMVN	-0.00401155	0.000280349	-14.3092
	NKB	SA	CMVN	-0.00358314	0.000439911	-8.14514
	NKB	THL	CMVN	-0.000630178	0.000160751	-3.9202
	PHI	SA	CMVN	-0.00435761	0.00034389	-12.6715
	PHI	THL	CMVN	-0.0009371	0.000130194	-7.19772
	SA	TSV	CMVN	-0.00634464	0.000360768	-17.5865
	SA	TIW	CMVN	-0.00480472	0.000283203	-16.9656
	THL	TSV	CMVN	-0.00101474	0.00013409	-7.56759
	THL	TIW	CMVN	-0.000706386	0.000105493	-6.69604
	BB	SLK	GC	-0.000237717	0.000278298	-0.854181
	BB	SA	GC	-0.000227122	0.000339745	-0.668508
	BB	KE	GC	-0.00015556	0.000341149	-0.45599
	EB	SLK	GC	-0.000277948	0.00029147	-0.953608
	EB	SA	GC	-0.00016777	0.000360571	-0.46529
	EB	KE	GC	-0.000134249	0.000362263	-0.370583
	KE	TSV	GC	-0.000451954	0.000357761	-1.26329
	SA	TSV	GC	-0.000483859	0.0003563	-1.35801
	SLK	TSV	GC	-0.000427674	0.000288347	-1.48319
	BB	SA	INDO	-0.00683809	0.000363417	-18.8161
	BB	KE	INDO	-0.00677549	0.000365343	-18.5456

Appendix 3.5 (cont.)

	Source 1	Source 2	Target	The *f3*-statistics	The standard error	The Z-score
f3-statistics were significantly positive (Z-score < 0)	BB	THL	INDO	-0.000784002	0.000132553	-5.91464
	CMVN	SLK	INDO	-0.00105941	0.000110817	-9.55999
	CMVN	SA	INDO	-0.000907726	0.000188901	-4.80529
	CMVN	KE	INDO	-0.000881598	0.000190387	-4.63056
	CMVN	THL	INDO	-8.67E-06	6.81E-05	-0.127209
	EB	SA	INDO	-0.0068131	0.000378869	-17.9827
	EB	KE	INDO	-0.00678855	0.000381449	-17.7968
	EB	THL	INDO	-0.000831328	0.000137698	-6.03731
	FJ	SA	INDO	-0.00668807	0.000464713	-14.3918
	FJ	KE	INDO	-0.00667155	0.000466026	-14.3158
	FJ	THL	INDO	-0.000924038	0.000168985	-5.46815
	GC	SA	INDO	-0.00571327	0.000314016	-18.1942
	GC	KE	INDO	-0.00572224	0.000315873	-18.1156
	GC	THL	INDO	-0.00060906	0.0001124	-5.41867
	JBG	SA	INDO	-0.00575893	0.000314662	-18.3019
	JBG	KE	INDO	-0.00574567	0.000316687	-18.143
	JBG	THL	INDO	-0.000688458	0.000113732	-6.05331
	KE	TIW	INDO	-0.00560843	0.000305048	-18.3854
	KE	TSV	INDO	-0.00697098	0.000382182	-18.2399
	KE	PHI	INDO	-0.00493266	0.000361505	-13.6448
	KE	NKB	INDO	-0.00445811	0.000443012	-10.0632
	KE	NTVN	INDO	-0.000950451	0.000252416	-3.76542
	NKB	SLK	INDO	-0.00507519	0.000268676	-18.8896
	NKB	SA	INDO	-0.00449509	0.00044088	-10.1957
	NKB	THL	INDO	-0.000643072	0.000156092	-4.11982
	NTVN	SLK	INDO	-0.00109549	0.000147415	-7.43135
	NTVN	SA	INDO	-0.00102373	0.000250778	-4.08223
	PHI	SA	INDO	-0.00496827	0.000358948	-13.8412
	PHI	THL	INDO	-0.000648703	0.000132628	-4.89115
	SA	TIW	INDO	-0.00562333	0.000303244	-18.5439
	SA	TSV	INDO	-0.00699392	0.000380522	-18.3798
	THL	TIW	INDO	-0.000625944	0.000109014	-5.74185
	THL	TSV	INDO	-0.000764961	0.000136328	-5.6112
	BB	SLK	JBG	-0.000451332	0.000261173	-1.72809
	BB	SA	JBG	-0.000522928	0.000318266	-1.64305
	BB	KE	JBG	-0.000473595	0.000319432	-1.48262

Appendix 3.5 (cont.)

	Source 1	Source 2	Target	The *f3*-statistics	The standard error	The Z-score
f3-statistics were significantly positive (Z-score < 0)	BB	GC	JBG	-7.55E-05	8.17E-05	-0.92452
	EB	SLK	JBG	-0.000469865	0.000276395	-1.69997
	EB	SA	JBG	-0.000441878	0.000340195	-1.29889
	EB	KE	JBG	-0.000430585	0.000342135	-1.25852
	EB	GC	JBG	-5.38E-05	8.38E-05	-0.642513
	FJ	GC	JBG	-0.000114828	0.00010745	-1.06867
	KE	TSV	JBG	-0.000685327	0.000347261	-1.97352
	SA	TSV	JBG	-0.000695003	0.000345991	-2.00873
	SLK	TSV	JBG	-0.000556627	0.000280221	-1.98639
	BB	SLK	NTVN	-0.00452098	0.000253016	-17.8684
	BB	KE	NTVN	-0.00437814	0.000377672	-11.5925
	BB	SA	NTVN	-0.00436746	0.000377236	-11.5775
	EB	SLK	NTVN	-0.00455818	0.000261926	-17.4026
	EB	KE	NTVN	-0.00435379	0.000393452	-11.0656
	EB	SA	NTVN	-0.00430507	0.000392461	-10.9694
	FJ	SLK	NTVN	-0.00437791	0.00031528	-13.8858
	FJ	KE	NTVN	-0.00409916	0.000492594	-8.32157
	FJ	SA	NTVN	-0.0040424	0.000492278	-8.21162
	GC	SLK	NTVN	-0.00314799	0.000218016	-14.4393
	GC	KE	NTVN	-0.00308731	0.000335519	-9.20157
	GC	SA	NTVN	-0.00300506	0.000334457	-8.98488
	JBG	SLK	NTVN	-0.00321716	0.000219023	-14.6887
	JBG	KE	NTVN	-0.00305205	0.000339957	-8.97775
	JBG	SA	NTVN	-0.00299203	0.000338999	-8.82609
	KE	TSV	NTVN	-0.00457973	0.00039389	-11.6269
	KE	TIW	NTVN	-0.00296203	0.000323834	-9.14673
	KE	PHI	NTVN	-0.00226193	0.000383793	-5.89363
	KE	NKB	NTVN	-0.00158131	0.000486791	-3.24843
	NKB	SLK	NTVN	-0.00205334	0.00030976	-6.62883
	NKB	SA	NTVN	-0.00154501	0.000486133	-3.17815
	PHI	SLK	NTVN	-0.00291841	0.000244208	-11.9505
	PHI	SA	NTVN	-0.00222426	0.000382946	-5.80829
	SA	TSV	NTVN	-0.00452939	0.000393488	-11.5109
	SA	TIW	NTVN	-0.00290365	0.000323308	-8.98107
	SLK	TSV	NTVN	-0.00461614	0.000262794	-17.5656
	SLK	TIW	NTVN	-0.00317477	0.000203604	-15.5929

Appendix 3.5 (cont.)

	Source 1	Source 2	Target	The *f3*-statistics	The standard error	The Z-score
f3-statistics were significantly positive (Z-score < 0)	BB	SLK	THL	-0.00272177	0.00026495	-10.2728
	BB	SA	THL	-0.00226456	0.000444688	-5.09248
	BB	KE	THL	-0.00221158	0.000446876	-4.94896
	EB	SLK	THL	-0.00274904	0.000271865	-10.1118
	EB	SA	THL	-0.00219225	0.000460835	-4.75713
	EB	KE	THL	-0.0021773	0.00046313	-4.70128
	FJ	SLK	THL	-0.0026137	0.000314456	-8.31184
	FJ	SA	THL	-0.00197451	0.000545738	-3.61805
	FJ	KE	THL	-0.0019676	0.000547186	-3.59585
	GC	SLK	THL	-0.0017613	0.000228247	-7.71662
	GC	KE	THL	-0.00133326	0.00041056	-3.24742
	GC	SA	THL	-0.00131469	0.000408076	-3.22168
	JBG	SLK	THL	-0.00180975	0.000224975	-8.04424
	JBG	SA	THL	-0.00128095	0.000407828	-3.14091
	JBG	KE	THL	-0.00127729	0.000410781	-3.10943
	KE	TSV	THL	-0.00242611	0.000465061	-5.21675
	KE	TIW	THL	-0.00120257	0.000397136	-3.02809
	KE	PHI	THL	-0.000504036	0.000455685	-1.10611
	KE	NKB	THL	-3.51E-05	0.000511083	-0.068724
	NKB	SLK	THL	-0.000874515	0.000292192	-2.99295
	NKB	SA	THL	-6.25E-05	0.000507621	-0.123114
	PHI	SLK	THL	-0.00152786	0.000258096	-5.91975
	PHI	SA	THL	-0.00053004	0.00045407	-1.16731
	SA	TSV	THL	-0.00243944	0.000463242	-5.26601
	SA	TIW	THL	-0.00120787	0.000394489	-3.06184
	SLK	TSV	THL	-0.00282986	0.000275041	-10.2889
	SLK	TIW	THL	-0.00178266	0.000216292	-8.24191
	BB	SLK	TIW	-0.00103188	0.000231678	-4.45394
	BB	GC	TIW	-0.000348069	8.10E-05	-4.29751
	BB	SA	TIW	-0.00114947	0.000284513	-4.04014
	BB	KE	TIW	-0.00110178	0.000285632	-3.85735
	BB	JBG	TIW	-0.000172703	6.49E-05	-2.66305
	BB	NKB	TIW	-0.000205635	0.000151572	-1.35668
	BB	THL	TIW	-9.28E-05	0.000165673	-0.559971
	CMVN	JBG	TIW	-0.000181384	6.11E-05	-2.96779
	CMVN	GC	TIW	-8.78E-05	7.01E-05	-1.25345
	EB	SLK	TIW	-0.00108103	0.0002458	-4.398

Appendix 3.5 (cont.)

	Source 1	Source 2	Target	The *f3*-statistics	The standard error	The Z-score
f3-statistics were significantly positive (Z-score < 0)	EB	GC	TIW	-0.000356989	8.53E-05	-4.18558
	EB	SA	TIW	-0.00109904	0.000308808	-3.55897
	EB	KE	TIW	-0.00108939	0.000310304	-3.51072
	EB	JBG	TIW	-0.00020332	6.92E-05	-2.93996
	EB	NKB	TIW	-0.000312739	0.00015998	-1.95486
	EB	THL	TIW	-0.000114653	0.000172768	-0.663624
	FJ	GC	TIW	-0.000457474	0.000103467	-4.42145
	FJ	JBG	TIW	-0.000242806	8.86E-05	-2.73957
	FJ	SLK	TIW	-0.000561342	0.000356227	-1.5758
	FJ	SA	TIW	-0.000496943	0.0004736	-1.04929
	FJ	KE	TIW	-0.000495333	0.000473867	-1.0453
	GC	TSV	TIW	-0.000328841	8.40E-05	-3.91665
	GC	PHI	TIW	-0.000125957	9.11E-05	-1.38293
	GC	KE	TIW	-0.000177819	0.000192778	-0.922403
	GC	INDO	TIW	-6.40E-05	7.20E-05	-0.888838
	GC	SA	TIW	-0.000153946	0.000192242	-0.80079
	GC	NTVN	TIW	-5.25E-05	7.95E-05	-0.660566
	GC	THL	TIW	-4.71E-05	8.79E-05	-0.536274
	GC	SLK	TIW	-2.58E-05	0.000136453	-0.188775
	INDO	JBG	TIW	-9.09E-05	6.39E-05	-1.42347
	INDO	TSV	TIW	-1.64E-05	0.000153045	-0.107094
	JBG	TSV	TIW	-0.000238136	6.74E-05	-3.53323
	JBG	NTVN	TIW	-0.000138139	6.96E-05	-1.98588
	JBG	THL	TIW	-0.000153439	8.29E-05	-1.85013
	JBG	SLK	TIW	-0.000180525	0.000127464	-1.41628
	JBG	KE	TIW	-0.000228166	0.000180783	-1.2621
	JBG	SA	TIW	-0.000226521	0.000180401	-1.25566
	JBG	PHI	TIW	-7.99E-05	7.88E-05	-1.01425
	KE	TSV	TIW	-0.00137895	0.000308396	-4.47135
	NKB	TSV	TIW	-0.000367688	0.000160708	-2.28792
	NKB	PHI	TIW	-5.40E-05	0.00018225	-0.296437
	SA	TSV	TIW	-0.00138698	0.000307349	-4.51272
	SLK	TSV	TIW	-0.00120261	0.000246175	-4.88517
	THL	TSV	TIW	-0.000155407	0.00017182	-0.904475

Appendix 3.6 Characterization of high-quality BLAST matches of 663 outlier SNPs to *Penaeus monodon* transcriptome contigs

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_354	contig9198.1	Serine threonine- kinase D3|GO:0016020 GO:0000166 GO:0004674 GO:0043167 GO:0016310	2.01E-29	100	100.00%
2019PM_354	contig9272.1	serine threonine- kinase D3 isoform X1|GO:0016020 GO:0000166 GO:0004674 GO:0043167 GO:0016310	2.01E-29	100	100.00%
2019PM_7022	contig200831.1	---NA---	2.01E-29	100	100.00%
2019PM_4058	contig92766.1	AChain Apocrustacyanin C1 Crystals|GO:0005615 GO:0031409 GO:0036094 GO:0006810	2.01E-29	100	100.00%
2019PM_4058	contig97472.1	AChain Apocrustacyanin C1 Crystals|GO:0005615 GO:0031409 GO:0036094 GO:0006810	2.01E-29	100	100.00%
2019PM_4058	contig86923.1	AChain Apocrustacyanin C1 Crystals|GO:0005615 GO:0031409 GO:0036094 GO:0006810	2.01E-29	100	100.00%
2019PM_4058	contig94787.1	AChain Apocrustacyanin C1 Crystals|GO:0005615 GO:0031409 GO:0036094 GO:0006810	2.01E-29	100	100.00%
2019PM_4058	contig147401.1	AChain Apocrustacyanin C1 Crystals|GO:0005615 GO:0031409 GO:0036094 GO:0006810	2.01E-29	100	100.00%
2019PM_10536	contig2616.1	Neuropilin and tolloid	2.01E-29	100	100.00%
2019PM_6523	contig42167.1	C-type lectin	2.01E-29	100	100.00%
2019PM_6523	contig48544.1	---NA---	2.01E-29	100	100.00%
2019PM_6523	contig50319.1	---NA---	2.01E-29	100	100.00%
2019PM_7307	contig29609.1	xylose isomerase-like|GO:0005737 GO:0016020 GO:0044238 GO:0071704	2.01E-29	100	100.00%
2019PM_7307	contig8599.1	xylose isomerase-like|GO:0005737 GO:0016020 GO:0044238 GO:0071704	2.01E-29	100	100.00%
2019PM_1019	contig54930.1	rna-directed dna polymerase from mobile element jockey	2.01E-29	100	100.00%
2019PM_1019	contig64187.1	rna-directed dna polymerase from mobile element	2.01E-29	100	100.00%
2019PM_7468	contig1470.1	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	2.01E-29	100	100.00%
2019PM_7468	contig1470.2	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	2.01E-29	100	100.00%
2019PM_7468	contig1729.1	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	2.01E-29	100	100.00%
2019PM_7468	contig1729.2	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	2.01E-29	100	100.00%
2019PM_7468	contig2018.1	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	2.01E-29	100	100.00%
2019PM_7468	contig2018.2	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	2.01E-29	100	100.00%
2019PM_7468	contig2269.1	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	2.01E-29	100	100.00%
2019PM_7067	contig9015.1	---NA---	2.01E-29	100	100.00%
2019PM_7067	contig13398.1	PREDICTED: uncharacterized protein LOC108669519 isoform X1	2.01E-29	100	100.00%
2019PM_7067	contig13398.2	---NA---	2.01E-29	100	100.00%
2019PM_7067	contig32911.1	PREDICTED: uncharacterized protein LOC108669519 isoform X2	2.01E-29	100	100.00%
2019PM_9268	contig33948.1	---NA---	2.01E-29	100	100.00%
2019PM_4727	contig117214.1	zinc C2H2|GO:0003676 GO:0046872	2.01E-29	100	100.00%
2019PM_941	contig26548.2	ligand of Numb X 2-like|GO:0008270	2.01E-29	100	100.00%
2019PM_941	contig26548.3	ligand of Numb X 2-like	2.01E-29	100	100.00%
2019PM_941	contig28846.1	ligand of Numb X 2-like	2.01E-29	100	100.00%
2019PM_941	contig49350.1	ligand of Numb X 2-like	2.01E-29	100	100.00%
2019PM_469	contig5079.1	PREDICTED: uncharacterized protein LOC108679759	2.01E-29	100	100.00%
2019PM_7837	contig170033.1	---NA---	2.01E-29	100	100.00%
2019PM_9964	contig3323.1	extracellular sulfatase SULF-1 homolog isoform X1|GO:0008449 GO:0008152	2.01E-29	100	100.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_6229	contig75648.1	---NA---	2.01E-29	100	100.00%
2019PM_6229	contig82925.1	---NA---	2.01E-29	100	100.00%
2019PM_9056	contig7456.1	cell adhesion molecule 3	2.01E-29	100	100.00%
2019PM_9056	contig16669.1	cell adhesion molecule 3	2.01E-29	100	100.00%
2019PM_9056	contig15435.1	cell adhesion molecule 3	2.01E-29	100	100.00%
2019PM_9056	contig17054.1	tyrosine- kinase receptor TYRO3-like isoform X3	2.01E-29	100	100.00%
2019PM_9056	contig21634.1	cell adhesion molecule 3	2.01E-29	100	100.00%
2019PM_9056	contig39363.1	cell adhesion molecule 3	2.01E-29	100	100.00%
2019PM_6944	contig51401.1	PREDICTED: uncharacterized protein LOC108680202	2.01E-29	100	100.00%
2019PM_6944	contig59928.1	PREDICTED: uncharacterized protein LOC108680202	2.01E-29	100	100.00%
2019PM_6944	contig83041.1	PREDICTED: uncharacterized protein LOC108680202	2.01E-29	100	100.00%
2019PM_10398	contig134349.1	---NA---	2.01E-29	100	100.00%
2019PM_3043	contig29545.1	chitinase 1|GO:0005576 GO:0004568 GO:0008061 GO:0006032	2.01E-29	100	100.00%
2019PM_3144	contig3945.1	PREDICTED: uncharacterized protein LOC108669467	2.01E-29	100	100.00%
2019PM_3144	contig3945.2	PREDICTED: uncharacterized protein LOC108669467	2.01E-29	100	100.00%
2019PM_1623	contig93906.1	---NA---	2.01E-29	100	100.00%
2019PM_3050	contig2672.1	isoforms A C F G H-like isoform X7	2.01E-29	100	100.00%
2019PM_574	contig188.1	bark beetle isoform X1	2.01E-29	100	100.00%
2019PM_574	contig188.2	bark beetle isoform X1	2.01E-29	100	100.00%
2019PM_574	contig4754.1	bark beetle isoform X1	2.01E-29	100	100.00%
2019PM_4073	contig50488.1	---NA---	2.01E-29	100	100.00%
2019PM_3913	contig15321.1	UDP-glucuronosyltransferase 2B14-like	2.01E-29	100	100.00%
2019PM_3913	contig15708.1	UDP-glucuronosyltransferase 2B14-like	2.01E-29	100	100.00%
2019PM_3913	contig15708.2	UDP-glucuronosyltransferase 2B14-like	2.01E-29	100	100.00%
2019PM_5152	contig34473.1	transcriptional regulatory -like	2.01E-29	100	100.00%
2019PM_5152	contig41161.1	---NA---	2.01E-29	100	100.00%
2019PM_354	contig9198.1	Serine threonine- kinase D3|GO:0016020 GO:0000166 GO:0004674 GO:0043167 GO:0016310	9.33E-28	100	98.60%
2019PM_354	contig9272.1	serine threonine- kinase D3 isoform X1|GO:0016020 GO:0000166 GO:0004674 GO:0043167 GO:0016310	9.33E-28	100	98.60%
2019PM_7022	contig200831.1	---NA---	9.33E-28	100	98.60%
2019PM_4058	contig86923.1		9.33E-28	100	98.60%
2019PM_4058	contig94787.1		9.33E-28	100	98.60%
2019PM_4058	contig147401.1		9.33E-28	100	98.60%
2019PM_4058	contig92766.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615	9.33E-28	100	98.60%
2019PM_4058	contig97472.1	GO:0031409 GO:0036094 GO:0006810	9.33E-28	100	98.60%
2019PM_10536	contig2616.1	Neuropilin and tolloid	9.33E-28	100	98.60%
2019PM_6523	contig42167.1	C-type lectin	9.33E-28	100	98.60%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_10536	contig2616.1	Neuropilin and tolloid	9.33E-28	100	98.60%
2019PM_6523	contig42167.1	C-type lectin	9.33E-28	100	98.60%
2019PM_6523	contig48544.1	---NA---	9.33E-28	100	98.60%
2019PM_6523	contig50319.1	---NA---	9.33E-28	100	98.60%
2019PM_7307	contig8599.1	xylose isomerase-like|GO:0005737 GO:0016020 GO:0044238 GO:0071704	9.33E-28	100	98.60%
2019PM_7307	contig29609.1	xylose isomerase-like|GO:0005737 GO:0016020 GO:0044238 GO:0071704	9.33E-28	100	98.60%
2019PM_1019	contig54930.1	rna-directed dna polymerase from mobile element jockey	9.33E-28	100	98.60%
2019PM_1019	contig64187.1	rna-directed dna polymerase from mobile element	9.33E-28	100	98.60%
2019PM_7468	contig1470.1	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	9.33E-28	100	98.60%
2019PM_7468	contig1470.2	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	9.33E-28	100	98.60%
2019PM_7468	contig1729.1	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	9.33E-28	100	98.60%
2019PM_7468	contig1729.2	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	9.33E-28	100	98.60%
2019PM_7468	contig2018.1	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	9.33E-28	100	98.60%
2019PM_7468	contig2018.2	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	9.33E-28	100	98.60%
2019PM_7468	contig2269.1	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	9.33E-28	100	98.60%
2019PM_9268	contig33948.1	---NA---	9.33E-28	100	98.60%
2019PM_4727	contig117214.1	zinc C2H2|GO:0003676 GO:0046872	9.33E-28	100	98.60%
2019PM_941	contig26548.2	ligand of Numb X 2-like|GO:0008270	9.33E-28	100	98.60%
2019PM_941	contig26548.3	ligand of Numb X 2-like	9.33E-28	100	98.60%
2019PM_941	contig28846.1	ligand of Numb X 2-like	9.33E-28	100	98.60%
2019PM_941	contig49350.1	ligand of Numb X 2-like	9.33E-28	100	98.60%
2019PM_469	contig5079.1	PREDICTED: uncharacterized protein LOC108679759	9.33E-28	100	98.60%
2019PM_7837	contig170033.1	---NA---	9.33E-28	100	98.60%
2019PM_9964	contig3323.1	extracellular sulfatase SULF-1 homolog isoform X1|GO:0008449 GO:0008152	9.33E-28	100	98.60%
2019PM_6229	contig75648.1	---NA---	9.33E-28	100	98.60%
2019PM_6229	contig82925.1	---NA---	9.33E-28	100	98.60%
2019PM_9056	contig15435.1	cell adhesion molecule 3	9.33E-28	100	98.60%
2019PM_9056	contig17054.1	tyrosine- kinase receptor TYRO3-like isoform X3	9.33E-28	100	98.60%
2019PM_9056	contig21634.1	cell adhesion molecule 3	9.33E-28	100	98.60%
2019PM_9056	contig39363.1	cell adhesion molecule 3	9.33E-28	100	98.60%
2019PM_9056	contig7456.1	cell adhesion molecule 3	9.33E-28	100	98.60%
2019PM_9056	contig16669.1	cell adhesion molecule 3	9.33E-28	100	98.60%
2019PM_8238	contig33915.1	diacyglycerol O-acyltransferase MT1809|GO:0005886 GO:0004144 GO:0047196 GO:0019432	9.33E-28	100	98.60%
2019PM_6944	contig51401.1	PREDICTED: uncharacterized protein LOC108680202	9.33E-28	100	98.60%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_6944	contig59928.1	PREDICTED: uncharacterized protein LOC108680202	9.33E-28	100	98.60%
2019PM_6944	contig83041.1	PREDICTED: uncharacterized protein LOC108680202	9.33E-28	100	98.60%
2019PM_3043	contig29545.1	chitinase 1|GO:0005576 GO:0004568 GO:0008061 GO:0006032	9.33E-28	100	98.60%
2019PM_3144	contig3945.1	PREDICTED: uncharacterized protein LOC108669467	9.33E-28	100	98.60%
2019PM_3144	contig3945.2	PREDICTED: uncharacterized protein LOC108669467	9.33E-28	100	98.60%
2019PM_1623	contig93906.1	---NA---	9.33E-28	100	98.60%
2019PM_3050	contig2672.1	isoforms A C F G H-like isoform X7	9.33E-28	100	98.60%
2019PM_574	contig188.1	bark beetle isoform X1	9.33E-28	100	98.60%
2019PM_574	contig188.2	bark beetle isoform X1	9.33E-28	100	98.60%
2019PM_574	contig4754.1	bark beetle isoform X1	9.33E-28	100	98.60%
2019PM_4073	contig50488.1	---NA---	9.33E-28	100	98.60%
2019PM_3913	contig15321.1	UDP-glucuronosyltransferase 2B14-like	9.33E-28	100	98.60%
2019PM_3913	contig15708.1	UDP-glucuronosyltransferase 2B14-like	9.33E-28	100	98.60%
2019PM_3913	contig15708.2	UDP-glucuronosyltransferase 2B14-like	9.33E-28	100	98.60%
2019PM_5152	contig34473.1	transcriptional regulatory -like	9.33E-28	100	98.60%
2019PM_5152	contig41161.1	---NA---	9.33E-28	100	98.60%
2019PM_4058	contig75561.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615 GO:0031409 GO:0036094 GO:0006810	4.34E-26	100	97.10%
2019PM_7307	contig10918.1	xylose isomerase-like|GO:0044238 GO:0071704	4.34E-26	100	97.10%
2019PM_7307	contig16991.1	xylose isomerase-like|GO:0005737 GO:0016020 GO:0044238 GO:0071704	4.34E-26	100	97.10%
2019PM_7307	contig18179.1	xylose isomerase-like|GO:0005737 GO:0016020 GO:0044238 GO:0071704	4.34E-26	100	97.10%
2019PM_7307	contig18179.2	xylose isomerase-like|GO:0044238 GO:0071704	4.34E-26	100	97.10%
2019PM_7307	contig29609.2	xylose isomerase-like|GO:0005737 GO:0016020 GO:0044238 GO:0071704	4.34E-26	100	97.10%
2019PM_10398	contig134349.1	---NA---	4.34E-26	100	97.10%
2019PM_5570	contig15817.1	esterase FE4	2.59E-28	97.1	100.00%
2019PM_7307	contig10918.1	xylose isomerase-like|GO:0044238 GO:0071704	2.02E-24	100	95.70%
2019PM_7307	contig16991.1	xylose isomerase-like|GO:0005737 GO:0016020 GO:0044238 GO:0071704	2.02E-24	100	95.70%
2019PM_7307	contig18179.1	xylose isomerase-like|GO:0005737 GO:0016020 GO:0044238 GO:0071704	2.02E-24	100	95.70%
2019PM_7307	contig18179.2	xylose isomerase-like|GO:0044238 GO:0071704	2.02E-24	100	95.70%
2019PM_7307	contig29609.2	xylose isomerase-like|GO:0005737 GO:0016020 GO:0044238 GO:0071704	2.02E-24	100	95.70%
2019PM_10046	contig17009.1	endonuclease-reverse transcriptase	2.02E-24	100	95.70%
2019PM_10046	contig17325.1	endonuclease-reverse transcriptase	2.02E-24	100	95.70%
2019PM_10046	contig17737.1	endonuclease-reverse transcriptase	2.02E-24	100	95.70%
2019PM_10046	contig19250.1	endonuclease-reverse transcriptase	2.02E-24	100	95.70%
2019PM_10046	contig33987.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	2.02E-24	100	95.70%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_10046	contig59618.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	2.02E-24	100	95.70%
2019PM_10046	contig62730.1	endonuclease-reverse transcriptase	2.02E-24	100	95.70%
2019PM_10046	contig66320.1	endonuclease-reverse transcriptase	2.02E-24	100	95.70%
2019PM_4058	contig75561.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615 GO:0031409 GO:0036094 GO:0006810	2.02E-24	100	95.70%
2019PM_7067	contig9015.1	---NA---	2.02E-24	100	95.70%
2019PM_7067	contig13398.1	PREDICTED: uncharacterized protein LOC108669519 isoform X1	2.02E-24	100	95.70%
2019PM_7067	contig13398.2	---NA---	2.02E-24	100	95.70%
2019PM_7067	contig32911.1	PREDICTED: uncharacterized protein LOC108669519 isoform X2	2.02E-24	100	95.70%
2019PM_3780	contig154245.1	---NA---	2.02E-24	100	95.70%
2019PM_9059	contig20496.1	hypothetical protein g.43282	9.33E-28	95.65	100.00%
2019PM_2406	contig32222.1	quiver-like	1.21E-26	97.1	98.50%
2019PM_8183	contig69798.1	Zinc finger MYND domain-containing 19|GO:0046872	9.33E-28	95.65	100.00%
2019PM_2716	contig22424.2	---NA---	3.36E-27	94.2	100.00%
2019PM_2716	contig22424.1	---NA---	3.36E-27	94.2	100.00%
2019PM_2716	contig58256.1	---NA---	3.36E-27	94.2	100.00%
2019PM_9059	contig20496.1	hypothetical protein g.43282	4.34E-26	95.65	98.50%
2019PM_8238	contig33915.1	diacyglycerol O-acyltransferase MT1809|GO:0005886 GO:0004144 GO:0047196 GO:0019432	3.36E-27	94.2	100.00%
2019PM_2406	contig32222.1	quiver-like	5.62E-25	97.1	97.00%
2019PM_5570	contig15817.1	esterase FE4	5.62E-25	97.1	97.00%
2019PM_8183	contig69798.1	Zinc finger MYND domain-containing 19|GO:0046872	4.34E-26	95.65	98.50%
2019PM_3780	contig154245.1	---NA---	9.40E-23	100	94.20%
2019PM_10046	contig8184.1	endonuclease-reverse transcriptase	9.40E-23	100	94.30%
2019PM_10046	contig59564.1	endonuclease-reverse transcriptase	9.40E-23	100	94.30%
2019PM_10046	contig17009.1	endonuclease-reverse transcriptase	9.40E-23	100	94.30%
2019PM_10046	contig17325.1	endonuclease-reverse transcriptase	9.40E-23	100	94.30%
2019PM_10046	contig17737.1	endonuclease-reverse transcriptase	9.40E-23	100	94.30%
2019PM_10046	contig19250.1	endonuclease-reverse transcriptase	9.40E-23	100	94.30%
2019PM_10046	contig33987.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	9.40E-23	100	94.30%
2019PM_10046	contig59618.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	9.40E-23	100	94.30%
2019PM_10046	contig62730.1	endonuclease-reverse transcriptase	9.40E-23	100	94.30%
2019PM_10046	contig66320.1	endonuclease-reverse transcriptase	9.40E-23	100	94.30%
2019PM_2434	contig33002.1	zinc finger 423-like|GO:0005488	1.21E-26	92.75	100.00%
2019PM_2716	contig22424.1	---NA---	1.56E-25	94.2	98.50%
2019PM_2716	contig58256.1	---NA---	1.56E-25	94.2	98.50%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_2716	contig22424.2	---NA---	1.56E-25	94.2	98.50%
2019PM_3453	contig1494.1	protocadherin-like wing polarity stan isoform X1	5.62E-25	94.2	98.50%
2019PM_3453	contig3495.1	protocadherin-like wing polarity stan isoform X1|GO:0016020 GO:0004871 GO:0007165	5.62E-25	94.2	98.50%
2019PM_9768	contig8082.1	lian-aa1 retrotransposon	3.38E-22	100	94.20%
2019PM_9768	contig73604.1	rna-directed dna polymerase from mobile element jockey	3.38E-22	100	94.20%
2019PM_6570	contig2127.1	RNA-directed DNA polymerase from mobile element jockey-like	2.61E-23	97.1	95.50%
2019PM_3313	contig46639.1	PREDICTED: uncharacterized protein LOC108667351	4.34E-26	91.3	100.00%
2019PM_3313	contig74145.1	PREDICTED: uncharacterized protein LOC108667351	4.34E-26	91.3	100.00%
2019PM_3530	contig9752.1	endonuclease-reverse transcriptase	5.62E-25	92.75	98.40%
2019PM_3530	contig55304.1	endonuclease-reverse transcriptase|GO:0003964 GO:0004519 GO:0006278 GO:0090305	5.62E-25	92.75	98.40%
2019PM_2434	contig33002.1	zinc finger 423-like|GO:0005488	5.62E-25	92.75	98.40%
2019PM_3453	contig1494.1	protocadherin-like wing polarity stan isoform X1	2.61E-23	94.2	96.90%
2019PM_3453	contig3495.1	protocadherin-like wing polarity stan isoform X1|GO:0016020 GO:0004871 GO:0007165	2.61E-23	94.2	96.90%
2019PM_3313	contig46639.1	PREDICTED: uncharacterized protein LOC108667351	2.02E-24	91.3	98.40%
2019PM_3313	contig74145.1	PREDICTED: uncharacterized protein LOC108667351	2.02E-24	91.3	98.40%
2019PM_6519	contig6427.1	PR domain zinc finger 1 isoform X1|GO:0003676 GO:0046872	2.02E-24	91.3	98.40%
2019PM_6519	contig6427.2	PR domain zinc finger 1 isoform X1|GO:0003676 GO:0046872	2.02E-24	91.3	98.40%
2019PM_3530	contig9752.1	endonuclease-reverse transcriptase	2.61E-23	92.75	96.90%
2019PM_3530	contig55304.1	endonuclease-reverse transcriptase|GO:0003964 GO:0004519 GO:0006278 GO:0090305	2.61E-23	92.75	96.90%
2019PM_2960	contig84151.1	---NA---	5.62E-25	88.41	100.00%
2019PM_2960	contig31454.1	Carboxyl-ester lipase	5.62E-25	88.41	100.00%
2019PM_9913	contig3257.1	Tissue factor pathway inhibitor|GO:0004867 GO:0010951	5.62E-25	88.41	100.00%
2019PM_9913	contig3257.3	Tissue factor pathway inhibitor|GO:0004867 GO:0010951	5.62E-25	88.41	100.00%
2019PM_9913	contig3260.2	Tissue factor pathway inhibitor|GO:0004867 GO:0010951	5.62E-25	88.41	100.00%
2019PM_9913	contig4639.1	Tissue factor pathway inhibitor|GO:0004867 GO:0010951	5.62E-25	88.41	100.00%
2019PM_9913	contig7219.1	pacifastin light chain-like serine ase inhibitor|GO:0004867 GO:0010951	5.62E-25	88.41	100.00%
2019PM_9913	contig7219.3	serine protease inhibitor i ii|GO:0004867 GO:0010951	5.62E-25	88.41	100.00%
2019PM_9913	contig9774.1	serine protease inhibitor i ii|GO:0004867 GO:0010951	5.62E-25	88.41	100.00%
2019PM_9913	contig9774.2	pacifastin light chain-like serine ase inhibitor|GO:0004867 GO:0010951	5.62E-25	88.41	100.00%
2019PM_3494	contig52331.1	---NA---	5.62E-25	88.41	100.00%
2019PM_3494	contig85168.1	---NA---	5.62E-25	88.41	100.00%
2019PM_3494	contig96126.1	---NA---	5.62E-25	88.41	100.00%
2019PM_5207	contig40533.1	macrophage mannose receptor 1-like	5.62E-25	88.41	100.00%
2019PM_5207	contig58541.1	macrophage mannose receptor 1-like	5.62E-25	88.41	100.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_3530	contig66188.1	endonuclease-reverse transcriptase	5.62E-25	88.41	100.00%
2019PM_1947	contig138606.1	Zinc finger	5.62E-25	88.41	100.00%
2019PM_8186	contig184950.1	---NA---	7.27E-24	89.86	98.40%
2019PM_10046	contig42824.1	endonuclease-reverse transcriptase	3.38E-22	94.2	95.50%
2019PM_6519	contig6427.1	PR domain zinc finger 1 isoform X1|GO:0003676 GO:0046872	9.40E-23	91.3	96.80%
2019PM_6519	contig6427.2	PR domain zinc finger 1 isoform X1|GO:0003676 GO:0046872	9.40E-23	91.3	96.80%
2019PM_9913	contig3260.1	Tissue factor pathway inhibitor|GO:0004867 GO:0010951	2.02E-24	86.96	100.00%
2019PM_4096	contig23637.1	spatzle 6 precursor	2.02E-24	86.96	100.00%
2019PM_2960	contig31454.1	Carboxyl-ester lipase	2.61E-23	88.41	98.40%
2019PM_2960	contig84151.1	---NA---	2.61E-23	88.41	98.40%
2019PM_9913	contig3257.1	Tissue factor pathway inhibitor|GO:0004867 GO:0010951	2.61E-23	88.41	98.40%
2019PM_9913	contig3257.2	Tissue factor pathway inhibitor|GO:0004867 GO:0010951	2.61E-23	88.41	98.40%
2019PM_9913	contig3257.3	Tissue factor pathway inhibitor|GO:0004867 GO:0010951	2.61E-23	88.41	98.40%
2019PM_9913	contig3260.2	Tissue factor pathway inhibitor|GO:0004867 GO:0010951	2.61E-23	88.41	98.40%
2019PM_9913	contig4639.1	Tissue factor pathway inhibitor|GO:0004867 GO:0010951	2.61E-23	88.41	98.40%
2019PM_9913	contig7219.1	pacifastin light chain-like serine ase inhibitor|GO:0004867 GO:0010951	2.61E-23	88.41	98.40%
2019PM_9913	contig7219.2	pacifastin light chain-like serine ase inhibitor|GO:0004867 GO:0010951	2.61E-23	88.41	98.40%
2019PM_9913	contig7219.3	serine protease inhibitor i ii|GO:0004867 GO:0010951	2.61E-23	88.41	98.40%
2019PM_9913	contig9774.1	serine protease inhibitor i ii|GO:0004867 GO:0010951	2.61E-23	88.41	98.40%
2019PM_9913	contig9774.2	pacifastin light chain-like serine ase inhibitor|GO:0004867 GO:0010951	2.61E-23	88.41	98.40%
2019PM_3494	contig12013.2	---NA---	2.61E-23	88.41	98.40%
2019PM_3494	contig166844.1	---NA---	2.61E-23	88.41	98.40%
2019PM_3494	contig212688.1	---NA---	2.61E-23	88.41	98.40%
2019PM_3494	contig52331.1	---NA---	2.61E-23	88.41	98.40%
2019PM_3494	contig85168.1	---NA---	2.61E-23	88.41	98.40%
2019PM_3494	contig96126.1	---NA---	2.61E-23	88.41	98.40%
2019PM_3530	contig66188.1	endonuclease-reverse transcriptase	2.61E-23	88.41	98.40%
2019PM_1947	contig138606.1	Zinc finger	2.61E-23	88.41	98.40%
2019PM_7022	contig50860.1	---NA---	1.22E-21	97.1	94.00%
2019PM_7022	contig52616.1	---NA---	1.22E-21	97.1	94.00%
2019PM_7022	contig55134.1	---NA---	1.22E-21	97.1	94.00%
2019PM_7022	contig127406.1	---NA---	1.22E-21	97.1	94.00%
2019PM_4058	contig96167.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615 GO:0031409 GO:0036094 GO:0006810	1.22E-21	97.1	94.00%
2019PM_4058	contig97213.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615 GO:0031409 GO:0036094 GO:0006810	1.22E-21	97.1	94.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_5354	contig51325.1	---NA---	1.22E-21	97.1	94.00%
2019PM_5354	contig89221.1	---NA---	1.22E-21	97.1	94.00%
2019PM_5354	contig102878.1	---NA---	1.22E-21	97.1	94.00%
2019PM_6488	contig41915.1	---NA---	3.38E-22	89.86	96.80%
2019PM_6488	contig46619.1	serine ase inhibitor	3.38E-22	89.86	96.80%
2019PM_6488	contig56254.1	---NA---	3.38E-22	89.86	96.80%
2019PM_6488	contig159278.1	---NA---	3.38E-22	89.86	96.80%
2019PM_6570	contig12110.1	---NA---	1.22E-21	97.1	94.00%
2019PM_6570	contig57522.1	---NA---	1.22E-21	97.1	94.00%
2019PM_8186	contig184950.1	---NA---	3.38E-22	89.86	96.80%
2019PM_9913	contig3260.1	Tissue factor pathway inhibitor|GO:0004867 GO:0010951	9.40E-23	86.96	98.30%
2019PM_4096	contig23637.1	spatzle 6 precursor	9.40E-23	86.96	98.30%
2019PM_5203	contig58636.1	---NA---	2.61E-23	84.06	100.00%
2019PM_9945	contig25935.1	---NA---	2.61E-23	84.06	100.00%
2019PM_9945	contig14415.1	---NA---	2.61E-23	84.06	100.00%
2019PM_9945	contig38870.1	---NA---	2.61E-23	84.06	100.00%
2019PM_10160	contig179887.1	---NA---	9.40E-23	82.61	100.00%
2019PM_5372	contig136133.1	---NA---	1.22E-21	92.75	95.30%
2019PM_5372	contig162684.1	---NA---	1.22E-21	92.75	95.30%
2019PM_5372	contig164859.1	---NA---	1.22E-21	92.75	95.30%
2019PM_9694	contig82930.1	---NA---	3.38E-22	81.16	100.00%
2019PM_9694	contig82930.1	---NA---	3.38E-22	81.16	100.00%
2019PM_7891	contig5907.1	DNA-directed RNA polymerase I subunit RPA2|GO:0016740	3.38E-22	81.16	100.00%
2019PM_9913	contig3257.2	Tissue factor pathway inhibitor|GO:0004867 GO:0010951	1.22E-21	88.41	96.70%
2019PM_9913	contig7219.2	pacifastin light chain-like serine ase inhibitor|GO:0004867 GO:0010951	1.22E-21	88.41	96.70%
2019PM_3494	contig2208.1	TANC2 isoform X2	1.22E-21	88.41	96.70%
2019PM_3494	contig12013.1	craniofacial development 2-like	1.22E-21	88.41	96.70%
2019PM_3494	contig64855.1	---NA---	1.22E-21	88.41	96.70%
2019PM_3494	contig68760.1	---NA---	1.22E-21	88.41	96.70%
2019PM_3494	contig128553.1	---NA---	1.22E-21	88.41	96.70%
2019PM_3494	contig163029.1	---NA---	1.22E-21	88.41	96.70%
2019PM_3494	contig12013.2	---NA---	1.22E-21	88.41	96.70%
2019PM_3494	contig166844.1	---NA---	1.22E-21	88.41	96.70%
2019PM_3494	contig212688.1	---NA---	1.22E-21	88.41	96.70%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_5207	contig40533.1	macrophage mannose receptor 1-like	1.22E-21	88.41	96.70%
2019PM_5207	contig58541.1	macrophage mannose receptor 1-like	1.22E-21	88.41	96.70%
2019PM_5203	contig58636.1	---NA---	1.22E-21	84.06	98.30%
2019PM_9945	contig26284.1	---NA---	1.22E-21	84.06	98.30%
2019PM_9945	contig26284.1	---NA---	1.22E-21	84.06	98.30%
2019PM_10046	contig56083.1	endonuclease-reverse transcriptase	4.37E-21	100	92.90%
2019PM_10046	contig8184.1	endonuclease-reverse transcriptase	4.37E-21	100	92.90%
2019PM_10046	contig59564.1	endonuclease-reverse transcriptase	4.37E-21	100	92.90%
2019PM_4058	contig132001.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615 GO:0031409 GO:0036094 GO:0006810	4.37E-21	100	92.80%
2019PM_10046	contig30213.1	endonuclease-reverse transcriptase	4.37E-21	95.65	94.00%
2019PM_4058	contig124618.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615 GO:0031409 GO:0036094 GO:0006810	4.37E-21	91.3	95.20%
2019PM_6488	contig49772.1	---NA---	4.37E-21	86.96	96.70%
2019PM_6638	contig8010.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig8588.1	craniofacial development 2-like|GO:0003824 GO:0090304	4.37E-21	86.96	96.70%
2019PM_6638	contig9352.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig10338.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig11118.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig11122.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig11294.1	mitochondrial inner membrane protease ATP23 homolog|GO:0003964 GO:0004222 GO:0006278 GO:0006508	4.37E-21	86.96	96.70%
2019PM_6638	contig11294.2	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig12697.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig13130.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig13790.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig13854.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig14754.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig19357.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig19357.2	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig21122.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig30774.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig36317.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig36423.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig39402.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig43349.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig82623.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_6638	contig83288.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig91868.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_6638	contig168627.1	LINE-1 reverse transcriptase|GO:0003964 GO:0006278	4.37E-21	86.96	96.70%
2019PM_10160	contig179887.1	---NA---	4.37E-21	82.61	98.20%
2019PM_3494	contig9387.1	Craniofacial development 2|GO:0003964 GO:0006278	4.37E-21	78.26	100.00%
2019PM_3494	contig48217.1	---NA---	4.37E-21	78.26	100.00%
2019PM_3494	contig48217.2	---NA---	4.37E-21	78.26	100.00%
2019PM_3492	contig47106.1	glutamate receptor NMDA 2B-|GO:0016021 GO:0030054 GO:0045211 GO:0004970 GO:0005234 GO:0034220 GO:0035235	4.37E-21	78.26	100.00%
2019PM_3492	contig47236.1	ionotropic glutamate receptor NMDA 2|GO:0016021 GO:0030054 GO:0045211 GO:0004970 GO:0005234 GO:0034220 GO:0035235	4.37E-21	78.26	100.00%
2019PM_9922	contig21299.1	---NA---	4.37E-21	78.26	100.00%
2019PM_9768	contig2758.1	rna-directed dna polymerase from mobile element	1.57E-20	100	92.80%
2019PM_9768	contig4605.1	transcription elongation factor SPT4|GO:0005634 GO:0003746 GO:0008270 GO:0006351 GO:0006414 GO:0032786	1.57E-20	100	92.80%
2019PM_9768	contig4911.1		1.57E-20	100	92.80%
2019PM_9768	contig4911.2		1.57E-20	100	92.80%
2019PM_9768	contig4911.3		1.57E-20	100	92.80%
2019PM_9768	contig9497.1	RNA-directed DNA polymerase from mobile element jockey-like	1.57E-20	100	92.80%
2019PM_9768	contig18302.1	RNA-directed DNA polymerase from mobile element jockey-	1.57E-20	100	92.80%
2019PM_9768	contig23768.1	RNA-directed DNA polymerase from mobile element jockey-like	1.57E-20	100	92.80%
2019PM_9768	contig32468.1	RNA-directed DNA polymerase from mobile element jockey-	1.57E-20	100	92.80%
2019PM_9768	contig56529.1	RNA-directed DNA polymerase from mobile element jockey-like	1.57E-20	100	92.80%
2019PM_9768	contig56529.2	RNA-directed DNA polymerase from mobile element jockey-like	1.57E-20	100	92.80%
2019PM_9768	contig56963.1	RNA-directed DNA polymerase from mobile element jockey-like	1.57E-20	100	92.80%
2019PM_9768	contig90807.1	PREDICTED: uncharacterized protein LOC108681254	1.57E-20	100	92.80%
2019PM_9768	contig95778.1	---NA---	1.57E-20	100	92.80%
2019PM_9768	contig98017.1	RNA-directed DNA polymerase from mobile element jockey-like	1.57E-20	100	92.80%
2019PM_9768	contig121761.1	---NA---	1.57E-20	100	92.80%
2019PM_9768	contig196289.1	RNA-directed DNA polymerase from mobile element jockey-like	1.57E-20	100	92.80%
2019PM_6677	contig123960.1	---NA---	5.66E-20	100	91.50%
2019PM_6730	contig1983.1	rna-directed dna polymerase from mobile element jockey	5.66E-20	100	91.70%
2019PM_6730	contig2399.1	rna-directed dna polymerase from mobile element jockey	5.66E-20	100	91.70%
2019PM_6730	contig3954.1	rna-directed dna polymerase from mobile element jockey	5.66E-20	100	91.70%
2019PM_6730	contig11695.1	RNA-directed DNA polymerase from mobile element jockey-	5.66E-20	100	91.70%
2019PM_6730	contig13732.1	RNA-directed DNA polymerase from mobile element jockey-like	5.66E-20	100	91.70%
2019PM_6730	contig13732.2	RNA-directed DNA polymerase from mobile element jockey-like	5.66E-20	100	91.70%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_6730	contig21017.1	---NA---	5.66E-20	100	91.70%
2019PM_6730	contig31924.1	RNA-directed DNA polymerase from mobile element jockey-like	5.66E-20	100	91.70%
2019PM_6730	contig58135.1	---NA---	5.66E-20	100	91.70%
2019PM_4058	contig132001.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615	2.03E-19	100	91.30%
2019PM_4058	contig140097.1	GO:0031409 GO:0036094 GO:0006810	2.03E-19	100	91.30%
2019PM_10046	contig127944.1	endonuclease-reverse transcriptase	1.57E-20	98.55	92.80%
2019PM_10046	contig56083.1	endonuclease-reverse transcriptase	2.03E-19	100	91.40%
2019PM_7022	contig21017.1	---NA---	1.57E-20	98.55	92.60%
2019PM_7022	contig51877.1	---NA---	1.57E-20	98.55	92.60%
2019PM_6730	contig1983.1	rna-directed dna polymerase from mobile element jockey	2.63E-18	100	90.30%
2019PM_6730	contig2399.1	rna-directed dna polymerase from mobile element jockey	2.63E-18	100	90.30%
2019PM_6730	contig3954.1	rna-directed dna polymerase from mobile element jockey	2.63E-18	100	90.30%
2019PM_6730	contig11695.1	RNA-directed DNA polymerase from mobile element jockey-	2.63E-18	100	90.30%
2019PM_6730	contig13732.1	RNA-directed DNA polymerase from mobile element jockey-like	2.63E-18	100	90.30%
2019PM_6730	contig13732.2	RNA-directed DNA polymerase from mobile element jockey-like	2.63E-18	100	90.30%
2019PM_6730	contig21017.1	---NA---	2.63E-18	100	90.30%
2019PM_6730	contig31924.1	RNA-directed DNA polymerase from mobile element jockey-like	2.63E-18	100	90.30%
2019PM_6730	contig58135.1	---NA---	2.63E-18	100	90.30%
2019PM_10046	contig129537.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	1.22E-16	100	90.00%
2019PM_7022	contig64309.1	---NA---	7.32E-19	98.55	91.20%
2019PM_7022	contig192310.1	---NA---	7.32E-19	98.55	91.20%
2019PM_4058	contig140097.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615 GO:0031409 GO:0036094 GO:0006810	9.47E-18	100	89.90%
2019PM_3848	contig176756.1	---NA---	9.47E-18	100	89.90%
2019PM_3845	contig22490.1	reverse transcriptase	7.32E-19	98.55	91.20%
2019PM_3845	contig24482.1	reverse transcriptase	7.32E-19	98.55	91.20%
2019PM_3845	contig42030.1	reverse transcriptase	7.32E-19	98.55	91.20%
2019PM_3845	contig48529.1	reverse transcriptase	7.32E-19	98.55	91.20%
2019PM_10046	contig127944.1	endonuclease-reverse transcriptase	7.32E-19	98.55	91.30%
2019PM_9768	contig8125.1	rna-directed dna polymerase from mobile element jockey	3.40E-17	100	89.90%
2019PM_9768	contig74646.1	reverse transcriptase	3.40E-17	100	89.90%
2019PM_7022	contig66334.1	---NA---	5.66E-20	97.1	92.50%
2019PM_7022	contig93902.1	---NA---	5.66E-20	97.1	92.50%
2019PM_7022	contig106521.1	---NA---	2.03E-19	97.1	92.50%
2019PM_7022	contig144050.1	---NA---	2.03E-19	97.1	92.50%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_7022	contig153702.1	---NA---	2.03E-19	97.1	92.50%
2019PM_4058	contig96167.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615 GO:0031409 GO:0036094 GO:0006810	5.66E-20	97.1	92.50%
2019PM_4058	contig97213.1		5.66E-20	97.1	92.50%
2019PM_4058	contig62445.1		5.66E-20	97.1	92.50%
2019PM_4058	contig82201.1		5.66E-20	97.1	92.50%
2019PM_6677	contig123960.1	---NA---	1.22E-16	100	88.70%
2019PM_10046	contig132769.1	endonuclease-reverse transcriptase	4.40E-16	100	88.60%
2019PM_10046	contig129537.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	5.70E-15	100	88.60%
2019PM_3801	contig12885.2	---NA---	4.40E-16	100	88.40%
2019PM_3801	contig175635.1	---NA---	4.40E-16	100	88.40%
2019PM_3848	contig4474.1	reverse transcriptase	4.40E-16	100	88.40%
2019PM_3848	contig4884.1	cytochrome P450 9e2-like isoform X2|GO:0016020 GO:0043227	4.40E-16	100	88.40%
2019PM_3848	contig4884.3	reverse transcriptase	4.40E-16	100	88.40%
2019PM_3848	contig44358.1	PREDICTED: uncharacterized protein LOC105566822	4.40E-16	100	88.40%
2019PM_3848	contig66537.1	PREDICTED: uncharacterized protein LOC106467757	4.40E-16	100	88.40%
2019PM_7022	contig164168.1	---NA---	2.63E-18	97.1	91.20%
2019PM_7022	contig11529.1	---NA---	1.22E-16	98.55	89.70%
2019PM_4058	contig62445.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615 GO:0031409 GO:0036094 GO:0006810	2.63E-18	97.1	91.00%
2019PM_4058	contig82201.1		2.63E-18	97.1	91.00%
2019PM_6651	contig7298.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.63E-18	97.1	91.00%
2019PM_6651	contig9352.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.63E-18	97.1	91.00%
2019PM_6651	contig51447.1	craniofacial development 2-like|GO:0003964 GO:0006278	2.63E-18	97.1	91.00%
2019PM_6651	contig51447.2	reverse transcriptase, partial|GO:0003964 GO:0006278	2.63E-18	97.1	91.00%
2019PM_3845	contig4474.1	reverse transcriptase	3.40E-17	98.55	89.70%
2019PM_3845	contig5811.1	reverse transcriptase	3.40E-17	98.55	89.70%
2019PM_3845	contig5811.2	reverse transcriptase	3.40E-17	98.55	89.70%
2019PM_3845	contig15586.1	reverse transcriptase	3.40E-17	98.55	89.70%
2019PM_3845	contig17521.1	reverse transcriptase	3.40E-17	98.55	89.70%
2019PM_3845	contig30116.1	reverse transcriptase	3.40E-17	98.55	89.70%
2019PM_3845	contig30315.1	reverse transcriptase	3.40E-17	98.55	89.70%
2019PM_3845	contig56458.1	reverse transcriptase	3.40E-17	98.55	89.70%
2019PM_3845	contig22490.1	reverse transcriptase	3.40E-17	98.55	89.70%
2019PM_3845	contig24482.1	reverse transcriptase	3.40E-17	98.55	89.70%
2019PM_3845	contig42030.1	reverse transcriptase	3.40E-17	98.55	89.70%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_3845	contig48529.1	reverse transcriptase	3.40E-17	98.55	89.70%
2019PM_10046	contig42824.1	endonuclease-reverse transcriptase	1.57E-20	94.2	93.90%
2019PM_10046	contig30213.1	endonuclease-reverse transcriptase	2.03E-19	95.65	92.50%
2019PM_7022	contig76341.1	---NA---	1.57E-20	94.2	93.80%
2019PM_7022	contig76341.2	---NA---	1.57E-20	94.2	93.80%
2019PM_3754	contig155085.1	Nucleic-acid-binding from mobile element	2.03E-19	95.65	92.40%
2019PM_5354	contig51325.1	---NA---	1.57E-20	94.2	93.80%
2019PM_5354	contig89221.1	---NA---	1.57E-20	94.2	93.80%
2019PM_5354	contig102878.1	---NA---	1.57E-20	94.2	93.80%
2019PM_3845	contig4474.1	reverse transcriptase	1.58E-15	98.55	88.20%
2019PM_3845	contig5811.1	reverse transcriptase	1.58E-15	98.55	88.20%
2019PM_3845	contig5811.2	reverse transcriptase	1.58E-15	98.55	88.20%
2019PM_3845	contig15586.1	reverse transcriptase	1.58E-15	98.55	88.20%
2019PM_3845	contig17521.1	reverse transcriptase	1.58E-15	98.55	88.20%
2019PM_3845	contig30116.1	reverse transcriptase	1.58E-15	98.55	88.20%
2019PM_3845	contig30315.1	reverse transcriptase	1.58E-15	98.55	88.20%
2019PM_3845	contig56458.1	reverse transcriptase	1.58E-15	98.55	88.20%
2019PM_7022	contig2399.1	rna-directed dna polymerase from mobile element jockey	7.32E-19	94.2	92.30%
2019PM_7022	contig75105.1	---NA---	7.32E-19	94.2	92.30%
2019PM_4058	contig110278.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615 GO:0031409 GO:0036094 GO:0006810	1.22E-16	97.1	89.60%
2019PM_4058	contig208405.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615 GO:0031409 GO:0036094 GO:0006810	1.22E-16	97.1	89.60%
2019PM_5354	contig86597.1	---NA---	1.22E-16	97.1	89.60%
2019PM_5354	contig168619.1	---NA---	1.22E-16	97.1	89.60%
2019PM_3845	contig29835.1	reverse transcriptase	7.32E-19	94.2	92.30%
2019PM_6863	contig164538.1	---NA---	5.66E-20	92.75	93.80%
2019PM_6677	contig81436.1	---NA---	2.65E-13	100	85.90%
2019PM_3848	contig7255.1	cytochrome P450 9e2-like isoform X2|GO:0016020 GO:0043227	7.37E-14	98.55	86.80%
2019PM_3848	contig7378.1	reverse transcriptase	7.37E-14	98.55	86.80%
2019PM_3848	contig23221.1	reverse transcriptase	7.37E-14	98.55	86.80%
2019PM_3848	contig24482.1	reverse transcriptase	7.37E-14	98.55	86.80%
2019PM_3848	contig26094.1	reverse transcriptase	7.37E-14	98.55	86.80%
2019PM_3848	contig29835.1	reverse transcriptase	7.37E-14	98.55	86.80%
2019PM_3848	contig30116.1	reverse transcriptase	7.37E-14	98.55	86.80%
2019PM_3848	contig33033.1	reverse transcriptase	7.37E-14	98.55	86.80%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_3848	contig41249.1	hypothetical protein ALC56_06634, partial	7.37E-14	98.55	86.80%
2019PM_1635	contig155224.1	---NA---	7.37E-14	98.55	86.80%
2019PM_1635	contig169701.1	RNA-directed DNA polymerase from mobile element jockey	7.37E-14	98.55	86.80%
2019PM_4058	contig110278.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615 GO:0031409 GO:0036094 GO:0006810	5.70E-15	97.1	88.10%
2019PM_4058	contig208405.1		5.70E-15	97.1	88.10%
2019PM_4058	contig124618.1		2.03E-19	91.3	93.70%
2019PM_3754	contig155085.1	Nucleic-acid-binding from mobile element	4.40E-16	95.65	89.40%
2019PM_6488	contig46712.1	craniofacial development 2-like|GO:0003824 GO:0090304	1.57E-20	89.86	95.20%
2019PM_6488	contig116311.1	---NA---	1.57E-20	89.86	95.20%
2019PM_6488	contig41915.1	---NA---	1.57E-20	89.86	95.20%
2019PM_6488	contig46619.1	serine ase inhibitor	1.57E-20	89.86	95.20%
2019PM_6488	contig56254.1	---NA---	1.57E-20	89.86	95.20%
2019PM_6488	contig159278.1	---NA---	1.57E-20	89.86	95.20%
2019PM_7837	contig199395.1	endonuclease-reverse transcriptase	2.63E-18	92.75	92.30%
2019PM_3845	contig24304.1	reverse transcriptase	2.03E-19	91.3	93.70%
2019PM_3845	contig27826.1	Lactosylceramide 4-alpha-galactosyltransferase	2.03E-19	91.3	93.70%
2019PM_3845	contig57678.1	reverse transcriptase	2.03E-19	91.3	93.70%
2019PM_3845	contig21126.1	reverse transcriptase	3.40E-17	94.2	90.80%
2019PM_3845	contig32590.1	reverse transcriptase	3.40E-17	94.2	90.80%
2019PM_3845	contig34650.1	reverse transcriptase	3.40E-17	94.2	90.80%
2019PM_3845	contig35587.1	reverse transcriptase	3.40E-17	94.2	90.80%
2019PM_3845	contig35587.2	reverse transcriptase	3.40E-17	94.2	90.80%
2019PM_3845	contig52218.1	reverse transcriptase	3.40E-17	94.2	90.80%
2019PM_3845	contig52218.2	reverse transcriptase	3.40E-17	94.2	90.80%
2019PM_3845	contig52797.1	reverse transcriptase	3.40E-17	94.2	90.80%
2019PM_3845	contig29835.1	reverse transcriptase	3.40E-17	94.2	90.80%
2019PM_5879	contig11529.1	---NA---	1.23E-11	100	84.60%
2019PM_6730	contig2758.1	rna-directed dna polymerase from mobile element	1.23E-11	100	84.60%
2019PM_6730	contig4605.1	transcription elongation factor SPT4|GO:0005634 GO:0003746 GO:0008270 GO:0006351 GO:0006414 GO:0032786	1.23E-11	100	84.60%
2019PM_6730	contig4911.1		1.23E-11	100	84.60%
2019PM_6730	contig4911.2		1.23E-11	100	84.60%
2019PM_6730	contig4911.3		1.23E-11	100	84.60%
2019PM_6730	contig9497.1	RNA-directed DNA polymerase from mobile element jockey-like	1.23E-11	100	84.60%
2019PM_6730	contig18302.1	RNA-directed DNA polymerase from mobile element jockey-	1.23E-11	100	84.60%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_6730	contig32468.1	RNA-directed DNA polymerase from mobile element jockey-	1.23E-11	100	84.60%
2019PM_6730	contig90807.1	PREDICTED: uncharacterized protein LOC108681254	1.23E-11	100	84.60%
2019PM_1635	contig155224.1	---NA---	3.43E-12	98.55	85.30%
2019PM_1635	contig169701.1	RNA-directed DNA polymerase from mobile element jockey	3.43E-12	98.55	85.30%
2019PM_9618	contig30925.1	---NA---	5.66E-20	88.41	95.20%
2019PM_9618	contig101329.1	---NA---	5.66E-20	88.41	95.20%
2019PM_3494	contig33517.1	---NA---	5.66E-20	88.41	95.20%
2019PM_3494	contig74435.1	---NA---	5.66E-20	88.41	95.20%
2019PM_6488	contig181435.1	---NA---	2.63E-18	91.3	92.30%
2019PM_7837	contig190762.1	reverse transcriptase, partial|GO:0003964 GO:0006278	1.22E-16	92.75	90.80%
2019PM_9893	contig42570.1	craniofacial development 2-like	1.22E-16	92.75	90.60%
2019PM_9893	contig58908.1	craniofacial development 2-like	1.22E-16	92.75	90.60%
2019PM_5354	contig86597.1	---NA---	1.58E-15	94.2	89.20%
2019PM_5354	contig168619.1	---NA---	1.58E-15	94.2	89.20%
2019PM_3494	contig4673.1	craniofacial development 2-like	5.66E-20	88.41	95.10%
2019PM_3494	contig8588.1	craniofacial development 2-like|GO:0003824 GO:0090304	5.66E-20	88.41	95.10%
2019PM_3494	contig22038.1	craniofacial development 2-like	5.66E-20	88.41	95.10%
2019PM_3494	contig108523.1	---NA---	5.66E-20	88.41	95.10%
2019PM_3494	contig12013.1	craniofacial development 2-like	5.66E-20	88.41	95.10%
2019PM_3494	contig64855.1	---NA---	5.66E-20	88.41	95.10%
2019PM_3494	contig68760.1	---NA---	5.66E-20	88.41	95.10%
2019PM_3494	contig108523.1	---NA---	5.66E-20	88.41	95.10%
2019PM_3494	contig128553.1	---NA---	5.66E-20	88.41	95.10%
2019PM_6488	contig46712.1	craniofacial development 2-like|GO:0003824 GO:0090304	7.32E-19	89.86	93.50%
2019PM_6488	contig116311.1	---NA---	7.32E-19	89.86	93.50%
2019PM_3845	contig53726.1	reverse transcriptase	1.58E-15	94.2	89.20%
2019PM_3845	contig24304.1	reverse transcriptase	9.47E-18	91.3	92.10%
2019PM_3845	contig27826.1	Lactosylceramide 4-alpha-galactosyltransferase	9.47E-18	91.3	92.10%
2019PM_3845	contig57678.1	reverse transcriptase	9.47E-18	91.3	92.10%
2019PM_3845	contig21126.1	reverse transcriptase	1.58E-15	94.2	89.20%
2019PM_3845	contig32590.1	reverse transcriptase	1.58E-15	94.2	89.20%
2019PM_3845	contig34650.1	reverse transcriptase	1.58E-15	94.2	89.20%
2019PM_3845	contig35587.1	reverse transcriptase	1.58E-15	94.2	89.20%
2019PM_3845	contig35587.2	reverse transcriptase	1.58E-15	94.2	89.20%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_3845	contig52218.1	reverse transcriptase	1.58E-15	94.2	89.20%
2019PM_3845	contig52218.2	reverse transcriptase	1.58E-15	94.2	89.20%
2019PM_3845	contig52797.1	reverse transcriptase	1.58E-15	94.2	89.20%
2019PM_9768	contig35556.1	RNA-directed DNA polymerase from mobile element jockey-	5.70E-15	94.2	89.20%
2019PM_9768	contig38733.1	RNA-directed DNA polymerase from mobile element jockey-	5.70E-15	94.2	89.20%
2019PM_9618	contig43933.1	---NA---	7.32E-19	88.41	93.70%
2019PM_9618	contig43933.2	---NA---	2.63E-18	88.41	93.50%
2019PM_9618	contig49299.1	---NA---	7.32E-19	88.41	93.70%
2019PM_9618	contig91827.1	---NA---	2.63E-18	88.41	93.50%
2019PM_3494	contig33517.1	---NA---	2.63E-18	88.41	93.50%
2019PM_3494	contig74435.1	---NA---	2.63E-18	88.41	93.50%
2019PM_6488	contig49772.1	---NA---	2.03E-19	86.96	95.00%
2019PM_6488	contig181435.1	---NA---	1.22E-16	91.3	90.80%
2019PM_10046	contig112216.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	3.40E-17	89.86	92.10%
2019PM_7022	contig32241.1	---NA---	1.22E-16	89.86	91.90%
2019PM_7022	contig45025.1	---NA---	1.22E-16	89.86	91.90%
2019PM_4058	contig41993.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615 GO:0031409 GO:0036094 GO:0006810	7.37E-14	94.2	87.70%
2019PM_9618	contig44056.1	---NA---	2.63E-18	88.41	93.40%
2019PM_9893	contig215323.1	---NA---	5.70E-15	92.75	89.10%
2019PM_8143	contig21955.1	histone-lysine n-methyltransferase setmar	2.63E-18	88.41	93.40%
2019PM_3494	contig4673.1	craniofacial development 2-like	2.63E-18	88.41	93.40%
2019PM_3494	contig8588.1	craniofacial development 2-like|GO:0003824 GO:0090304	2.63E-18	88.41	93.40%
2019PM_3494	contig22038.1	craniofacial development 2-like	2.63E-18	88.41	93.40%
2019PM_6488	contig185065.1	---NA---	3.40E-17	89.86	91.90%
2019PM_6488	contig191186.1	---NA---	3.40E-17	89.86	91.90%
2019PM_6488	contig214317.1	---NA---	3.40E-17	89.86	91.90%
2019PM_6488	contig185065.1	---NA---	3.40E-17	89.86	91.90%
2019PM_6488	contig191186.1	---NA---	3.40E-17	89.86	91.90%
2019PM_6488	contig214317.1	---NA---	3.40E-17	89.86	91.90%
2019PM_3845	contig53726.1	reverse transcriptase	7.37E-14	94.2	87.70%
2019PM_9618	contig275.1	aminopeptidase N	5.66E-20	84.06	96.60%
2019PM_9618	contig925.1	aminopeptidase N	5.66E-20	84.06	96.60%
2019PM_9618	contig1774.1	aminopeptidase N-like	5.66E-20	84.06	96.60%
2019PM_9618	contig42740.1	---NA---	5.66E-20	84.06	96.60%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_9618	contig56295.1	---NA---	5.66E-20	84.06	96.60%
2019PM_9618	contig79071.1	---NA---	5.66E-20	84.06	96.60%
2019PM_9618	contig82569.1	---NA---	5.66E-20	84.06	96.60%
2019PM_9618	contig100662.1	---NA---	5.66E-20	84.06	96.60%
2019PM_9618	contig36599.1	---NA---	2.63E-18	86.96	93.50%
2019PM_8143	contig77398.1	---NA---	5.66E-20	84.06	96.60%
2019PM_4058	contig41993.1	AChain Apocrustacyanin C1 Crystals Grown In Space And Earth Using Vapour Diffusion Geometry|GO:0005615 GO:0031409 GO:0036094 GO:0006810	3.43E-12	94.2	86.20%
2019PM_9618	contig31839.1	---NA---	3.40E-17	88.41	92.10%
2019PM_9618	contig53200.1	---NA---	3.40E-17	88.41	92.10%
2019PM_8143	contig21377.1	RNA-directed DNA polymerase from	7.32E-19	85.51	94.90%
2019PM_8143	contig31354.1	---NA---	7.32E-19	85.51	94.90%
2019PM_3494	contig139262.1	---NA---	2.63E-18	85.51	94.90%
2019PM_10046	contig112216.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	1.58E-15	89.86	90.50%
2019PM_7022	contig85217.1	---NA---	4.40E-16	88.41	91.80%
2019PM_9618	contig210762.1	---NA---	1.22E-16	88.41	91.80%
2019PM_9748	contig99594.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.47E-18	86.96	93.30%
2019PM_7891	contig5907.1	DNA-directed RNA polymerase I subunit RPA2|GO:0016740	1.57E-20	81.16	98.20%
2019PM_9748	contig135699.1	reverse transcriptase, partial	2.03E-19	82.61	96.50%
2019PM_9748	contig192441.1	Retrovirus-related Pol poly LINE-|GO:0003964 GO:0006278	2.03E-19	82.61	96.50%
2019PM_10046	contig39732.1	endonuclease-reverse transcriptase	2.63E-18	84.06	94.90%
2019PM_10046	contig57241.1	endonuclease-reverse transcriptase	2.63E-18	84.06	94.90%
2019PM_9618	contig46946.1	---NA---	2.63E-18	84.06	94.80%
2019PM_9618	contig59376.1	---NA---	2.63E-18	84.06	94.80%
2019PM_9618	contig138308.1	---NA---	2.63E-18	84.06	94.80%
2019PM_8143	contig118524.1	---NA---	2.63E-18	84.06	94.80%
2019PM_8143	contig118524.1	---NA---	2.63E-18	84.06	94.80%
2019PM_8143	contig21377.1	RNA-directed DNA polymerase from	3.40E-17	85.51	93.20%
2019PM_8143	contig31354.1	---NA---	3.40E-17	85.51	93.20%
2019PM_3494	contig139262.1	---NA---	1.22E-16	85.51	93.20%
2019PM_6775	contig12501.1	RNA-directed DNA polymerase from mobile element jockey-like	5.70E-15	88.41	90.20%
2019PM_3494	contig89713.1	---NA---	1.58E-15	86.96	91.70%
2019PM_6488	contig110202.1	---NA---	2.65E-13	89.86	88.70%
2019PM_9748	contig9172.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.66E-20	79.71	98.20%
2019PM_9748	contig9172.2	craniofacial development 2-like|GO:0003964 GO:0006278	5.66E-20	79.71	98.20%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_9748	contig21690.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.47E-18	82.61	94.70%
2019PM_9748	contig12130.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.47E-18	82.61	94.70%
2019PM_9748	contig65233.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.47E-18	82.61	94.70%
2019PM_7022	contig54969.1	---NA---	4.40E-16	84.06	93.10%
2019PM_9618	contig219477.1	---NA---	1.22E-16	84.06	93.10%
2019PM_3406	contig25740.1	RNA-directed DNA polymerase from	1.22E-16	84.06	93.20%
2019PM_3406	contig35636.1	RNA-directed DNA polymerase from	1.22E-16	84.06	93.20%
2019PM_10046	contig39732.1	endonuclease-reverse transcriptase	1.22E-16	84.06	93.20%
2019PM_10046	contig57241.1	endonuclease-reverse transcriptase	1.22E-16	84.06	93.20%
2019PM_6775	contig12501.1	RNA-directed DNA polymerase from mobile element jockey-like	2.65E-13	88.41	88.50%
2019PM_7022	contig65981.1	---NA---	1.58E-15	85.51	91.50%
2019PM_9893	contig172528.1	---NA---	2.65E-13	88.41	88.50%
2019PM_10348	contig60958.1	reverse transcriptase, partial	5.70E-15	85.51	91.50%
2019PM_6449	contig170249.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	1.58E-15	85.51	91.50%
2019PM_6488	contig110202.1	---NA---	1.23E-11	89.86	87.10%
2019PM_3494	contig108635.1	---NA---	1.57E-20	76.81	100.00%
2019PM_3494	contig9387.1	Craniofacial development 2|GO:0003964 GO:0006278	2.03E-19	78.26	98.10%
2019PM_3494	contig48217.1	---NA---	2.03E-19	78.26	98.10%
2019PM_3494	contig48217.2	---NA---	2.03E-19	78.26	98.10%
2019PM_3492	contig47106.1	glutamate receptor NMDA 2B-|GO:0016021 GO:0030054 GO:0045211 GO:0004970 GO:0005234 GO:0034220 GO:0035235	2.03E-19	78.26	98.10%
2019PM_3492	contig47236.1	ionotropic glutamate receptor NMDA 2|GO:0016021 GO:0030054 GO:0045211 GO:0004970 GO:0005234 GO:0034220 GO:0035235	2.03E-19	78.26	98.10%
2019PM_9922	contig21299.1	---NA---	2.03E-19	78.26	98.10%
2019PM_9748	contig21119.1	craniofacial development 2-like|GO:0003964 GO:0006278	2.63E-18	79.71	96.40%
2019PM_9748	contig50200.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.63E-18	79.71	96.40%
2019PM_9748	contig65560.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	2.63E-18	79.71	96.40%
2019PM_3406	contig15718.1	rna-directed dna polymerase from mobile element jockey	3.40E-17	81.16	94.60%
2019PM_3406	contig25740.1	RNA-directed DNA polymerase from	5.70E-15	84.06	91.50%
2019PM_3406	contig35636.1	RNA-directed DNA polymerase from	5.70E-15	84.06	91.50%
2019PM_10046	contig132769.1	endonuclease-reverse transcriptase	5.70E-15	84.06	91.50%
2019PM_6211	contig2934.2	---NA---	5.70E-15	84.06	91.50%
2019PM_6211	contig25827.1	---NA---	5.70E-15	84.06	91.50%
2019PM_6211	contig25827.2	---NA---	5.70E-15	84.06	91.50%
2019PM_6211	contig38385.1	---NA---	5.70E-15	84.06	91.50%
2019PM_6211	contig201777.1	---NA---	5.70E-15	84.06	91.50%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_6211	contig204275.1	---NA---	5.70E-15	84.06	91.50%
2019PM_6211	contig43642.1	---NA---	7.37E-14	85.51	90.00%
2019PM_6211	contig50432.1	reverse transcriptase, partial|GO:0003964 GO:0006278	7.37E-14	85.51	90.00%
2019PM_6211	contig54086.2	---NA---	7.37E-14	85.51	90.00%
2019PM_6211	contig54086.3	---NA---	7.37E-14	85.51	90.00%
2019PM_6211	contig64067.1	---NA---	7.37E-14	85.51	90.00%
2019PM_6211	contig127920.1	---NA---	7.37E-14	85.51	90.00%
2019PM_6449	contig100612.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	84.06	91.40%
2019PM_6449	contig212013.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	84.06	91.40%
2019PM_6449	contig108767.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	7.37E-14	85.51	89.80%
2019PM_6449	contig195976.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	7.37E-14	85.51	89.80%
2019PM_8284	contig26478.1	---NA---	5.66E-20	75.36	100.00%
2019PM_3494	contig108635.1	---NA---	7.32E-19	76.81	98.10%
2019PM_9618	contig163670.1	---NA---	1.22E-16	79.71	94.60%
2019PM_3406	contig15718.1	rna-directed dna polymerase from mobile element jockey	1.58E-15	81.16	92.90%
2019PM_6449	contig15540.1	endonuclease-reverse transcriptase	1.58E-15	81.16	92.90%
2019PM_4372	contig554.1	Transposon Ty3-I Gag-Pol poly|GO:0005488 GO:0043170 GO:0044238	2.03E-19	73.91	100.00%
2019PM_4372	contig554.2	Transposon Ty3-I Gag-Pol poly|GO:0005488 GO:0043170 GO:0044238	2.03E-19	73.91	100.00%
2019PM_4372	contig2797.1	Transposon Ty3-I Gag-Pol poly|GO:0005488 GO:0043170 GO:0044238	2.03E-19	73.91	100.00%
2019PM_4372	contig3075.1	39S ribosomal mitochondrial	2.03E-19	73.91	100.00%
2019PM_4372	contig4683.1	Retroviral-like aspartic protease	2.03E-19	73.91	100.00%
2019PM_4372	contig4683.2	Retroviral-like aspartic protease	2.03E-19	73.91	100.00%
2019PM_4372	contig6869.1	Retrovirus-related Pol poly from transposon|GO:0005488	2.03E-19	73.91	100.00%
2019PM_4372	contig6869.2	Retrovirus-related Pol poly from transposon	2.03E-19	73.91	100.00%
2019PM_4372	contig6869.3	Retrovirus-related Pol poly from transposon	2.03E-19	73.91	100.00%
2019PM_4372	contig6869.5	retroviral aspartyl protease family	2.03E-19	73.91	100.00%
2019PM_4372	contig10330.1	Retrovirus-related Pol poly from transposon	2.03E-19	73.91	100.00%
2019PM_4372	contig10330.2	Retrovirus-related Pol poly from transposon	2.03E-19	73.91	100.00%
2019PM_4372	contig10330.4	Retrovirus-related Pol poly from transposon	2.03E-19	73.91	100.00%
2019PM_4372	contig10330.5	Retrovirus-related Pol poly from transposon	2.03E-19	73.91	100.00%
2019PM_4372	contig11262.1	Retrovirus-related Pol poly from transposon	2.03E-19	73.91	100.00%
2019PM_4372	contig11262.2	Retrovirus-related Pol poly from transposon	2.03E-19	73.91	100.00%
2019PM_4372	contig20767.1	Retrovirus-related Pol poly from transposon|GO:0005488	2.03E-19	73.91	100.00%
2019PM_4372	contig20767.2	Retrovirus-related Pol poly from transposon	2.03E-19	73.91	100.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_4372	contig22767.1	Retrovirus-related Pol poly from transposon|GO:0005488 GO:0043170 GO:0044238	2.03E-19	73.91	100.00%
2019PM_4372	contig22767.2	Retrovirus-related Pol poly from transposon|GO:0005488	2.03E-19	73.91	100.00%
2019PM_4372	contig37226.1	Retrovirus-related Pol poly from transposon	2.03E-19	73.91	100.00%
2019PM_4372	contig126763.1	---NA---	2.03E-19	73.91	100.00%
2019PM_4372	contig161608.1	---NA---	2.03E-19	73.91	100.00%
2019PM_125	contig7158.1	transcription factor SPT20 homolog isoform X1	2.03E-19	73.91	100.00%
2019PM_125	contig26093.1	abrupt-like isoform X1	2.03E-19	73.91	100.00%
2019PM_6449	contig20642.1	endonuclease-reverse transcriptase	2.65E-13	84.06	89.70%
2019PM_6488	contig216821.1	---NA---	2.65E-13	84.06	89.70%
2019PM_6488	contig216821.1	---NA---	2.65E-13	84.06	89.70%
2019PM_6211	contig54086.1	---NA---	3.43E-12	85.51	88.30%
2019PM_6449	contig210575.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	3.43E-12	85.51	88.10%
2019PM_8284	contig26478.1	---NA---	2.63E-18	75.36	98.10%
2019PM_9748	contig55518.1	endonuclease-reverse transcriptase	3.40E-17	76.81	96.20%
2019PM_3494	contig8335.1	---NA---	3.40E-17	76.81	96.20%
2019PM_3406	contig9721.1	RNA-directed DNA polymerase from mobile element jockey-	4.40E-16	78.26	94.40%
2019PM_3406	contig15527.1	---NA---	4.40E-16	78.26	94.40%
2019PM_3406	contig97121.1	---NA---	4.40E-16	78.26	94.40%
2019PM_3406	contig131255.1	---NA---	4.40E-16	78.26	94.40%
2019PM_3406	contig218656.1	pol-like protein, partial	4.40E-16	78.26	94.40%
2019PM_10046	contig67431.1	endonuclease-reverse transcriptase	5.70E-15	79.71	92.90%
2019PM_10046	contig67431.1	endonuclease-reverse transcriptase	5.70E-15	79.71	92.90%
2019PM_4464	contig9844.1	Chorion peroxidase|GO:0016491 GO:0050896	7.32E-19	72.46	100.00%
2019PM_4372	contig93663.1	Retrovirus-related Pol poly from transposon	7.32E-19	72.46	100.00%
2019PM_3214	contig15527.1	---NA---	7.32E-19	72.46	100.00%
2019PM_9912	contig43354.1	Retinol dehydrogenase 14	7.32E-19	72.46	100.00%
2019PM_7472	contig11419.1	---NA---	7.32E-19	72.46	100.00%
2019PM_7472	contig11419.2	---NA---	7.32E-19	72.46	100.00%
2019PM_7472	contig12692.1	phosphomethylethanolamine N-methyltransferase-like|GO:0016740	7.32E-19	72.46	100.00%
2019PM_7472	contig15151.1	phosphomethylethanolamine N-methyltransferase-like	7.32E-19	72.46	100.00%
2019PM_7472	contig17082.1	---NA---	7.32E-19	72.46	100.00%
2019PM_7472	contig22476.1	phosphomethylethanolamine N-methyltransferase-like	7.32E-19	72.46	100.00%
2019PM_7472	contig22476.2	phosphomethylethanolamine N-methyltransferase-like	7.32E-19	72.46	100.00%
2019PM_7472	contig39135.1	phosphomethylethanolamine N-methyltransferase-like|GO:0008757 GO:0008152	7.32E-19	72.46	100.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_7472	contig85494.1	---NA---	7.32E-19	72.46	100.00%
2019PM_6449	contig99594.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	7.37E-14	81.16	91.10%
2019PM_6449	contig151251.1	RNA-directed DNA polymerase from mobile element jockey|GO:0003964 GO:0006278	7.37E-14	81.16	91.10%
2019PM_10551	contig1316.1	RNA-directed DNA polymerase from	7.37E-14	81.16	91.10%
2019PM_10551	contig2758.1	rna-directed dna polymerase from mobile element	7.37E-14	81.16	91.10%
2019PM_10551	contig3177.1	RNA-directed DNA polymerase from mobile element jockey-like	7.37E-14	81.16	91.10%
2019PM_10551	contig3177.2	RNA-directed DNA polymerase from mobile element jockey-like	7.37E-14	81.16	91.10%
2019PM_10551	contig9949.1	RNA-directed DNA polymerase from transposon	7.37E-14	81.16	91.10%
2019PM_10551	contig28122.1	Gag-Pol poly	7.37E-14	81.16	91.10%
2019PM_10551	contig33912.1	Gag-Pol poly	7.37E-14	81.16	91.10%
2019PM_10551	contig38733.1	RNA-directed DNA polymerase from mobile element jockey-	7.37E-14	81.16	91.10%
2019PM_10551	contig95310.1	---NA---	7.37E-14	81.16	91.10%
2019PM_10551	contig117383.1	---NA---	7.37E-14	81.16	91.10%
2019PM_4372	contig126764.1	---NA---	9.47E-18	73.91	98.00%
2019PM_125	contig7158.1	transcription factor SPT20 homolog isoform X1	9.47E-18	73.91	98.00%
2019PM_125	contig26093.1	abrupt-like isoform X1	9.47E-18	73.91	98.00%
2019PM_7344	contig29059.1	endonuclease-reverse transcriptase	1.22E-16	75.36	96.20%
2019PM_7344	contig29059.2	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	1.22E-16	75.36	96.20%
2019PM_7344	contig34716.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	1.22E-16	75.36	96.20%
2019PM_7344	contig39733.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	1.22E-16	75.36	96.20%
2019PM_7344	contig56169.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	1.22E-16	75.36	96.20%
2019PM_7344	contig105287.1	Craniofacial development|GO:0003964 GO:0006278	1.22E-16	75.36	96.20%
2019PM_9618	contig196241.1	---NA---	1.22E-16	75.36	96.20%
2019PM_9748	contig39990.1	reverse transcriptase, partial|GO:0003964 GO:0006278	1.58E-15	76.81	94.30%
2019PM_9748	contig120288.1	outcast ele5 orf1 -h 1e-40 -j|GO:0003964 GO:0006278	1.58E-15	76.81	94.30%
2019PM_3494	contig8335.1	---NA---	1.58E-15	76.81	94.30%
2019PM_3043	contig5811.1	reverse transcriptase	1.58E-15	76.81	94.30%
2019PM_3043	contig5811.2	reverse transcriptase	1.58E-15	76.81	94.30%
2019PM_8143	contig44143.2	pol-like protein, partial	2.63E-18	71.01	100.00%
2019PM_8143	contig111297.1	---NA---	2.63E-18	71.01	100.00%
2019PM_404	contig31966.1	Kv channel-interacting 2-like|GO:0016021 GO:0005509	2.63E-18	71.01	100.00%
2019PM_404	contig122781.1	Kv channel-interacting 2-like isoform X2|GO:0016020 GO:0005509	2.63E-18	71.01	100.00%
2019PM_6211	contig2934.1	RNA-directed DNA polymerase from mobile element jockey-like	2.05E-14	78.26	92.70%
2019PM_6211	contig15192.1	blastopia poly	2.05E-14	78.26	92.70%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_6211	contig16465.1	adenylate cyclase type 3	2.05E-14	78.26	92.70%
2019PM_6211	contig30266.1	PREDICTED: uncharacterized protein LOC109418678	2.05E-14	78.26	92.70%
2019PM_6211	contig64412.1	---NA---	2.05E-14	78.26	92.70%
2019PM_6211	contig175000.1	---NA---	2.05E-14	78.26	92.70%
2019PM_3648	contig191543.1	endonuclease-reverse transcriptase	7.37E-14	79.71	91.20%
2019PM_3648	contig9387.1	Craniofacial development 2|GO:0003964 GO:0006278	9.53E-13	81.16	89.80%
2019PM_3648	contig14204.1	peroxisomal trans-2-enoyl- reductase-like|GO:0005777 GO:0019166 GO:0033306 GO:0055114	9.53E-13	81.16	89.80%
2019PM_3406	contig9721.1	RNA-directed DNA polymerase from mobile element jockey-	2.05E-14	78.26	92.60%
2019PM_3406	contig15527.1	---NA---	2.05E-14	78.26	92.60%
2019PM_3406	contig97121.1	---NA---	2.05E-14	78.26	92.60%
2019PM_3406	contig131255.1	---NA---	2.05E-14	78.26	92.60%
2019PM_3406	contig218656.1	pol-like protein, partial	2.05E-14	78.26	92.60%
2019PM_1316	contig23768.1	RNA-directed DNA polymerase from mobile element jockey-like	3.40E-17	72.46	98.00%
2019PM_1316	contig56529.1	RNA-directed DNA polymerase from mobile element jockey-like	3.40E-17	72.46	98.00%
2019PM_1316	contig56529.2	RNA-directed DNA polymerase from mobile element jockey-like	3.40E-17	72.46	98.00%
2019PM_1316	contig56963.1	RNA-directed DNA polymerase from mobile element jockey-like	3.40E-17	72.46	98.00%
2019PM_1316	contig196289.1	RNA-directed DNA polymerase from mobile element jockey-like	3.40E-17	72.46	98.00%
2019PM_3214	contig15527.1	---NA---	3.40E-17	72.46	98.00%
2019PM_3214	contig8771.1	transcription elongation factor SPT4|GO:0005634 GO:0003746 GO:0008270 GO:0006351 GO:0006414 GO:0032786	3.40E-17	72.46	98.00%
2019PM_3214	contig8771.2	non-ltr retrotransposon r1bmks orf2	3.40E-17	72.46	98.00%
2019PM_3214	contig9721.1	RNA-directed DNA polymerase from mobile element jockey-	3.40E-17	72.46	98.00%
2019PM_3214	contig9949.1	RNA-directed DNA polymerase from transposon	3.40E-17	72.46	98.00%
2019PM_3214	contig11495.1	---NA---	3.40E-17	72.46	98.00%
2019PM_3214	contig15805.1	RNA-directed DNA polymerase from mobile element	3.40E-17	72.46	98.00%
2019PM_3214	contig16953.1	RNA-directed DNA polymerase from mobile element jockey	3.40E-17	72.46	98.00%
2019PM_3214	contig17467.1	RNA-directed DNA polymerase from mobile element jockey-like	3.40E-17	72.46	98.00%
2019PM_3214	contig21522.1	non-ltr retrotransposon r1bmks orf2	3.40E-17	72.46	98.00%
2019PM_3214	contig31530.1	RNA-directed DNA polymerase from mobile element jockey-like	3.40E-17	72.46	98.00%
2019PM_3214	contig31961.1	RNA-directed DNA polymerase from mobile element jockey	3.40E-17	72.46	98.00%
2019PM_3214	contig47006.1	rna-directed dna polymerase from mobile element	3.40E-17	72.46	98.00%
2019PM_3214	contig50960.1	RNA-directed DNA polymerase from transposon X-	3.40E-17	72.46	98.00%
2019PM_3214	contig108961.1	---NA---	3.40E-17	72.46	98.00%
2019PM_3214	contig120051.1	reverse transcriptase	3.40E-17	72.46	98.00%
2019PM_7472	contig11419.1	---NA---	3.40E-17	72.46	98.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_7472	contig11419.2	---NA---	3.40E-17	72.46	98.00%
2019PM_7472	contig12692.1	phosphomethylethanolamine N-methyltransferase-like|GO:0016740	3.40E-17	72.46	98.00%
2019PM_7472	contig15151.1	phosphomethylethanolamine N-methyltransferase-like	3.40E-17	72.46	98.00%
2019PM_7472	contig17082.1	---NA---	3.40E-17	72.46	98.00%
2019PM_7472	contig22476.1	phosphomethylethanolamine N-methyltransferase-like	3.40E-17	72.46	98.00%
2019PM_7472	contig22476.2	phosphomethylethanolamine N-methyltransferase-like	3.40E-17	72.46	98.00%
2019PM_7472	contig39135.1	phosphomethylethanolamine N-methyltransferase-like|GO:0008757 GO:0008152	3.40E-17	72.46	98.00%
2019PM_7472	contig85494.1	---NA---	3.40E-17	72.46	98.00%
2019PM_7344	contig29059.1	endonuclease-reverse transcriptase	4.40E-16	73.91	96.10%
2019PM_7344	contig29059.2	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.40E-16	73.91	96.10%
2019PM_7344	contig34716.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	4.40E-16	73.91	96.10%
2019PM_7344	contig39733.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	4.40E-16	73.91	96.10%
2019PM_7344	contig56169.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.40E-16	73.91	96.10%
2019PM_7344	contig105287.1	Craniofacial development|GO:0003964 GO:0006278	4.40E-16	73.91	96.10%
2019PM_7344	contig206.1	aminopeptidase N	5.70E-15	75.36	94.20%
2019PM_7344	contig283.1	aminopeptidase N	5.70E-15	75.36	94.20%
2019PM_7344	contig9530.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	75.36	94.20%
2019PM_7344	contig15911.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	75.36	94.20%
2019PM_7344	contig21122.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	75.36	94.20%
2019PM_7344	contig36423.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	75.36	94.20%
2019PM_7344	contig77924.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	75.36	94.20%
2019PM_7344	contig86870.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	5.70E-15	75.36	94.20%
2019PM_7344	contig87111.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	75.36	94.20%
2019PM_7344	contig118789.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	75.36	94.20%
2019PM_8143	contig104448.1	---NA---	9.47E-18	69.57	100.00%
2019PM_5354	contig77779.1	---NA---	5.70E-15	75.36	94.20%
2019PM_5354	contig202419.1	---NA---	5.70E-15	75.36	94.20%
2019PM_5354	contig77779.1	---NA---	5.70E-15	75.36	94.20%
2019PM_5354	contig202419.1	---NA---	5.70E-15	75.36	94.20%
2019PM_6651	contig151006.1	reverse transcriptase, partial|GO:0003964 GO:0006278	2.05E-14	75.36	94.20%
2019PM_404	contig17872.1	Kv channel-interacting 2-like isoform X2|GO:0016021 GO:0005509	9.47E-18	69.57	100.00%
2019PM_404	contig51495.1	Kv channel-interacting 2-like|GO:0016021 GO:0005509	9.47E-18	69.57	100.00%
2019PM_404	contig65508.1	Kv channel-interacting 2-like|GO:0016021 GO:0005509	9.47E-18	69.57	100.00%
2019PM_3214	contig14286.1	rna-directed dna polymerase from mobile element	9.47E-18	69.57	100.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_6211	contig63842.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	7.37E-14	76.81	92.60%
2019PM_3043	contig5811.1	reverse transcriptase	7.37E-14	76.81	92.50%
2019PM_3043	contig5811.2	reverse transcriptase	7.37E-14	76.81	92.50%
2019PM_8143	contig44143.2	pol-like protein, partial	1.22E-16	71.01	98.00%
2019PM_8143	contig111297.1	---NA---	1.22E-16	71.01	98.00%
2019PM_3406	contig11723.1	RNA-directed DNA polymerase from mobile element jockey	9.53E-13	78.26	90.70%
2019PM_3406	contig134915.1	---NA---	9.53E-13	78.26	90.70%
2019PM_5354	contig40314.1	---NA---	1.22E-16	71.01	98.00%
2019PM_9748	contig72167.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	1.22E-16	71.01	98.00%
2019PM_4418	contig204635.1	---NA---	4.44E-11	79.71	89.10%
2019PM_1316	contig23768.1	RNA-directed DNA polymerase from mobile element jockey-like	1.58E-15	72.46	96.00%
2019PM_1316	contig56529.1	RNA-directed DNA polymerase from mobile element jockey-like	1.58E-15	72.46	96.00%
2019PM_1316	contig56529.2	RNA-directed DNA polymerase from mobile element jockey-like	1.58E-15	72.46	96.00%
2019PM_1316	contig56963.1	RNA-directed DNA polymerase from mobile element jockey-like	1.58E-15	72.46	96.00%
2019PM_1316	contig196289.1	RNA-directed DNA polymerase from mobile element jockey-like	1.58E-15	72.46	96.00%
2019PM_7344	contig99459.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	1.58E-15	72.46	96.00%
2019PM_7344	contig99459.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	1.58E-15	72.46	96.00%
2019PM_4464	contig9844.1	Chorion peroxidase|GO:0016491 GO:0050896	1.58E-15	72.46	96.00%
2019PM_3043	contig26094.1	reverse transcriptase	1.58E-15	72.46	96.00%
2019PM_3214	contig11723.1	RNA-directed DNA polymerase from mobile element jockey	1.58E-15	72.46	96.00%
2019PM_3214	contig45666.1	RNA-directed DNA polymerase from mobile element jockey-like	1.58E-15	72.46	96.00%
2019PM_3214	contig87523.1	PREDICTED: uncharacterized protein LOC109401447	1.58E-15	72.46	96.00%
2019PM_3214	contig154710.1	---NA---	1.58E-15	72.46	96.00%
2019PM_354	contig137397.1	---NA---	3.40E-17	68.12	100.00%
2019PM_8143	contig44143.1	pol-like protein, partial	3.40E-17	68.12	100.00%
2019PM_3494	contig78121.1	craniofacial development 2-like	3.40E-17	68.12	100.00%
2019PM_9986	contig120903.1	Retrovirus-related Pol poly from type-2 retrotransposable element R2DM	3.40E-17	68.12	100.00%
2019PM_964	contig28979.1	---NA---	3.40E-17	68.12	100.00%
2019PM_7344	contig206.1	aminopeptidase N	2.05E-14	73.91	94.10%
2019PM_7344	contig283.1	aminopeptidase N	2.05E-14	73.91	94.10%
2019PM_7344	contig9530.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	73.91	94.10%
2019PM_7344	contig15911.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	73.91	94.10%
2019PM_7344	contig21122.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	73.91	94.10%
2019PM_7344	contig36423.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	73.91	94.10%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_7344	contig77924.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	73.91	94.10%
2019PM_7344	contig86870.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	2.05E-14	73.91	94.10%
2019PM_7344	contig87111.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	73.91	94.10%
2019PM_7344	contig118789.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	73.91	94.10%
2019PM_9893	contig98916.1	---NA---	2.05E-14	73.91	94.10%
2019PM_7344	contig51447.2	reverse transcriptase, partial|GO:0003964 GO:0006278	2.65E-13	75.36	92.30%
2019PM_7344	contig56598.1	craniofacial development 2-like|GO:0003964 GO:0006278	2.65E-13	75.36	92.30%
2019PM_7344	contig102769.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.65E-13	75.36	92.30%
2019PM_7344	contig120288.1	outcast ele5 orf1 -h 1e-40 -j|GO:0003964 GO:0006278	2.65E-13	75.36	92.30%
2019PM_7344	contig92457.1	reverse transcriptase, partial	4.40E-16	69.57	97.90%
2019PM_8143	contig104448.1	---NA---	4.40E-16	69.57	97.90%
2019PM_3406	contig42028.1	---NA---	4.44E-11	78.26	89.10%
2019PM_3648	contig21122.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.40E-16	69.57	97.90%
2019PM_3648	contig205111.1	craniofacial development 2-like	4.40E-16	69.57	97.90%
2019PM_3214	contig14286.1	rna-directed dna polymerase from mobile element	4.40E-16	69.57	97.90%
2019PM_3406	contig11723.1	RNA-directed DNA polymerase from mobile element jockey	4.44E-11	78.26	88.90%
2019PM_3406	contig134915.1	---NA---	4.44E-11	78.26	88.90%
2019PM_8143	contig126600.1	---NA---	5.70E-15	71.01	95.90%
2019PM_8143	contig199562.1	---NA---	5.70E-15	71.01	95.90%
2019PM_4418	contig204635.1	---NA---	2.06E-09	79.71	87.30%
2019PM_9748	contig36241.1	reverse transcriptase, partial|GO:0003964 GO:0006278	5.70E-15	71.01	95.90%
2019PM_1316	contig4882.1	rna-directed dna polymerase from mobile element jockey	1.22E-16	66.67	100.00%
2019PM_1316	contig4911.1	transcription elongation factor SPT4|GO:0005634 GO:0003746 GO:0008270 GO:0006351 GO:0006414 GO:0032786	1.22E-16	66.67	100.00%
2019PM_1316	contig8125.2	RNA-directed DNA polymerase from mobile element jockey	1.22E-16	66.67	100.00%
2019PM_1316	contig13732.2	RNA-directed DNA polymerase from mobile element jockey-like	1.22E-16	66.67	100.00%
2019PM_1316	contig31924.1	RNA-directed DNA polymerase from mobile element jockey-like	1.22E-16	66.67	100.00%
2019PM_4573	contig64050.1	---NA---	1.22E-16	66.67	100.00%
2019PM_4573	contig68161.1	---NA---	1.22E-16	66.67	100.00%
2019PM_4573	contig127735.1	---NA---	1.22E-16	66.67	100.00%
2019PM_9249	contig1189.1	Transposon Ty3-I Gag-Pol poly	1.22E-16	66.67	100.00%
2019PM_9249	contig1189.2	Retrovirus-related Pol poly from transposon|GO:0003824 GO:0090304	1.22E-16	66.67	100.00%
2019PM_9270	contig57715.1	---NA---	1.22E-16	66.67	100.00%
2019PM_9270	contig57715.2	---NA---	1.22E-16	66.67	100.00%
2019PM_4109	contig10762.1	Zinc finger and BTB domain-containing	1.22E-16	66.67	100.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_1316	contig116189.1	---NA---	7.37E-14	72.46	94.00%
2019PM_1316	contig143617.1	115 kDa in type-1 retrotransposable element R1DM	7.37E-14	72.46	94.00%
2019PM_7344	contig8588.1	craniofacial development 2-like|GO:0003824 GO:0090304	7.37E-14	72.46	94.00%
2019PM_7344	contig8588.1	craniofacial development 2-like|GO:0003824 GO:0090304	7.37E-14	72.46	94.00%
2019PM_9748	contig80242.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	7.37E-14	72.46	94.00%
2019PM_9748	contig190762.1	reverse transcriptase, partial|GO:0003964 GO:0006278	7.37E-14	72.46	94.00%
2019PM_3043	contig26094.1	reverse transcriptase	7.37E-14	72.46	94.00%
2019PM_3530	contig62355.1	endonuclease-reverse transcriptase|GO:0003964 GO:0004519 GO:0006278 GO:0090305	2.65E-13	72.46	94.00%
2019PM_3530	contig63059.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	2.65E-13	72.46	94.00%
2019PM_3214	contig11723.1	RNA-directed DNA polymerase from mobile element jockey	7.37E-14	72.46	94.00%
2019PM_3214	contig45666.1	RNA-directed DNA polymerase from mobile element jockey-like	7.37E-14	72.46	94.00%
2019PM_3214	contig87523.1	PREDICTED: uncharacterized protein LOC109401447	7.37E-14	72.46	94.00%
2019PM_3214	contig154710.1	---NA---	7.37E-14	72.46	94.00%
2019PM_6638	contig62820.1	endonuclease-reverse transcriptase	7.37E-14	72.46	94.00%
2019PM_7344	contig92457.1	reverse transcriptase, partial	1.58E-15	68.12	97.90%
2019PM_7344	contig51447.2	reverse transcriptase, partial|GO:0003964 GO:0006278	9.53E-13	73.91	92.20%
2019PM_7344	contig56598.1	craniofacial development 2-like|GO:0003964 GO:0006278	9.53E-13	73.91	92.20%
2019PM_7344	contig102769.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	73.91	92.20%
2019PM_7344	contig120288.1	outcast ele5 orf1 -h 1e-40 -j|GO:0003964 GO:0006278	9.53E-13	73.91	92.20%
2019PM_8143	contig153787.1	---NA---	1.58E-15	68.12	97.90%
2019PM_8143	contig44143.1	pol-like protein, partial	1.58E-15	68.12	97.90%
2019PM_5354	contig40314.1	---NA---	1.58E-15	68.12	97.90%
2019PM_9986	contig120903.1	Retrovirus-related Pol poly from type-2 retrotransposable element R2DM	1.58E-15	68.12	97.90%
2019PM_9986	contig44692.1	Retrovirus-related Pol poly from type-2 retrotransposable element R2DM	1.58E-15	68.12	97.90%
2019PM_964	contig28979.1	---NA---	1.58E-15	68.12	97.90%
2019PM_3406	contig42028.1	---NA---	2.06E-09	78.26	87.30%
2019PM_3648	contig9411.1	C-type lectin|GO:0030246	2.05E-14	69.57	95.80%
2019PM_3648	contig9414.1	craniofacial development 2-like|GO:0030246	2.05E-14	69.57	95.80%
2019PM_3648	contig16404.1	craniofacial development 2-like	2.05E-14	69.57	95.80%
2019PM_3648	contig16528.1	craniofacial development 2-like|GO:0003964 GO:0006278	2.05E-14	69.57	95.80%
2019PM_3648	contig23900.1	craniofacial development 2-like|GO:0030246	2.05E-14	69.57	95.80%
2019PM_3648	contig45528.1	craniofacial development 2-like	2.05E-14	69.57	95.80%
2019PM_3648	contig61111.1	---NA---	2.05E-14	69.57	95.80%
2019PM_3648	contig134694.1	craniofacial development 2-like	2.05E-14	69.57	95.80%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_10348	contig210.2	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.40E-16	65.22	100.00%
2019PM_10348	contig2385.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.40E-16	65.22	100.00%
2019PM_10348	contig8010.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.40E-16	65.22	100.00%
2019PM_10348	contig12264.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.40E-16	65.22	100.00%
2019PM_10348	contig12264.2	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.40E-16	65.22	100.00%
2019PM_10348	contig13854.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.40E-16	65.22	100.00%
2019PM_10348	contig14754.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.40E-16	65.22	100.00%
2019PM_10348	contig20383.1	reverse transcriptase, partial	4.40E-16	65.22	100.00%
2019PM_10348	contig30774.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.40E-16	65.22	100.00%
2019PM_10348	contig37527.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	4.40E-16	65.22	100.00%
2019PM_10348	contig38937.1	craniofacial development 2-like|GO:0003964 GO:0006278	4.40E-16	65.22	100.00%
2019PM_10348	contig49245.1	endonuclease-reverse transcriptase	4.40E-16	65.22	100.00%
2019PM_10348	contig82623.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.40E-16	65.22	100.00%
2019PM_10348	contig91465.1	reverse transcriptase, partial|GO:0003964 GO:0006278	4.40E-16	65.22	100.00%
2019PM_1233	contig50442.1	serine-rich adhesin for platelets	4.40E-16	65.22	100.00%
2019PM_7344	contig154046.1	Craniofacial development|GO:0003964 GO:0006278	2.65E-13	71.01	93.90%
2019PM_6677	contig126611.1	---NA---	2.65E-13	71.01	93.90%
2019PM_1316	contig8082.1	lian-aa1 retrotransposon	5.70E-15	66.67	97.80%
2019PM_1316	contig11695.1	RNA-directed DNA polymerase from mobile element jockey-	5.70E-15	66.67	97.80%
2019PM_1316	contig13732.1	RNA-directed DNA polymerase from mobile element jockey-like	5.70E-15	66.67	97.80%
2019PM_1316	contig32468.1	RNA-directed DNA polymerase from mobile element jockey-	5.70E-15	66.67	97.80%
2019PM_1316	contig90807.1	PREDICTED: uncharacterized protein LOC108681254	5.70E-15	66.67	97.80%
2019PM_1316	contig4882.1	rna-directed dna polymerase from mobile element jockey	5.70E-15	66.67	97.80%
2019PM_1316	contig4911.1	transcription elongation factor SPT4|GO:0005634 GO:0003746 GO:0008270 GO:0006351 GO:0006414 GO:0032786	5.70E-15	66.67	97.80%
2019PM_1316	contig8125.2	RNA-directed DNA polymerase from mobile element jockey	5.70E-15	66.67	97.80%
2019PM_1316	contig13732.2	RNA-directed DNA polymerase from mobile element jockey-like	5.70E-15	66.67	97.80%
2019PM_1316	contig31924.1	RNA-directed DNA polymerase from mobile element jockey-like	5.70E-15	66.67	97.80%
2019PM_1316	contig87201.1	PREDICTED: uncharacterized protein LOC106685760	5.70E-15	66.67	97.80%
2019PM_1316	contig87202.1	RNA-directed DNA polymerase from mobile element jockey-	5.70E-15	66.67	97.80%
2019PM_1316	contig116189.1	---NA---	3.43E-12	72.46	92.00%
2019PM_1316	contig143617.1	115 kDa in type-1 retrotransposable element R1DM	3.43E-12	72.46	92.00%
2019PM_8143	contig15724.1	Gag-Pol poly|GO:0003824 GO:0044260 GO:0090304	5.70E-15	66.67	97.80%
2019PM_8143	contig33912.1	Gag-Pol poly	5.70E-15	66.67	97.80%
2019PM_8143	contig35845.1	reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	66.67	97.80%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_8143	contig38405.1	RNA-directed DNA polymerase from mobile element	5.70E-15	66.67	97.80%
2019PM_8143	contig55830.1	---NA---	5.70E-15	66.67	97.80%
2019PM_8143	contig72052.1	---NA---	5.70E-15	66.67	97.80%
2019PM_4573	contig64050.1	---NA---	5.70E-15	66.67	97.80%
2019PM_4573	contig68161.1	---NA---	5.70E-15	66.67	97.80%
2019PM_4573	contig127735.1	---NA---	5.70E-15	66.67	97.80%
2019PM_10348	contig10338.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	66.67	97.80%
2019PM_10348	contig11118.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	66.67	97.80%
2019PM_10348	contig11122.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	66.67	97.80%
2019PM_10348	contig12697.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	66.67	97.80%
2019PM_10348	contig13130.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	66.67	97.80%
2019PM_10348	contig13790.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	66.67	97.80%
2019PM_10348	contig38675.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	66.67	97.80%
2019PM_10348	contig38675.2	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	5.70E-15	66.67	97.80%
2019PM_9249	contig1189.1	Transposon Ty3-I Gag-Pol poly	5.70E-15	66.67	97.80%
2019PM_9249	contig1189.2	Retrovirus-related Pol poly from transposon|GO:0003824 GO:0090304	5.70E-15	66.67	97.80%
2019PM_9270	contig57715.1	---NA---	5.70E-15	66.67	97.80%
2019PM_9270	contig57715.2	---NA---	5.70E-15	66.67	97.80%
2019PM_9729	contig180018.1	---NA---	3.43E-12	72.46	92.00%
2019PM_3214	contig95778.1	---NA---	3.43E-12	72.46	92.00%
2019PM_3214	contig98017.1	RNA-directed DNA polymerase from mobile element jockey-like	3.43E-12	72.46	92.00%
2019PM_3214	contig121761.1	---NA---	3.43E-12	72.46	92.00%
2019PM_3214	contig196289.1	RNA-directed DNA polymerase from mobile element jockey-like	3.43E-12	72.46	92.00%
2019PM_3406	contig6357.1	Gag-Pol poly	4.44E-11	73.91	90.20%
2019PM_7344	contig8551.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	7.37E-14	68.12	95.70%
2019PM_7344	contig8551.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	7.37E-14	68.12	95.70%
2019PM_4418	contig107538.1	---NA---	2.65E-13	68.12	95.70%
2019PM_7673	contig52099.1	---NA---	1.58E-15	63.77	100.00%
2019PM_5602	contig151783.1	---NA---	1.58E-15	63.77	100.00%
2019PM_6223	contig1026.1	PREDICTED: uncharacterized protein CG43867-like	1.58E-15	63.77	100.00%
2019PM_9986	contig44692.1	Retrovirus-related Pol poly from type-2 retrotransposable element R2DM	7.37E-14	68.12	95.70%
2019PM_3530	contig115129.1	endonuclease-reverse transcriptase|GO:0003964 GO:0004519 GO:0006278 GO:0090305	7.37E-14	68.12	95.70%
2019PM_7344	contig154046.1	Craniofacial development|GO:0003964 GO:0006278	9.53E-13	69.57	93.80%
2019PM_10046	contig123905.1	endonuclease-reverse transcriptase	9.53E-13	69.57	93.90%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_10046	contig123905.1	endonuclease-reverse transcriptase	9.53E-13	69.57	93.90%
2019PM_3648	contig6494.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	69.57	93.80%
2019PM_3648	contig12013.1	craniofacial development 2-like	9.53E-13	69.57	93.80%
2019PM_3648	contig22038.1	craniofacial development 2-like	9.53E-13	69.57	93.80%
2019PM_3648	contig42607.1	craniofacial development 2-like	9.53E-13	69.57	93.80%
2019PM_3648	contig44233.1	craniofacial development 2-like	9.53E-13	69.57	93.80%
2019PM_3648	contig58064.1	craniofacial development 2-like	9.53E-13	69.57	93.80%
2019PM_3648	contig62552.1	craniofacial development 2-like	9.53E-13	69.57	93.80%
2019PM_3648	contig63934.1	craniofacial development 2-like	9.53E-13	69.57	93.80%
2019PM_3648	contig88568.1	craniofacial development 2-like	9.53E-13	69.57	93.80%
2019PM_3648	contig92199.1	craniofacial development 2-like	9.53E-13	69.57	93.80%
2019PM_3648	contig95950.1	craniofacial development 2-like	9.53E-13	69.57	93.80%
2019PM_3648	contig107290.1	craniofacial development 2-like	9.53E-13	69.57	93.80%
2019PM_8143	contig15724.1	Gag-Pol poly|GO:0003824 GO:0044260 GO:0090304	2.05E-14	65.22	97.80%
2019PM_8143	contig33912.1	Gag-Pol poly	2.05E-14	65.22	97.80%
2019PM_8143	contig35845.1	reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	65.22	97.80%
2019PM_8143	contig38405.1	RNA-directed DNA polymerase from mobile element	2.05E-14	65.22	97.80%
2019PM_8143	contig55830.1	---NA---	2.05E-14	65.22	97.80%
2019PM_8143	contig72052.1	---NA---	2.05E-14	65.22	97.80%
2019PM_8143	contig126600.1	---NA---	2.05E-14	65.22	97.80%
2019PM_8143	contig199562.1	---NA---	2.05E-14	65.22	97.80%
2019PM_10348	contig210.2	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	65.22	97.80%
2019PM_10348	contig2385.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	65.22	97.80%
2019PM_10348	contig8010.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	65.22	97.80%
2019PM_10348	contig12264.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	65.22	97.80%
2019PM_10348	contig12264.2	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	65.22	97.80%
2019PM_10348	contig13854.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	65.22	97.80%
2019PM_10348	contig14754.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	65.22	97.80%
2019PM_10348	contig20383.1	reverse transcriptase, partial	2.05E-14	65.22	97.80%
2019PM_10348	contig30774.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	65.22	97.80%
2019PM_10348	contig37527.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	2.05E-14	65.22	97.80%
2019PM_10348	contig38937.1	craniofacial development 2-like|GO:0003964 GO:0006278	2.05E-14	65.22	97.80%
2019PM_10348	contig49245.1	endonuclease-reverse transcriptase	2.05E-14	65.22	97.80%
2019PM_10348	contig82623.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	65.22	97.80%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_10348	contig91465.1	reverse transcriptase, partial|GO:0003964 GO:0006278	2.05E-14	65.22	97.80%
2019PM_10348	contig43349.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	65.22	97.80%
2019PM_10348	contig55171.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.05E-14	65.22	97.80%
2019PM_10348	contig86870.1	endonuclease-reverse transcriptase|GO:0003824 GO:0090304	2.05E-14	65.22	97.80%
2019PM_1233	contig50442.1	serine-rich adhesin for platelets	2.05E-14	65.22	97.80%
2019PM_7333	contig14656.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	5.70E-15	62.32	100.00%
2019PM_7333	contig15205.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	5.70E-15	62.32	100.00%
2019PM_1316	contig87201.1	PREDICTED: uncharacterized protein LOC106685760	2.65E-13	66.67	95.70%
2019PM_1316	contig87202.1	RNA-directed DNA polymerase from mobile element jockey-	2.65E-13	66.67	95.70%
2019PM_1316	contig8082.1	lian-aa1 retrotransposon	2.65E-13	66.67	95.70%
2019PM_1316	contig11695.1	RNA-directed DNA polymerase from mobile element jockey-	2.65E-13	66.67	95.70%
2019PM_1316	contig13732.1	RNA-directed DNA polymerase from mobile element jockey-like	2.65E-13	66.67	95.70%
2019PM_1316	contig32468.1	RNA-directed DNA polymerase from mobile element jockey-	2.65E-13	66.67	95.70%
2019PM_1316	contig90807.1	PREDICTED: uncharacterized protein LOC108681254	2.65E-13	66.67	95.70%
2019PM_8143	contig64062.1	reverse transcriptase	5.70E-15	62.32	100.00%
2019PM_8143	contig64062.1	reverse transcriptase	5.70E-15	62.32	100.00%
2019PM_3406	contig6614.1	Gag-Pol poly	1.60E-10	72.46	90.00%
2019PM_3406	contig6721.1	Gag-Pol poly	1.60E-10	72.46	90.00%
2019PM_3406	contig6721.2	Gag-Pol poly	1.60E-10	72.46	90.00%
2019PM_10348	contig10338.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.65E-13	66.67	95.70%
2019PM_10348	contig11118.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.65E-13	66.67	95.70%
2019PM_10348	contig11122.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.65E-13	66.67	95.70%
2019PM_10348	contig12697.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.65E-13	66.67	95.70%
2019PM_10348	contig13130.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.65E-13	66.67	95.70%
2019PM_10348	contig13790.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.65E-13	66.67	95.70%
2019PM_10348	contig38675.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.65E-13	66.67	95.70%
2019PM_10348	contig38675.2	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.65E-13	66.67	95.70%
2019PM_2915	contig195280.1	---NA---	5.70E-15	62.32	100.00%
2019PM_7837	contig55171.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.65E-13	66.67	95.70%
2019PM_4158	contig75326.1	---NA---	9.53E-13	66.67	95.70%
2019PM_3214	contig2758.1	rna-directed dna polymerase from mobile element	1.60E-10	72.46	90.00%
2019PM_3214	contig4605.1	transcription elongation factor SPT4|GO:0005634 GO:0003746 GO:0008270 GO:0006351 GO:0006414 GO:0032786	1.60E-10	72.46	90.00%
2019PM_3214	contig4911.1		1.60E-10	72.46	90.00%
2019PM_3214	contig4911.2		1.60E-10	72.46	90.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_3214	contig4911.3	transcription elongation factor SPT4|GO:0005634 GO:0003746 GO:0008270 GO:0006351 GO:0006414 GO:0032786	1.60E-10	72.46	90.00%
2019PM_3214	contig9497.1	RNA-directed DNA polymerase from mobile element jockey-like	1.60E-10	72.46	90.00%
2019PM_3214	contig18302.1	RNA-directed DNA polymerase from mobile element jockey-	1.60E-10	72.46	90.00%
2019PM_3214	contig23768.1	RNA-directed DNA polymerase from mobile element jockey-like	1.60E-10	72.46	90.00%
2019PM_3214	contig32468.1	RNA-directed DNA polymerase from mobile element jockey-	1.60E-10	72.46	90.00%
2019PM_3214	contig56529.1	RNA-directed DNA polymerase from mobile element jockey-like	1.60E-10	72.46	90.00%
2019PM_3214	contig56529.2	RNA-directed DNA polymerase from mobile element jockey-like	1.60E-10	72.46	90.00%
2019PM_3214	contig56963.1	RNA-directed DNA polymerase from mobile element jockey-like	1.60E-10	72.46	90.00%
2019PM_3214	contig90807.1	PREDICTED: uncharacterized protein LOC108681254	1.60E-10	72.46	90.00%
2019PM_3214	contig95778.1	---NA---	1.60E-10	72.46	90.00%
2019PM_3214	contig98017.1	RNA-directed DNA polymerase from mobile element jockey-like	1.60E-10	72.46	90.00%
2019PM_3214	contig121761.1	---NA---	1.60E-10	72.46	90.00%
2019PM_3214	contig196289.1	RNA-directed DNA polymerase from mobile element jockey-like	1.60E-10	72.46	90.00%
2019PM_3406	contig6357.1	Gag-Pol poly	2.06E-09	73.91	88.20%
2019PM_4418	contig107538.1	---NA---	2.65E-13	63.77	97.70%
2019PM_6677	contig98880.1	---NA---	7.37E-14	63.77	97.70%
2019PM_6223	contig1026.1	PREDICTED: uncharacterized protein CG43867-like	7.37E-14	63.77	97.70%
2019PM_10471	contig11118.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig11122.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig11294.1	mitochondrial inner membrane protease ATP23 homolog|GO:0003964 GO:0004222 GO:0006278 GO:0006508	9.53E-13	65.22	95.60%
2019PM_10471	contig11294.2	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig12697.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig13130.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig13790.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig13854.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig14754.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig19357.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig19357.2	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig21122.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig26512.1	endonuclease-reverse transcriptase	9.53E-13	65.22	95.60%
2019PM_10471	contig30774.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig36317.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig36423.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig39402.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_10471	contig43349.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig77924.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig80033.1	outcast ele5 orf1 -h 1e-40 -j|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig82623.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig83288.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig91308.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig91465.1	reverse transcriptase, partial|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig118789.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig11118.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig11122.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig11294.1	mitochondrial inner membrane protease ATP23 homolog|GO:0003964 GO:0004222 GO:0006278 GO:0006508	9.53E-13	65.22	95.60%
2019PM_10471	contig11294.2	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig12697.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig13130.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig13790.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig13854.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig14754.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig19357.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig19357.2	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig21122.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig26512.1	endonuclease-reverse transcriptase	9.53E-13	65.22	95.60%
2019PM_10471	contig30774.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig36317.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig36423.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig39402.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig43349.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig77924.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig80033.1	outcast ele5 orf1 -h 1e-40 -j|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig82623.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig83288.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig91308.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig91465.1	reverse transcriptase, partial|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10471	contig118789.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_10348	contig43349.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_10348	contig55171.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.53E-13	65.22	95.60%
2019PM_6677	contig153857.1	---NA---	5.74E-10	71.01	89.80%
2019PM_3406	contig6614.1	Gag-Pol poly	7.42E-09	72.46	88.00%
2019PM_3406	contig6721.1	Gag-Pol poly	7.42E-09	72.46	88.00%
2019PM_3406	contig6721.2	Gag-Pol poly	7.42E-09	72.46	88.00%
2019PM_4158	contig75326.1	---NA---	4.44E-11	66.67	93.50%
2019PM_7333	contig14656.2	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	62.32	97.70%
2019PM_2915	contig195280.1	---NA---	2.65E-13	62.32	97.70%
2019PM_6919	contig103275.1	---NA---	7.37E-14	59.42	100.00%
2019PM_4656	contig195082.1	---NA---	7.37E-14	59.42	100.00%
2019PM_8986	contig37734.1	crustacean female sex	7.37E-14	59.42	100.00%
2019PM_9986	contig114623.1	Retrovirus-related Pol poly from type-2 retrotransposable element R2DM	3.43E-12	63.77	95.50%
2019PM_4418	contig40478.1	---NA---	3.43E-12	60.87	97.60%
2019PM_4418	contig62653.1	---NA---	3.43E-12	60.87	97.60%
2019PM_4418	contig168629.1	---NA---	3.43E-12	60.87	97.60%
2019PM_4418	contig40478.1	---NA---	3.43E-12	60.87	97.60%
2019PM_4418	contig62653.1	---NA---	3.43E-12	60.87	97.60%
2019PM_4418	contig168629.1	---NA---	3.43E-12	60.87	97.60%
2019PM_8676	contig166363.1	reverse transcriptase, partial|GO:0003964 GO:0006278	9.53E-13	60.87	97.60%
2019PM_6872	contig3739.1	RNA-directed DNA polymerase from mobile element jockey-like	9.53E-13	60.87	97.60%
2019PM_6872	contig11041.1	RNA-directed DNA polymerase from mobile element	9.53E-13	60.87	97.60%
2019PM_6872	contig16138.1	RNA-directed DNA polymerase from mobile element	9.53E-13	60.87	97.60%
2019PM_6872	contig16908.1	RNA-directed DNA polymerase from mobile element	9.53E-13	60.87	97.60%
2019PM_6872	contig17322.1	RNA-directed DNA polymerase from mobile element	9.53E-13	60.87	97.60%
2019PM_6872	contig27267.1	RNA-directed DNA polymerase from mobile element jockey-like	9.53E-13	60.87	97.60%
2019PM_6872	contig36930.1	RNA-directed DNA polymerase from mobile element	9.53E-13	60.87	97.60%
2019PM_6872	contig46671.1	RNA-directed DNA polymerase from mobile element|GO:0003824 GO:0005488 GO:0044260 GO:0090304	9.53E-13	60.87	97.60%
2019PM_6872	contig50315.1	RNA-directed DNA polymerase from mobile element jockey-like	9.53E-13	60.87	97.60%
2019PM_6872	contig53029.1	RNA-directed DNA polymerase from mobile element	9.53E-13	60.87	97.60%
2019PM_6872	contig63961.1	RNA-directed DNA polymerase from mobile element jockey-like	9.53E-13	60.87	97.60%
2019PM_6872	contig70581.1	RNA-directed DNA polymerase from mobile element	9.53E-13	60.87	97.60%
2019PM_6872	contig76559.1	RNA-directed DNA polymerase from mobile element jockey-like	9.53E-13	60.87	97.60%
2019PM_6872	contig98509.1	RNA-directed DNA polymerase from mobile element jockey-like	9.53E-13	60.87	97.60%
2019PM_6872	contig108301.1	RNA-directed DNA polymerase from mobile element jockey-like	9.53E-13	60.87	97.60%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_6872	contig117821.1	RNA-directed DNA polymerase from mobile element jockey-like	9.53E-13	60.87	97.60%
2019PM_6872	contig133003.1	RNA-directed DNA polymerase from transposon X-	9.53E-13	60.87	97.60%
2019PM_7333	contig1573.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig4558.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig10680.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig11381.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig13827.1	cytochrome P450 2L1-like|GO:0016020 GO:0043231 GO:0005488 GO:0016491	2.65E-13	57.97	100.00%
2019PM_7333	contig13915.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig14188.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig14188.2	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig16088.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig18688.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig21644.1	CP2L1_PANAR ame: Full=Cytochrome P450 2L1 ame: Full=CYPIIL1|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig26020.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig27233.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig28431.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig31222.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig31561.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_7333	contig63605.1	cytochrome P450 2L1-like|GO:0005783 GO:0016020 GO:0044422 GO:0004497 GO:0016705 GO:0046872	2.65E-13	57.97	100.00%
2019PM_6211	contig42801.1	---NA---	5.74E-10	66.67	91.30%
2019PM_6211	contig126492.1	---NA---	5.74E-10	66.67	91.30%
2019PM_2465	contig1232.1	partitioning defective 3 homolog isoform X3	2.65E-13	57.97	100.00%
2019PM_2465	contig1232.2	pdz domain-containing	2.65E-13	57.97	100.00%
2019PM_2465	contig1249.1	Partitioning defective 3	2.65E-13	57.97	100.00%
2019PM_2465	contig1794.1	partitioning defective 3 homolog isoform X5	2.65E-13	57.97	100.00%
2019PM_2465	contig2364.1	Partitioning defective 3	2.65E-13	57.97	100.00%
2019PM_2465	contig2556.1	partitioning defective 3 homolog isoform X5	2.65E-13	57.97	100.00%
2019PM_2465	contig3132.1	Partitioning defective 3	2.65E-13	57.97	100.00%
2019PM_2465	contig5397.1	partitioning defective 3 homolog isoform X1	2.65E-13	57.97	100.00%
2019PM_2465	contig5408.1	partitioning defective 3 homolog isoform X5	2.65E-13	57.97	100.00%
2019PM_2465	contig5484.1	partitioning defective 3 homolog isoform X5	2.65E-13	57.97	100.00%
2019PM_2465	contig6077.1	Partitioning defective 3	2.65E-13	57.97	100.00%
2019PM_2465	contig18354.1	partitioning defective 3 homolog B isoform X2	2.65E-13	57.97	100.00%
2019PM_2465	contig3496.1	partitioning defective 3 homolog isoform X2	2.65E-13	57.97	100.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_6211	contig79842.1	PREDICTED: uncharacterized protein LOC103506659	7.42E-09	68.12	89.40%
2019PM_4656	contig195082.1	---NA---	3.43E-12	59.42	97.60%
2019PM_3214	contig33912.1	Gag-Pol poly	1.60E-10	63.77	93.20%
2019PM_8676	contig197698.1	---NA---	9.53E-13	56.52	100.00%
2019PM_8676	contig31079.1	endonuclease-reverse transcriptase|GO:0003824 GO:0043170 GO:0044238	4.44E-11	60.87	95.20%
2019PM_8676	contig36423.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.44E-11	60.87	95.20%
2019PM_8676	contig91308.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	4.44E-11	60.87	95.20%
2019PM_8676	contig103281.1	reverse transcriptase, partial|GO:0003964 GO:0006278	4.44E-11	60.87	95.20%
2019PM_5879	contig23832.1	---NA---	4.44E-11	60.87	95.20%
2019PM_5879	contig133091.1	---NA---	4.44E-11	60.87	95.20%
2019PM_6872	contig11251.1	RNA-directed DNA polymerase from mobile element jockey-like	4.44E-11	60.87	95.20%
2019PM_6872	contig11460.1	RNA-directed DNA polymerase from mobile element jockey-like	4.44E-11	60.87	95.20%
2019PM_6872	contig12501.1	RNA-directed DNA polymerase from mobile element jockey-like	4.44E-11	60.87	95.20%
2019PM_4418	contig187381.1	---NA---	2.06E-09	62.32	93.00%
2019PM_4418	contig187381.1	---NA---	2.06E-09	62.32	93.00%
2019PM_1266	contig111647.1	---NA---	3.43E-12	55.07	100.00%
2019PM_3480	contig26801.1	sulfotransferase 1C4-like	3.43E-12	55.07	100.00%
2019PM_3480	contig35357.1	sulfotransferase 1C4-like	3.43E-12	55.07	100.00%
2019PM_8483	contig82348.1	---NA---	3.43E-12	55.07	100.00%
2019PM_5685	contig28515.1	coproporphyrinigen III|GO:0003824	3.43E-12	55.07	100.00%
2019PM_6730	contig4911.1	transcription elongation factor SPT4|GO:0005634 GO:0003746 GO:0008270 GO:0006351 GO:0006414 GO:0032786	3.43E-12	55.07	100.00%
2019PM_6730	contig4911.2		3.43E-12	55.07	100.00%
2019PM_6730	contig4911.3		3.43E-12	55.07	100.00%
2019PM_6730	contig8125.1	rna-directed dna polymerase from mobile element jockey	3.43E-12	55.07	100.00%
2019PM_6730	contig8125.2	RNA-directed DNA polymerase from mobile element jockey	3.43E-12	55.07	100.00%
2019PM_6730	contig9497.1	RNA-directed DNA polymerase from mobile element jockey-like	3.43E-12	55.07	100.00%
2019PM_6730	contig11723.1	RNA-directed DNA polymerase from mobile element jockey	3.43E-12	55.07	100.00%
2019PM_6730	contig18302.1	RNA-directed DNA polymerase from mobile element jockey-	3.43E-12	55.07	100.00%
2019PM_6730	contig20167.1	rna-directed dna polymerase from mobile element	3.43E-12	55.07	100.00%
2019PM_6730	contig32468.1	RNA-directed DNA polymerase from mobile element jockey-	3.43E-12	55.07	100.00%
2019PM_6730	contig35556.1	RNA-directed DNA polymerase from mobile element jockey-	3.43E-12	55.07	100.00%
2019PM_6730	contig38733.1	RNA-directed DNA polymerase from mobile element jockey-	3.43E-12	55.07	100.00%
2019PM_6730	contig45666.1	RNA-directed DNA polymerase from mobile element jockey-like	3.43E-12	55.07	100.00%
2019PM_6730	contig73604.1	rna-directed dna polymerase from mobile element jockey	3.43E-12	55.07	100.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_6730	contig74646.1	reverse transcriptase	3.43E-12	55.07	100.00%
2019PM_6730	contig90807.1	PREDICTED: uncharacterized protein LOC108681254	3.43E-12	55.07	100.00%
2019PM_2197	contig103332.1	---NA---	4.44E-11	59.42	95.30%
2019PM_6449	contig2742.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.67E-08	63.77	90.90%
2019PM_7837	contig221839.1	reverse transcriptase, partial|GO:0003964 GO:0006278	1.60E-10	59.42	95.10%
2019PM_3530	contig144358.1	endonuclease-reverse transcriptase	1.60E-10	59.42	95.10%
2019PM_4418	contig88394.1	---NA---	1.60E-10	56.52	97.40%
2019PM_4418	contig88394.1	---NA---	1.60E-10	56.52	97.40%
2019PM_8676	contig166363.1	reverse transcriptase, partial|GO:0003964 GO:0006278	4.44E-11	56.52	97.40%
2019PM_8676	contig197698.1	---NA---	4.44E-11	56.52	97.40%
2019PM_3043	contig25186.1	chitinase 1|GO:0005576 GO:0016021 GO:0004568 GO:0008061 GO:0005975 GO:0006032	1.23E-11	53.62	100.00%
2019PM_3043	contig25186.1	chitinase 1|GO:0005576 GO:0016021 GO:0004568 GO:0008061 GO:0005975 GO:0006032	1.23E-11	53.62	100.00%
2019PM_3480	contig26801.1	sulfotransferase 1C4-like	1.60E-10	55.07	97.40%
2019PM_3480	contig35357.1	sulfotransferase 1C4-like	1.60E-10	55.07	97.40%
2019PM_2197	contig103332.1	---NA---	2.06E-09	59.42	93.00%
2019PM_5685	contig23161.1	coproporphyrinigen III|GO:0016491	1.60E-10	55.07	97.40%
2019PM_5685	contig23161.2	radical S-adenosyl methionine domain-containing mitochondrial-	1.60E-10	55.07	97.40%
2019PM_5685	contig29267.1	coproporphyrinigen III	1.60E-10	55.07	97.40%
2019PM_5685	contig30260.1	coproporphyrinigen III|GO:0003824	1.60E-10	55.07	97.40%
2019PM_5685	contig39629.1	radical S-adenosyl methionine domain-containing mitochondrial-|GO:0003824	1.60E-10	55.07	97.40%
2019PM_6730	contig11723.1	RNA-directed DNA polymerase from mobile element jockey	1.60E-10	55.07	97.40%
2019PM_6730	contig20167.1	rna-directed dna polymerase from mobile element	1.60E-10	55.07	97.40%
2019PM_6730	contig35556.1	RNA-directed DNA polymerase from mobile element jockey-	1.60E-10	55.07	97.40%
2019PM_6730	contig38733.1	RNA-directed DNA polymerase from mobile element jockey-	1.60E-10	55.07	97.40%
2019PM_6730	contig45666.1	RNA-directed DNA polymerase from mobile element jockey-like	1.60E-10	55.07	97.40%
2019PM_6730	contig73604.1	rna-directed dna polymerase from mobile element jockey	1.60E-10	55.07	97.40%
2019PM_6730	contig74646.1	reverse transcriptase	1.60E-10	55.07	97.40%
2019PM_3043	contig15586.1	reverse transcriptase	2.67E-08	59.42	92.70%
2019PM_3043	contig56983.1	---NA---	2.67E-08	59.42	92.70%
2019PM_3043	contig76683.1	hypothetical protein, partial	2.67E-08	59.42	92.70%
2019PM_3043	contig103428.1	---NA---	2.67E-08	59.42	92.70%
2019PM_6449	contig191278.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	2.06E-09	56.52	94.90%
2019PM_6365	contig146529.1	---NA---	2.06E-09	56.52	94.90%
2019PM_7837	contig211386.1	reverse transcriptase, partial|GO:0003964 GO:0006278	5.74E-10	53.62	97.30%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_1316	contig1983.1	rna-directed dna polymerase from mobile element jockey	1.60E-10	50.72	100.00%
2019PM_1316	contig2399.1	rna-directed dna polymerase from mobile element jockey	1.60E-10	50.72	100.00%
2019PM_1316	contig3954.1	rna-directed dna polymerase from mobile element jockey	1.60E-10	50.72	100.00%
2019PM_1316	contig21017.1	---NA---	1.60E-10	50.72	100.00%
2019PM_1316	contig58135.1	---NA---	1.60E-10	50.72	100.00%
2019PM_1316	contig95778.1	---NA---	1.60E-10	50.72	100.00%
2019PM_1316	contig121761.1	---NA---	1.60E-10	50.72	100.00%
2019PM_6872	contig2136.1	RNA-directed DNA polymerase from mobile element|GO:0003824 GO:0044260 GO:0090304	1.60E-10	50.72	100.00%
2019PM_6872	contig2761.2	rna-directed dna polymerase from mobile element|GO:0016020 GO:0003824 GO:0005488 GO:0044260 GO:0090304	1.60E-10	50.72	100.00%
2019PM_6872	contig6738.1	RNA-directed DNA polymerase from mobile element jockey-like|GO:0003824 GO:0005488 GO:0044260 GO:0090304	1.60E-10	50.72	100.00%
2019PM_6872	contig15673.1	RNA-directed DNA polymerase from mobile element jockey-like	1.60E-10	50.72	100.00%
2019PM_6872	contig39038.1	RNA-directed DNA polymerase from mobile element	1.60E-10	50.72	100.00%
2019PM_9893	contig133670.1	---NA---	7.42E-09	55.07	94.70%
2019PM_3043	contig143785.1	---NA---	1.24E-06	59.42	90.20%
2019PM_1518	contig17794.1	PDF receptor isoform X1|GO:0016020 GO:0004888 GO:0007186	5.74E-10	49.28	100.00%
2019PM_4011	contig149691.1	---NA---	5.74E-10	49.28	100.00%
2019PM_3043	contig56983.2	---NA---	3.45E-07	56.52	92.30%
2019PM_3092	contig5699.1	---NA---	2.67E-08	53.62	94.60%
2019PM_3092	contig5699.1	---NA---	2.67E-08	53.62	94.60%
2019PM_1316	contig121761.1	---NA---	7.42E-09	50.72	97.10%
2019PM_1316	contig2399.1	rna-directed dna polymerase from mobile element jockey	7.42E-09	50.72	97.10%
2019PM_1316	contig3954.1	rna-directed dna polymerase from mobile element jockey	7.42E-09	50.72	97.10%
2019PM_1316	contig21017.1	---NA---	7.42E-09	50.72	97.10%
2019PM_1316	contig58135.1	---NA---	7.42E-09	50.72	97.10%
2019PM_6001	contig162797.1	---NA---	9.60E-08	55.07	92.50%
2019PM_3043	contig178925.1	---NA---	1.24E-06	55.07	92.10%
2019PM_1476	contig69267.1	Gag-Pol poly	9.60E-08	52.17	94.40%
2019PM_1476	contig74902.1	RNA-directed DNA polymerase from mobile element	9.60E-08	52.17	94.40%
2019PM_1476	contig69267.1	Gag-Pol poly	9.60E-08	52.17	94.40%
2019PM_1476	contig74902.1	RNA-directed DNA polymerase from mobile element	9.60E-08	52.17	94.40%
2019PM_3092	contig50860.1	---NA---	9.60E-08	52.17	94.40%
2019PM_3092	contig52616.1	---NA---	9.60E-08	52.17	94.40%
2019PM_3092	contig55134.1	---NA---	9.60E-08	52.17	94.40%
2019PM_3092	contig82868.1	---NA---	9.60E-08	52.17	94.40%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_3092	contig85958.1	---NA---	9.60E-08	52.17	94.40%
2019PM_3092	contig50860.1	---NA---	9.60E-08	52.17	94.40%
2019PM_3092	contig52616.1	---NA---	9.60E-08	52.17	94.40%
2019PM_3092	contig55134.1	---NA---	9.60E-08	52.17	94.40%
2019PM_3092	contig82868.1	---NA---	9.60E-08	52.17	94.40%
2019PM_3092	contig85958.1	---NA---	9.60E-08	52.17	94.40%
2019PM_4209	contig194757.1	reverse transcriptase, partial	7.42E-09	46.38	100.00%
2019PM_2673	contig32503.1	PREDICTED: uncharacterized protein LOC108669347	7.42E-09	46.38	100.00%
2019PM_2673	contig34695.1	PREDICTED: uncharacterized protein LOC108669347	7.42E-09	46.38	100.00%
2019PM_2673	contig35263.1	PREDICTED: uncharacterized protein LOC108669347	7.42E-09	46.38	100.00%
2019PM_2673	contig35729.1	PREDICTED: uncharacterized protein LOC108669347	7.42E-09	46.38	100.00%
2019PM_1552	contig40493.1	---NA---	7.42E-09	46.38	100.00%
2019PM_5879	contig170342.1	---NA---	2.67E-08	49.28	97.10%
2019PM_3043	contig15586.1	reverse transcriptase	7.42E-09	46.38	100.00%
2019PM_3043	contig56983.1	---NA---	7.42E-09	46.38	100.00%
2019PM_3043	contig76683.1	hypothetical protein, partial	7.42E-09	46.38	100.00%
2019PM_425	contig12597.1	PREDICTED: neurotrypsin-like	7.42E-09	46.38	100.00%
2019PM_425	contig12597.2	PREDICTED: neurotrypsin-like	7.42E-09	46.38	100.00%
2019PM_6001	contig23877.1	---NA---	3.45E-07	53.62	92.30%
2019PM_6001	contig30481.1	---NA---	3.45E-07	53.62	92.30%
2019PM_6001	contig34835.1	---NA---	3.45E-07	53.62	92.30%
2019PM_3848	contig142188.1	---NA---	3.45E-07	50.72	94.30%
2019PM_9859	contig89248.1	---NA---	2.67E-08	44.93	100.00%
2019PM_992	contig17419.1	trypsin-1-like|GO:0016787	2.67E-08	44.93	100.00%
2019PM_992	contig17419.2	trypsin-1-like|GO:0016787	2.67E-08	44.93	100.00%
2019PM_3214	contig58135.1	---NA---	2.67E-08	44.93	100.00%
2019PM_3043	contig14368.1	---NA---	9.60E-08	47.83	97.00%
2019PM_3043	contig95106.1	cuticle 19-like	9.60E-08	47.83	97.00%
2019PM_3043	contig104011.1	---NA---	9.60E-08	47.83	97.00%
2019PM_3043	contig113197.1	cuticle 19-like	9.60E-08	47.83	97.00%
2019PM_3043	contig120750.1	---NA---	9.60E-08	47.83	97.00%
2019PM_3043	contig132422.1	---NA---	9.60E-08	47.83	97.00%
2019PM_3043	contig150981.1	AF445324_1hypothetical cob intron 4|GO:0016021	9.60E-08	47.83	97.00%
2019PM_3043	contig162672.1	---NA---	9.60E-08	47.83	97.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_3043	contig169245.1	AF445324_1hypothetical cob intron 4	9.60E-08	47.83	97.00%
2019PM_3043	contig171974.1	---NA---	9.60E-08	47.83	97.00%
2019PM_4209	contig138721.1	endonuclease-reverse transcriptase|GO:0003964 GO:0006278	9.60E-08	43.48	100.00%
2019PM_1476	contig151736.1	RNA-directed DNA polymerase from transposon BS	3.45E-07	46.38	97.00%
2019PM_1476	contig151736.1	RNA-directed DNA polymerase from transposon BS	3.45E-07	46.38	97.00%
2019PM_3528	contig22241.1	Serine threonine- kinase PAK|GO:0005737 GO:0004702 GO:0005524 GO:0000278 GO:0006468 GO:0007409 GO:0030036	9.60E-08	43.48	100.00%
2019PM_3043	contig105236.1	---NA---	9.60E-08	43.48	100.00%
2019PM_3043	contig119426.1	---NA---	9.60E-08	43.48	100.00%
2019PM_3043	contig133508.1	---NA---	9.60E-08	43.48	100.00%
2019PM_3043	contig147437.1	---NA---	9.60E-08	43.48	100.00%
2019PM_3043	contig185302.1	---NA---	9.60E-08	43.48	100.00%
2019PM_3043	contig192064.1	---NA---	9.60E-08	43.48	100.00%
2019PM_3406	contig11495.1	---NA---	3.45E-07	46.38	96.90%
2019PM_3406	contig11495.1	---NA---	3.45E-07	46.38	96.90%
2019PM_8143	contig38946.1	---NA---	3.45E-07	42.03	100.00%
2019PM_8143	contig44602.1	---NA---	3.45E-07	42.03	100.00%
2019PM_3406	contig8771.1	transcription elongation factor SPT4|GO:0005634 GO:0003746 GO:0008270 GO:0006351 GO:0006414 GO:0032786	3.45E-07	42.03	100.00%
2019PM_3406	contig8771.2	non-ltr retrotransposon r1bmks orf2	3.45E-07	42.03	100.00%
2019PM_3406	contig15362.1	---NA---	3.45E-07	42.03	100.00%
2019PM_3406	contig15724.1	Gag-Pol poly|GO:0003824 GO:0044260 GO:0090304	3.45E-07	42.03	100.00%
2019PM_3406	contig49839.1	pol-like protein, partial	3.45E-07	42.03	100.00%
2019PM_3406	contig56246.1	---NA---	3.45E-07	42.03	100.00%
2019PM_3406	contig8771.1	transcription elongation factor SPT4|GO:0005634 GO:0003746 GO:0008270 GO:0006351 GO:0006414 GO:0032786	3.45E-07	42.03	100.00%
2019PM_3406	contig8771.2	non-ltr retrotransposon r1bmks orf2	3.45E-07	42.03	100.00%
2019PM_3406	contig15362.1	---NA---	3.45E-07	42.03	100.00%
2019PM_3406	contig15724.1	Gag-Pol poly|GO:0003824 GO:0044260 GO:0090304	3.45E-07	42.03	100.00%
2019PM_3406	contig49839.1	pol-like protein, partial	3.45E-07	42.03	100.00%
2019PM_3406	contig56246.1	---NA---	3.45E-07	42.03	100.00%
2019PM_1639	contig116796.1	---NA---	3.45E-07	42.03	100.00%
2019PM_1476	contig2758.1	rna-directed dna polymerase from mobile element	3.45E-07	42.03	100.00%
2019PM_1476	contig3120.1	rna-directed dna polymerase from mobile element jockey	3.45E-07	42.03	100.00%
2019PM_1476	contig6357.1	Gag-Pol poly	3.45E-07	42.03	100.00%
2019PM_1476	contig6478.1	rna-directed dna polymerase from mobile element jockey	3.45E-07	42.03	100.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_1476	contig6614.1	Gag-Pol poly	3.45E-07	42.03	100.00%
2019PM_1476	contig6721.1	Gag-Pol poly	3.45E-07	42.03	100.00%
2019PM_1476	contig6721.2	Gag-Pol poly	3.45E-07	42.03	100.00%
2019PM_1476	contig9497.1	RNA-directed DNA polymerase from mobile element jockey-like	3.45E-07	42.03	100.00%
2019PM_1476	contig11857.1	adenylate cyclase type 3	3.45E-07	42.03	100.00%
2019PM_1476	contig13459.1	histone-lysine N-methyltransferase SETMAR-like	3.45E-07	42.03	100.00%
2019PM_1476	contig16012.1	histone-lysine N-methyltransferase SETMAR-like	3.45E-07	42.03	100.00%
2019PM_1476	contig43327.1	---NA---	3.45E-07	42.03	100.00%
2019PM_1476	contig45058.1	pol-like protein, partial	3.45E-07	42.03	100.00%
2019PM_1476	contig47066.1	---NA---	3.45E-07	42.03	100.00%
2019PM_1476	contig55643.1	RNA-directed DNA polymerase from mobile element jockey	3.45E-07	42.03	100.00%
2019PM_1476	contig61824.1	---NA---	3.45E-07	42.03	100.00%
2019PM_1476	contig68662.1	---NA---	3.45E-07	42.03	100.00%
2019PM_1476	contig127633.1	---NA---	3.45E-07	42.03	100.00%
2019PM_1476	contig2758.1	rna-directed dna polymerase from mobile element	3.45E-07	42.03	100.00%
2019PM_1476	contig3120.1	rna-directed dna polymerase from mobile element jockey	3.45E-07	42.03	100.00%
2019PM_1476	contig6357.1	Gag-Pol poly	3.45E-07	42.03	100.00%
2019PM_1476	contig6478.1	rna-directed dna polymerase from mobile element jockey	3.45E-07	42.03	100.00%
2019PM_1476	contig6614.1	Gag-Pol poly	3.45E-07	42.03	100.00%
2019PM_1476	contig6721.1	Gag-Pol poly	3.45E-07	42.03	100.00%
2019PM_1476	contig6721.2	Gag-Pol poly	3.45E-07	42.03	100.00%
2019PM_1476	contig9497.1	RNA-directed DNA polymerase from mobile element jockey-like	3.45E-07	42.03	100.00%
2019PM_1476	contig11857.1	adenylate cyclase type 3	3.45E-07	42.03	100.00%
2019PM_1476	contig13459.1	histone-lysine N-methyltransferase SETMAR-like	3.45E-07	42.03	100.00%
2019PM_1476	contig16012.1	histone-lysine N-methyltransferase SETMAR-like	3.45E-07	42.03	100.00%
2019PM_1476	contig43327.1	---NA---	3.45E-07	42.03	100.00%
2019PM_1476	contig45058.1	pol-like protein, partial	3.45E-07	42.03	100.00%
2019PM_1476	contig47066.1	---NA---	3.45E-07	42.03	100.00%
2019PM_1476	contig55643.1	RNA-directed DNA polymerase from mobile element jockey	3.45E-07	42.03	100.00%
2019PM_1476	contig61824.1	---NA---	3.45E-07	42.03	100.00%
2019PM_1476	contig68662.1	---NA---	3.45E-07	42.03	100.00%
2019PM_1476	contig127633.1	---NA---	3.45E-07	42.03	100.00%
2019PM_6488	contig80930.1	Sorting nexin-17|GO:0035091	3.45E-07	42.03	100.00%
2019PM_6488	contig80930.1	Sorting nexin-17|GO:0035091	3.45E-07	42.03	100.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_9859	contig30652.1	---NA---	3.45E-07	42.03	100.00%
2019PM_9859	contig30798.1	mediator of RNA polymerase II transcription subunit 13-like	3.45E-07	42.03	100.00%
2019PM_9859	contig54777.1	---NA---	3.45E-07	42.03	100.00%
2019PM_9859	contig54777.2	---NA---	3.45E-07	42.03	100.00%
2019PM_9859	contig80157.1	---NA---	3.45E-07	42.03	100.00%
2019PM_9859	contig83920.1	hypothetical protein ALC62_14810	3.45E-07	42.03	100.00%
2019PM_9859	contig90199.1	---NA---	3.45E-07	42.03	100.00%
2019PM_2281	contig202000.1	---NA---	3.45E-07	42.03	100.00%
2019PM_5879	contig14298.1	lectin B isoform	3.45E-07	42.03	100.00%
2019PM_5879	contig32241.1	---NA---	3.45E-07	42.03	100.00%
2019PM_5879	contig45025.1	---NA---	3.45E-07	42.03	100.00%
2019PM_5879	contig64057.1	---NA---	3.45E-07	42.03	100.00%
2019PM_5879	contig147041.1	---NA---	3.45E-07	42.03	100.00%
2019PM_5879	contig14298.1	lectin B isoform	3.45E-07	42.03	100.00%
2019PM_5879	contig32241.1	---NA---	3.45E-07	42.03	100.00%
2019PM_5879	contig45025.1	---NA---	3.45E-07	42.03	100.00%
2019PM_5879	contig147041.1	---NA---	3.45E-07	42.03	100.00%
2019PM_665	contig4772.1	serine-rich adhesin for platelets-like	3.45E-07	42.03	100.00%
2019PM_665	contig7360.1	serine-rich adhesin for platelets-like	3.45E-07	42.03	100.00%
2019PM_3043	contig213519.1	---NA---	3.45E-07	42.03	100.00%
2019PM_7492	contig70238.1	lysozyme|GO:0003796 GO:0008152	3.45E-07	42.03	100.00%
2019PM_7492	contig107318.1	lysozyme|GO:0003796 GO:0008152	3.45E-07	42.03	100.00%
2019PM_3043	contig150981.1	AF445324_1hypothetical cob intron 4|GO:0016021	1.24E-06	44.93	96.80%
2019PM_3043	contig162672.1	---NA---	1.24E-06	44.93	96.80%
2019PM_3043	contig169245.1	AF445324_1hypothetical cob intron 4	1.24E-06	44.93	96.80%
2019PM_3043	contig171974.1	---NA---	1.24E-06	44.93	96.80%
2019PM_3043	contig174071.1	---NA---	1.24E-06	44.93	96.80%
2019PM_9618	contig196269.1	---NA---	1.24E-06	40.58	100.00%
2019PM_4354	contig1189.1	Transposon Ty3-I Gag-Pol poly	1.24E-06	40.58	100.00%
2019PM_5879	contig38430.1	---NA---	1.24E-06	40.58	100.00%
2019PM_5879	contig107437.1	---NA---	1.24E-06	40.58	100.00%
2019PM_5879	contig113814.1	---NA---	1.24E-06	40.58	100.00%
2019PM_5879	contig127989.1	---NA---	1.24E-06	40.58	100.00%
2019PM_5879	contig194195.1	---NA---	1.24E-06	40.58	100.00%

Appendix 3.6 (cont.)

SNP_ID	Contig matched	Description	E Value	Query (%) coverage	% Pairwise Identity
2019PM_5879	contig38430.1	---NA---	1.24E-06	40.58	100.00%
2019PM_5879	contig107437.1	---NA---	1.24E-06	40.58	100.00%
2019PM_5879	contig113814.1	---NA---	1.24E-06	40.58	100.00%
2019PM_5879	contig127989.1	---NA---	1.24E-06	40.58	100.00%
2019PM_5879	contig194195.1	---NA---	1.24E-06	40.58	100.00%
2019PM_3043	contig133508.1	---NA---	1.24E-06	40.58	100.00%
2019PM_3043	contig141082.1	---NA---	1.24E-06	40.58	100.00%
2019PM_3043	contig147437.1	---NA---	1.24E-06	40.58	100.00%
2019PM_3043	contig170482.1	---NA---	1.24E-06	40.58	100.00%
2019PM_3043	contig185302.1	---NA---	1.24E-06	40.58	100.00%
2019PM_3043	contig186285.1	---NA---	1.24E-06	40.58	100.00%
2019PM_3043	contig192064.1	---NA---	1.24E-06	40.58	100.00%

Appendix 3.7 Putative functional annotations for 62 contigs that exhibited 100% pairwise identity to 27 outlier SNPs

SNP_ID	Contig matched	Description	E_value	Query (%) coverage	% Pairwise Identity
2019PM_354	contig9198.1	Serine threonine- kinase D3|GO:0016020 GO:0000166 GO:0004674 GO:0043167 GO:0016310	2.01E-29	100	100
	contig9272.1	serine threonine- kinase D3 isoform X1|GO:0016020 GO:0000166 GO:0004674 GO:0043167 GO:0016310	2.01E-29	100	100
2019PM_7022	contig200831.1	---NA---	2.01E-29	100	100
2019PM_4058	contig92766.1	AChain Apocrustacyanin C1 Crystals|GO:0005615 GO:0031409 GO:0036094 GO:0006810	2.01E-29	100	100
	contig97472.1	AChain Apocrustacyanin C1 Crystals|GO:0005615 GO:0031409 GO:0036094 GO:0006810	2.01E-29	100	100
	contig86923.1	AChain Apocrustacyanin C1 Crystals|GO:0005615 GO:0031409 GO:0036094 GO:0006810	2.01E-29	100	100
	contig94787.1	AChain Apocrustacyanin C1 Crystals|GO:0005615 GO:0031409 GO:0036094 GO:0006810	2.01E-29	100	100
	contig147401.1	AChain Apocrustacyanin C1 Crystals|GO:0005615 GO:0031409 GO:0036094 GO:0006810	2.01E-29	100	100
2019PM_10536	contig2616.1	Neuropilin and tolloid	2.01E-29	100	100
2019PM_6523	contig42167.1	C-type lectin	2.01E-29	100	100
	contig48544.1	---NA---	2.01E-29	100	100
	contig50319.1	---NA---	2.01E-29	100	100
2019PM_1019	contig54930.1	rna-directed dna polymerase from mobile element jockey	2.01E-29	100	100
	contig64187.1	rna-directed dna polymerase from mobile element	2.01E-29	100	100
2019PM_7468	contig1470.1	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	2.01E-29	100	100
	contig1470.2	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	2.01E-29	100	100
	contig1729.1	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	2.01E-29	100	100
	contig1729.2	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	2.01E-29	100	100
	contig2018.1	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	2.01E-29	100	100
	contig2018.2	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	2.01E-29	100	100
	contig2269.1	pleckstrin homology domain-containing family G member 3 isoform X3|GO:0005089 GO:0035023 GO:0043547	2.01E-29	100	100

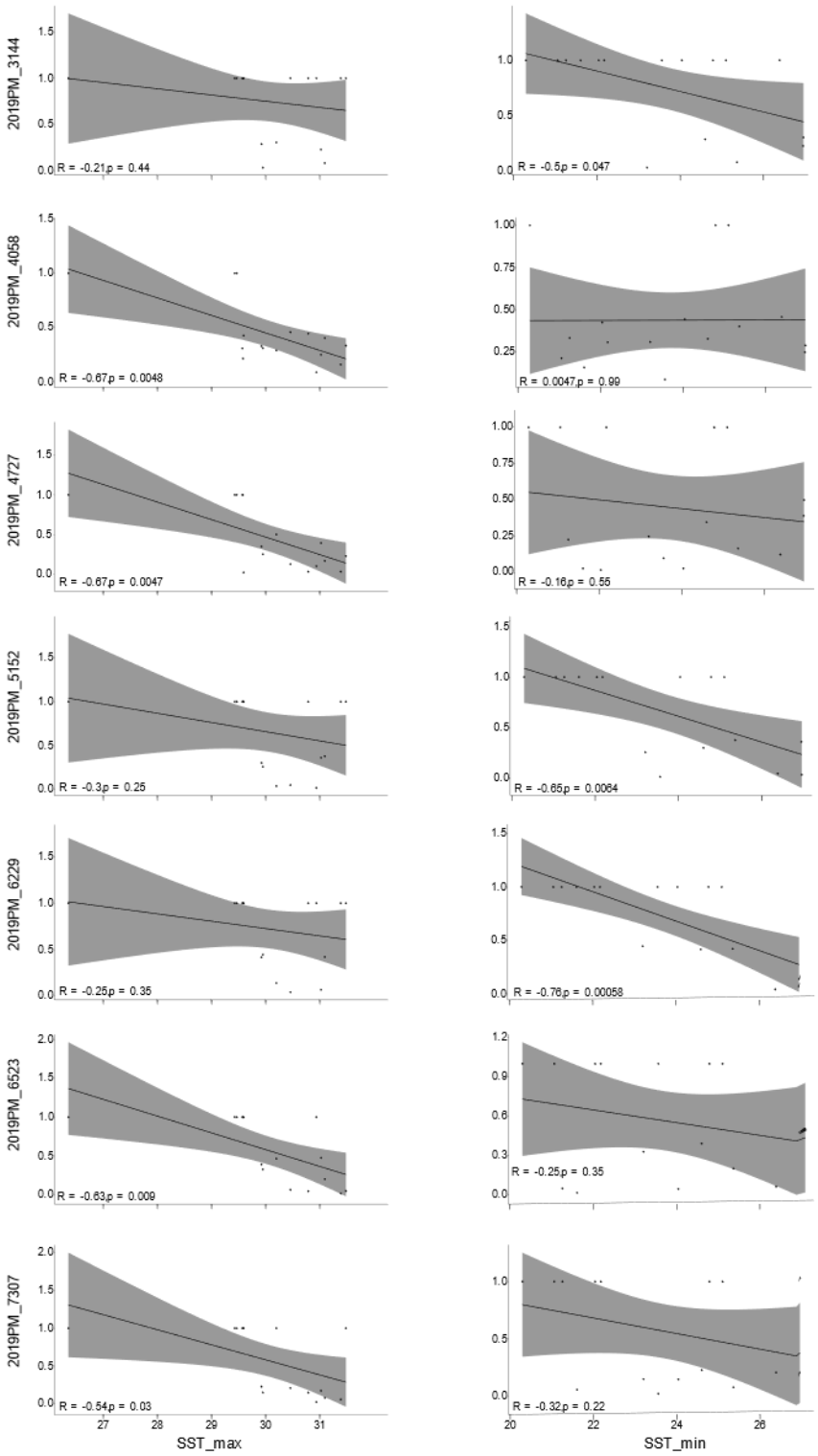
Appendix 3.7 (cont.)

SNP_ID	Contig matched	Description	E_value	Query (%) coverage	% Pairwise Identity
2019PM_7307	contig29609.1	xylose isomerase-like|GO:0005737 GO:0016020 GO:0044238 GO:0071704	2.01E-29	100	100%
	contig8599.1	xylose isomerase-like|GO:0005737 GO:0016020 GO:0044238 GO:0071704	2.01E-29	100	100%
2019PM_7067	contig9015.1	---NA---	2.01E-29	100	100%
	contig13398.1	PREDICTED: uncharacterized protein LOC108669519 isoform X1	2.01E-29	100	100%
	contig13398.2	---NA---	2.01E-29	100	100%
	contig32911.1	PREDICTED: uncharacterized protein LOC108669519 isoform X2	2.01E-29	100	100%
2019PM_9268	contig33948.1	---NA---	2.01E-29	100	100%
2019PM_4727	contig117214.1	zinc C2H2|GO:0003676 GO:0046872	2.01E-29	100	100%
2019PM_941	contig26548.2	ligand of Numb X 2-like|GO:0008270	2.01E-29	100	100%
	contig26548.3	ligand of Numb X 2-like	2.01E-29	100	100%
	contig28846.1	ligand of Numb X 2-like	2.01E-29	100	100%
	contig49350.1	ligand of Numb X 2-like	2.01E-29	100	100%
2019PM_469	contig5079.1	PREDICTED: uncharacterized protein LOC108679759	2.01E-29	100	100%
2019PM_7837	contig170033.1	---NA---	2.01E-29	100	100%
2019PM_9964	contig3323.1	extracellular sulfatase SULF-1 homolog isoform X1|GO:0008449 GO:0008152	2.01E-29	100	100%
2019PM_9056	contig7456.1	cell adhesion molecule 3	2.01E-29	100	100%
	contig16669.1	cell adhesion molecule 3	2.01E-29	100	100%
	contig15435.1	cell adhesion molecule 3	2.01E-29	100	100%
	contig17054.1	tyrosine- kinase receptor TYRO3-like isoform X3	2.01E-29	100	100%
	contig21634.1	cell adhesion molecule 3	2.01E-29	100	100%
	contig39363.1	cell adhesion molecule 3	2.01E-29	100	100%

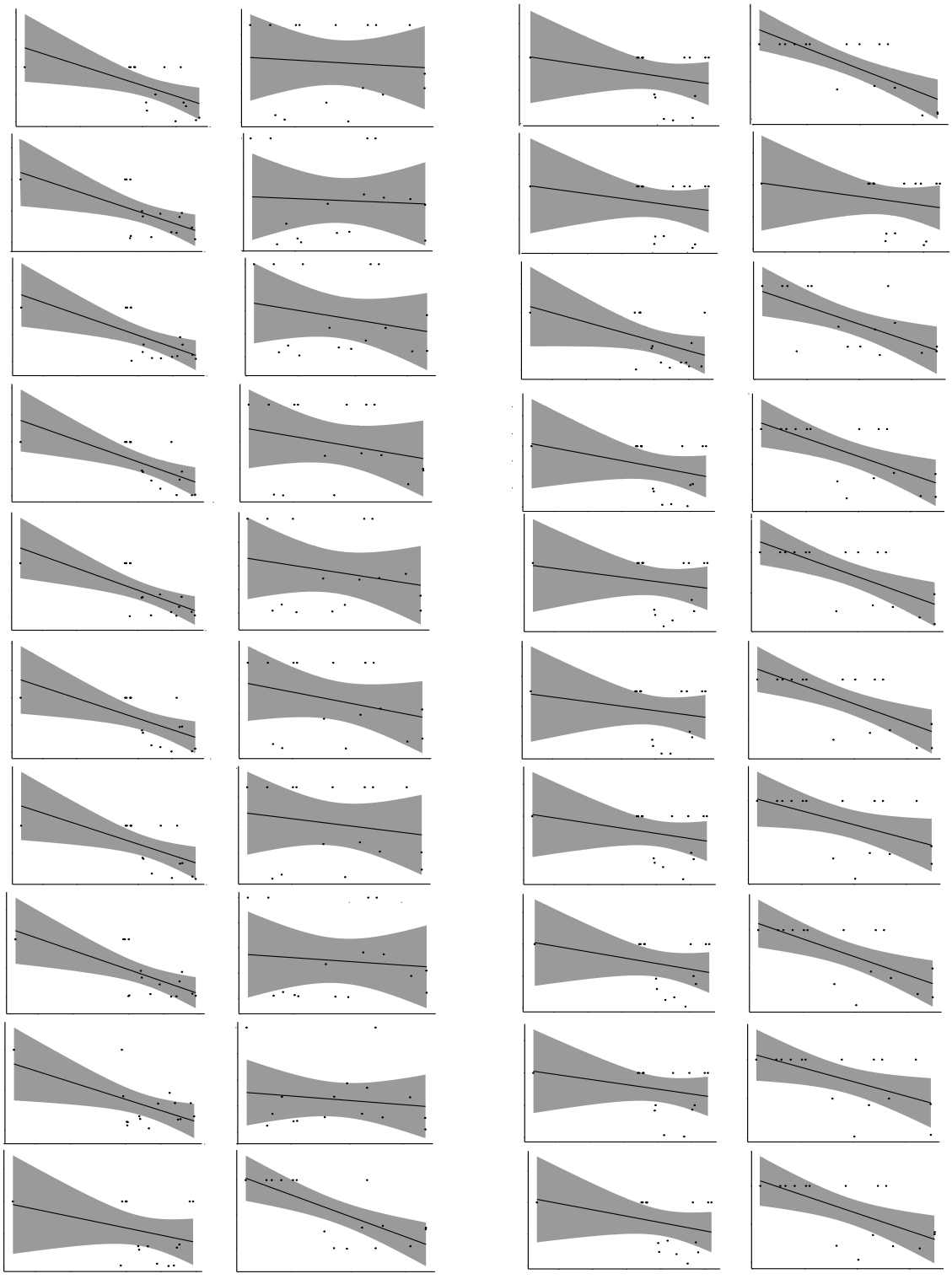
Appendix 3.7 (cont.)

SNP_ID	Contig matched	Description	E_value	Query (%) coverage	% Pairwise Identity
2019PM_6229	contig75648.1	---NA---	2.01E-29	100	100%
	contig82925.1	---NA---	2.01E-29	100	100%
2019PM_6944	contig51401.1	PREDICTED: uncharacterized protein LOC108680202	2.01E-29	100	100%
	contig59928.1	PREDICTED: uncharacterized protein LOC108680202	2.01E-29	100	100%
	contig83041.1	PREDICTED: uncharacterized protein LOC108680202	2.01E-29	100	100%
2019PM_10398	contig134349.1	---NA---	2.01E-29	100	100%
2019PM_3043	contig29545.1	chitinase 1|GO:0005576 GO:0004568 GO:0008061 GO:0006032	2.01E-29	100	100%
2019PM_3144	contig3945.1	PREDICTED: uncharacterized protein LOC108669467	2.01E-29	100	100%
	contig3945.2	PREDICTED: uncharacterized protein LOC108669467	2.01E-29	100	100%
2019PM_1623	contig93906.1	---NA---	2.01E-29	100	100%
2019PM_3050	contig2672.1	isoforms A C F G H-like isoform X7	2.01E-29	100	100%
2019PM_574	contig188.1	bark beetle isoform X1	2.01E-29	100	100%
	contig188.2	bark beetle isoform X1	2.01E-29	100	100%
	contig4754.1	bark beetle isoform X1	2.01E-29	100	100%
2019PM_4073	contig50488.1	---NA---	2.01E-29	100	100%
2019PM_3913	contig15321.1	UDP-glucuronosyltransferase 2B14-like	2.01E-29	100	100%
	contig15708.1	UDP-glucuronosyltransferase 2B14-like	2.01E-29	100	100%
	contig15708.2	UDP-glucuronosyltransferase 2B14-like	2.01E-29	100	100%
2019PM_5152 *	contig34473.1	transcriptional regulatory -like	2.01E-29	100	100%
	contig41161.1	---NA---	2.01E-29	100	100%

* overlaping SNP across both putative adaptive loci identification approaches



Appendix 3.8 Correlation between minor allele frequency (Y axis) and sea surface temprature maximum (SST_max) and minumum (SST_min) for each of the seven transcriptome BLAST identified outlier SNPs. Each plot includes loess smoothing function and 95% confidence intervals (grey area).



Appendix 3.9 Pearson correlations between minor allele frequency and sea surface temperature (maximum and minimum) for the 20 temperatures correlated putatively adaptive SNPs identified by combined EA and PD approach. Each plot includes loess smoothing function and 95% confidence intervals (grey area).

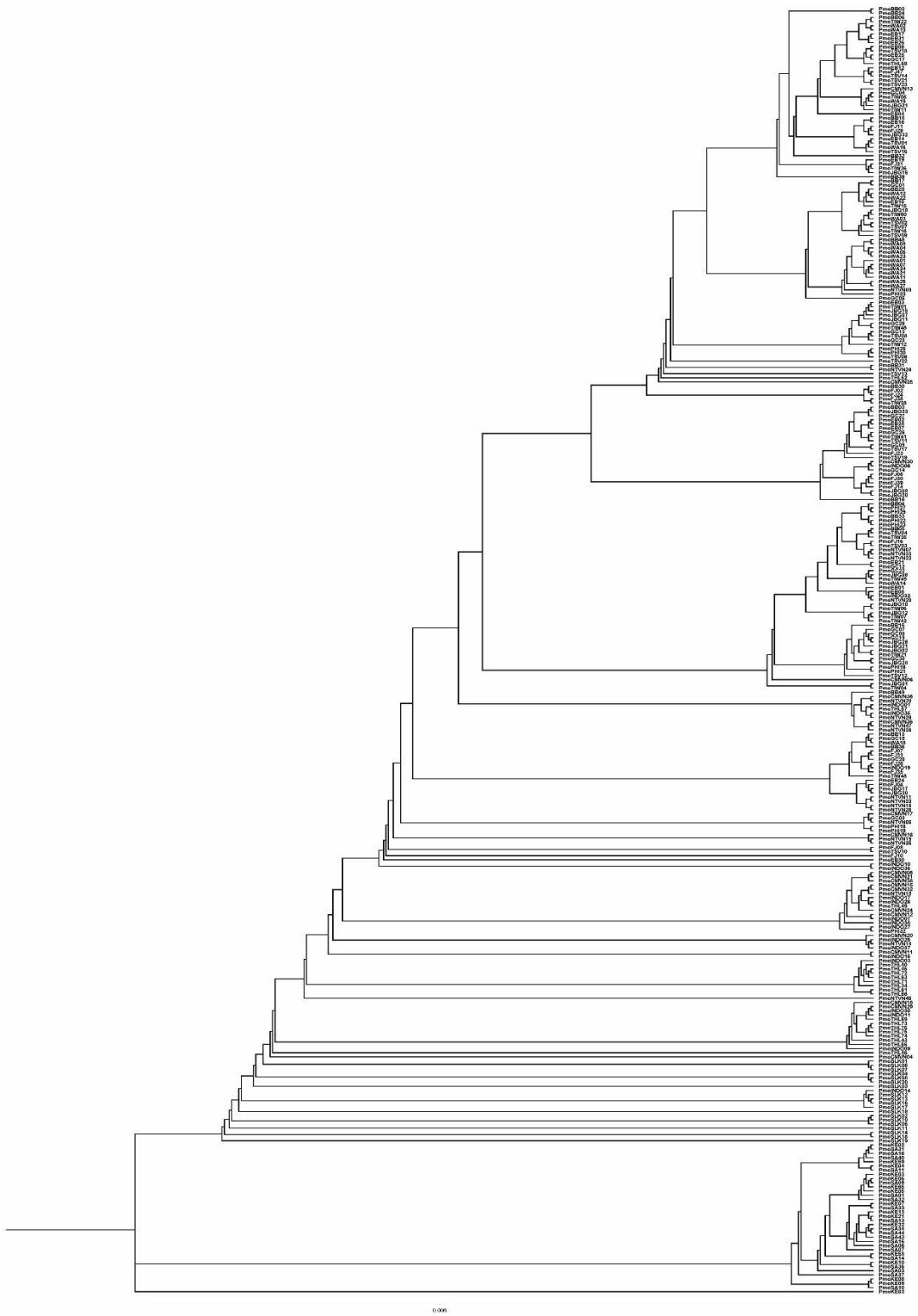
Appendix 3.10 Genetic diversity indices among 16 populations of *Penaeus monodon* assessed using 9,930 neutral loci

Population	*n*	A_R (± SD)	A_{PR} (± SD)	PPL (Av. MAF)	H_o (± SD)	H_E (± SD)	Av. MLH (± SD)	sMLH (±SD)	IR (± SD	F_{IS} ($p < 0.05$)
Bramston Beach (BB)	51	1.50 ± 0.43	0.002 ± 0.02	0.65	0.12 ± 0.15	0.13 ± 0.16	0.12 ± 0.01	1.02 ± 0.04	-0.03 ± 0.03	0.07
Etty Bay (EB)	31	1.50 ± 0.45	0.002 ± 0.03	0.66	0.12 ± 0.16	0.13 ± 0.16	0.12 ± 0.003	1.03 ± 0.03	-0.03 ± 0.02	0.06
Townsville (TSV)	22	1.50 ± 0.45	0.006 ± 0.05	0.57	0.12 ± 0.16	0.13 ± 0.16	0.12 ± 0.00	1.02 ± 0.03	-0.02 ± 0.02	0.07
Gulf of Carpentaria (GC)	30	1.54 ± 0.44	0.003 ± 0.03	0.65	0.13 ± 0.16	0.14 ± 0.16	0.13 ± 0.01	1.10 ± 0.07	-0.05 ± 0.04	0.07
Joseph Bonaparte Gulf (JBG)	34	1.54 ± 0.43	0.003 ± 0.03	0.59	0.13 ± 0.16	0.14 ± 0.16	0.13 ± 0.00	1.11 ± 0.03	-0.05 ± 0.03	0.07
Tiwi Island (TIW)	50	1.55 ± 0.42	0.003 ± 0.03	0.70	0.13 ± 0.15	0.14 ± 0.16	0.13 ± 0.00	1.10 ± 0.03	-0.05 ± 0.02	0.08
Nickol Bay (NKB)	22	1.46 ± 0.47	0.007 ± 0.06	0.51	0.14 ± 0.20	0.13 ± 0.17	0.14 ± 0.03	1.18 ± 0.27	-0.06 ± 0.05	-0.06
NhaTrang Vietnam (NTVN)	30	1.56 ± 0.44	0.008 ± 0.07	0.66	0.13 ± 0.16	0.14 ± 0.16	0.13 ± 0.004	1.10 ± 0.03	-0.08 ± 0.02	0.08
CaMau Vietnam (CMVN)	39	1.59 ± 0.41	0.004 ± 0.03	0.74	0.13 ± 0.15	0.15 ± 0.16	0.13 ± 0.004	1.09 ± 0.03	-0.07 ± 0.02	0.10
Fiji	49	1.41 ± 0.45	0.009 ± 0.08	0.49	0.11 ± 0.16	0.12 ± 0.17	0.11 ± 0.005	0.95 ± 0.04	0.00 ± 0.02	0.06
Java Indonesia (INDO)	38	1.60 ± 0.40	0.004 ± 0.04	0.74	0.13 ± 0.15	0.15 ± 0.16	0.13 ± 0.01	1.11 ± 0.10	-0.07 ± 0.04	0.10
Thailand (THL)	46	1.51 ± 0.46	0.011 ± 0.08	0.58	0.13 ± 0.17	0.14 ± 0.16	0.14 ± 0.004	1.13 ± 0.03	-0.10 ± 0.02	0.05
Philippines (PHI)	12	1.47 ± 0.50	0.005 ± 0.06	0.47	0.13 ± 0.19	0.13 ± 0.17	0.13 ± 0.00	1.12 ± 0.03	-0.07 ± 0.02	0.003
SriLanka (SLK)	19	1.46 ± 0.47	0.015 ± 0.09	0.51	0.10 ± 0.15	0.11 ± 0.16	0.10 ± 0.01	0.87 ± 0.11	0.03 ± 0.06	0.10
Kenya (KE)	16	1.11 ± 0.30	0.010 ± 0.08	0.18	0.02 ± 0.08	0.10 ± 0.27	0.02 ± 0.002	0.21 ± 0.02	0.64 ± 0.03	0.79
SouthAfrica (SA)	31	1.10 ± 0.29	0.005 ± 0.06	0.18	0.02 ± 0.08	0.09 ± 0.26	0.02 ± 0.002	0.21 ± 0.01	0.64 ± 0.02	0.77

Number of samples (*n*), mean allelic richness (A_R), private allelic richness (A_{RA}), percentage of polymorphic loci (PPL, mean MAF of polymorphic loci), mean observed heterozygosity (H_O), mean expected heterozygosity (H_E), average multi-locus heterozygosity (Av. MLH), standardised multi-locus heterozygosity (sMLH), internal relatedness (IR), and significant ($p < 0.05$) inbreeding coefficient (F_{IS}). SD: standard deviation.

Appendix 3.11 Genetic diversity indices among 16 populations of *Penaeus monodon* assessed using 2,155 SNP loci (> 99% concurrence in technical replicates with missing data per SNP < 2%).

Population	A_R (± SD)	H_o (± SD)	H_E (± SD)	F_{IS}
Fiji	1.34 ± 0.43	0.09 ± 0.16	0.10 ± 0.16	0.01
Bramston Beach (BB)	1.43 ± 0.42	0.10 ± 0.15	0.10 ± 0.14	0.01
Etty Bay (EB)	1.43 ± 0.43	0.10 ± 0.15	0.10 ± 0.15	-0.01
Townsville (TSV)	1.43 ± 0.44	0.10 ± 0.15	0.11 ± 0.15	0.01
Gulf of Carpentaria (GC)	1.47 ± 0.43	0.11 ± 0.15	0.11 ± 0.15	0.03
Joseph Bonaparte Gulf (JBG)	1.47 ± 0.43	0.11 ± 0.15	0.10 ± 0.15	0.01
Tiwi Island (TIW)	1.48 ± 0.41	0.11 ± 0.15	0.12 ± 0.15	0.03
Nickol Bay (NKB)	1.39 ± 0.45	0.11 ± 0.18	0.11 ± 0.16	-0.08
Philippines (PHI)	1.44 ± 0.50	0.12 ± 0.18	0.11 ± 0.15	-0.04
NhaTrang Vietnam (NTVN)	1.50 ± 0.44	0.13 ± 0.17	0.13 ± 0.16	0.01
CaMau Vietnam (CMVN)	1.55 ± 0.40	0.13 ± 0.16	0.13 ± 0.15	0.02
Java Indonesia (INDO)	1.55 ± 0.41	0.13 ± 0.16	0.13 ± 0.16	0.03
Thailand (THL)	1.47 ± 0.46	0.14 ± 0.18	0.13 ± 0.16	-0.03
SriLanka (SLK)	1.44 ± 0.46	0.10 ± 0.16	0.10 ± 0.15	0.02
Kenya (KE)	1.12 ± 0.32	0.03 ± 0.09	0.03 ± 0.09	0.03
SouthAfrica (SA)	1.11 ± 0.30	0.03 ± 0.09	0.03 ± 0.09	0.02



Appendix 4.1 Phylogeographic analyses of 317 *Penaeus monodon* individuals using SNP loci (*n* = 4,496) datasets. Tree topology corresponds to the best ML tree. ML tree generated in IQ-TREE v2.1.0 using 4,496 SNP loci under K3Pu+F+R6 model. Legend for locality codes is the same as defined in Figure 4.1.

Chapter 5.

Appendix 5.1 Custom R scripts to simulate progeny genotypes in R

```
## Simulation of progeny genotypes in R
## Set working directory

setwd(choose.dir())

# read the genotype file

SNP_2rows_genotypes <- read.csv("Please read DArT input file (SNP_2row.csv)",
header=FALSE, stringsAsFactor=FALSE, colClasses="character")

# remove population line

SNP_genotypes <- SNP_2rows_genotypes[-1,]

 # file containing all the sample ids from the possible parents and the offspring (1st column)
and their sex 1 for male or 2 for female (2nd column). The sex of offspring does not matter

samples_ID <- read.table("Sample_ID.txt",header=FALSE,stringsAsFactor=FALSE)

#  The list of SNPs in the panel. File containing all SNP id using for the simulation

  SNPsID <- read.table("MySNP_ID.txt",header=TRUE,stringsAsFactor=FALSE)

 # we used loop to repeat the progeny genotypes simulation

for(nrep in 1:5)

{

    ######## selecting the SNPs from the panel

    q <- match(SNP_genotypes$V1,SNPsID$V1)

  q[1] <- 1

    newdataset <- SNP_genotypes [-which(is.na(q)),]

      ## removing all the information columns in the DArT input file but keep the first 2 clumns
(AlleleID and CloneID)
  ## newdataset <- newdataset[,-length(newdataset)]
```

```
  newdataset <- newdataset[,-seq(2,18)]
  # transposing the matrix so that samples are in row and SNPs in column
  newdataset2 <- t(newdataset)
  # Count the number of SNPs
  nsnps <- (ncol(newdataset2) - 1)/2
# rename the SNPs LocXa and LocXb for CERVUS
  n <- c(0)
  for(i in 1:nsnps)
{
  if(i == 1) n <- c("Loc1a","Loc1b")
  else {
        n <- c(n,paste("Loc",i,"a",sep=""),paste("Loc",i,"b",sep=""))
  }
}
  n <- c("SampleID",n)
  # add the new SNPs names to the dataset
  colnames(newdataset2) <- n
  # remove the first 2 lines (AlleleID and CloneID) and keep new ID as Loci 1a, 1b
  newdataset2 <- newdataset2[-1,2]
  ###### Transforming DART format into CERVUS format
  newdataset3 <- matrix(NA,ncol=ncol(newdataset2),nrow=nrow(newdataset2))
  for(i in 1:nrow(newdataset2))
{ couSLKer <- 2
```

```
for(j in 1:floor(nsnps))
{
  if(newdataset2[i,couSLKer] == "1" & newdataset2[i,couSLKer+1] == "0")
  {
    newdataset3[i,couSLKer] <- "1"
    newdataset3[i,couSLKer+1] <- "1"
  }
  if(newdataset2[i,couSLKer] == "1" & newdataset2[i,couSLKer+1] == "1")
  {
    newdataset3[i,couSLKer] <- "1"
    newdataset3[i,couSLKer+1] <- "2"
  }
  if(newdataset2[i,couSLKer] =="0" & newdataset2[i,couSLKer+1] == "1")
  {
    newdataset3[i,couSLKer] <- "2"
    newdataset3[i,couSLKer+1] <-  "2"
  }
  if(newdataset2[i,couSLKer] =="-" & newdataset2[i,couSLKer+1] == "-")
  {
    newdataset3[i,couSLKer] <- "0"
    newdataset3[i,couSLKer+1] <- "0"
  }
  if(newdataset2[i,couSLKer] =="-" & newdataset2[i,couSLKer+1] != "-")
  {
```

```
    newdataset3[i,couSLKer] <- "0"
    newdataset3[i,couSLKer+1] <- "0"
  }
  if(newdataset2[i,couSLKer] !="-" & newdataset2[i,couSLKer+1] == "-")
  {
    newdataset3[i,couSLKer] <- "0"
    newdataset3[i,couSLKer+1] <- "0"
  }
  couSLKer <- couSLKer + 2
}
  }
  # Add column and row names to the converted dataset
  colnames(newdataset3) <- n
newdataset3[,1] <- newdataset2[,1]
  ## chek
  newdataset3[1:5,1:5]
  # Simulation of 18 full-sib families of 100 offspring each
# The ids of the parents should be in the sample id file
    allindiv <- read.table("Sample_ID.txt",stringsAsFactor=FALSE)
  dads <- allindiv[which(allindiv$V2 == 1),1]     ## dad code = 1
dads <- merge(dads,newdataset3[,1],1,1)
  mums <- allindiv[which(allindiv$V2 == 2),1]      ## mum code = 2
mums <- merge(mums,newdataset3[,1],1,1)
  dads$x <- as.character(dads$x)
```

```
mums$x <- as.character(mums$x)

  ################### 18 Families

  selecteddads <- sample(dads$x,18,replace=TRUE)

selectedmums<- sample(mums$x,18,replace=TRUE)

  ################### Creat Dad_Mums

p1 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[1]),])    # Sire family 1

p2 <- as.vector(newdataset3[which(newdataset3[,1] == selectedmums[1]),])  # Dam family 1

p3 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[2]),])   # Sire family 2

p4 <- as.vector(newdataset3[which(newdataset3[,1] == selectedmums[2]),])  # Dam family 2

p5 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[3]),])    # Sire family 3

p6 <- as.vector(newdataset3[which(newdataset3[,1] == selectedmums[3]),])  # Dam family 3

p7 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[4]),])    # Sire family 4

p8 <- as.vector(newdataset3[which(newdataset3[,1] == selectedmums[4]),])  # Dam family 4

 p9 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[5]),])   # Sire family 5

p10 <-as.vector(newdataset3[which(newdataset3[,1] == selectedmums[5]),])  # Dam family 5

p11 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[6]),])    # Sire family 6

p12 <-as.vector(newdataset3[which(newdataset3[,1] == selectedmums[6]),])  # Dam family 6

p13 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[7]),])   # Sire family 7

p14 <-as.vector(newdataset3[which(newdataset3[,1] == selectedmums[7]),])  # Dam family 7

p15 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[8]),])   # Sire family 8

p16 <-as.vector(newdataset3[which(newdataset3[,1] == selectedmums[8]),])  # Dam family 8

p17 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[9]),])    # Sire family 9

p18 <-as.vector(newdataset3[which(newdataset3[,1] == selectedmums[9]),])  # Dam family 9

p19 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[10]),])  # Sire family 10
```

```
p20 <-as.vector(newdataset3[which(newdataset3[,1] == selectedmums[10]),]) # Dam family 10
p21 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[11]),])  # Sire family 11
p22 <-as.vector(newdataset3[which(newdataset3[,1] == selectedmums[11]),]) # Dam family 11
p23 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[12]),])  # Sire family 12
p24 <-as.vector(newdataset3[which(newdataset3[,1] == selectedmums[12]),]) # Dam family 12
p25 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[13]),])    # Sire family 13
p26 <-as.vector(newdataset3[which(newdataset3[,1] == selectedmums[13]),])    # Dam family 13
p27 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[14]),])    # Sire family 14
p28 <-as.vector(newdataset3[which(newdataset3[,1] == selectedmums[14]),])    # Dam family 14
p29 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[15]),])    # Sire family 15
 p30 <-as.vector(newdataset3[which(newdataset3[,1] == selectedmums[15]),])    # Dam family 15
p31 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[16]),])    # Sire family 16
p32 <-as.vector(newdataset3[which(newdataset3[,1] == selectedmums[16]),])    # Dam family 16
p33 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[17]),])    # Sire family 17
p34 <-as.vector(newdataset3[which(newdataset3[,1] == selectedmums[17]),])    # Dam family 17
p35 <- as.vector(newdataset3[which(newdataset3[,1] == selecteddads[18]),])    # Sire family 18
p36 <-as.vector(newdataset3[which(newdataset3[,1] == selectedmums[18]),])    # Dam family 18
  sire <- list(p1,p3,p5,p7,p9,p11,p13,p15,p17,p19,p21,p23,p25,p27,p29,p31,p33,p35)
  dam <- list(p2,p4,p6,p8,p10,p12,p14,p16,p18,p20,p22,p24,p26,p28,p30,p32,p34,p36)
    offspring <- matrix(c(0),nrow=1800,ncol=ncol(newdataset3))
  siresim <- c(0)
  damsim <- c(0)
  idoffspring <- c(0)
  ######################## creating offspring
```

```
  g <- 1
for(i in 1:18)
{
     for(u in 1:100)
  {
         off <- rep(0,nsnps*2)
         couSLKer <- 2
    for(j in 1:nsnps)
    {
           s <- sire[[i]]
     d <- dam[[i]]
     q <- sample(c(couSLKer,couSLKer+1),1,replace=FALSE)
          off[couSLKer] <- s[q]
         q <- sample(c(couSLKer,couSLKer+1),1,replace=FALSE)
     off[couSLKer+1] <- d[q]
          if(is.na(off[couSLKer]) | is.na(off[couSLKer + 1]) )
     {
             off[couSLKer] <- "0"
       off[couSLKer+1] <- "0"
     }
     couSLKer <- couSLKer+2
         }
        # add id
    off[1] <- paste("OFFSIMU",g,sep="")
```

Appendix 5.2 Genomic distribution of five SNP panels exhibited the chromosome-scale genome assembly of *P. monodon* (Uengwetwanit et al., 2021)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_0015	NC_051397.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0021	NC_051406.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0026	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0027	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0028	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0029	NC_051424.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0030	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0031	NC_051404.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0032	NC_051414.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0033	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0034	NC_051394.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0037	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0038	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0039	NC_051414.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0043	NC_051411.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0044	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0045	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0047	NC_051400.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0050	NC_051425.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0051	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0052	NC_051419.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 34, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0053	NC_051397.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0054	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0055	NC_051416.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0057	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0058	NC_051425.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0060	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0061	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0062	NC_051425.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0064	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0067	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0068	NC_051426.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0070	NC_051418.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0071	NC_051396.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0073	NC_051423.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0074	NC_051395.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0075	NC_051400.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0076	NC_051410.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_0077	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0079	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0080	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0081	NC_051394.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0082	NC_051395.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0083	NC_051391.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0084	NC_051388.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0085	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0086	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0094	NC_051401.1	98	91	95	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0095	NC_051391.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0096	NC_051412.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0099	NC_051418.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0102	NC_051421.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0109	NC_051404.1	99	94	96	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0116	NC_051390.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0117	NC_051392.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0119	NC_051393.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0131	NC_051393.1	98	86	83	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0146	NC_051398.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0148	NC_051407.1	97	100	99	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0152	NC_051424.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0154	NC_051417.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0170	NC_051408.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0183	NC_051390.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0190	NC_051408.1	98	88	91	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0192	NC_051417.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0204	NC_051421.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0205	NC_051412.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0211	NC_051423.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0231	NC_051389.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0232	NC_051410.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0234	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0235	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0236	NC_051410.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0237	NC_051403.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0238	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0239	NC_051426.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_0240	NC_051424.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0241	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0242	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0243	NC_051428.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0244	NC_051412.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0246	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0247	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0248	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0251	NC_051392.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0254	NC_051424.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0255	NC_051420.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 35, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0256	NC_051415.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0257	NC_051406.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0258	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0260	NC_051426.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0261	NC_051404.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0262	NC_051406.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0263	NC_051425.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0264	NC_051424.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0265	NC_051397.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0266	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0267	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0269	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0270	NC_051409.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0271	NC_051388.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0273	NC_051398.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0278	NC_051407.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0284	NC_051411.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0285	NC_051410.1	98	86	83	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0294	NC_051410.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0297	NC_051408.1	98	86	83	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0298	NC_051408.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0300	NC_051426.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0305	NC_051402.1	97	93	92	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0326	NC_051402.1	98	91	95	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0332	NC_051404.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0333	NC_051407.1	98	87	90	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0341	NC_051400.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_0342	NC_051414.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0343	NC_051421.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0344	NC_051407.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0345	NC_051395.1	97	100	99	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0355	NC_051398.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0358	NC_051412.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0365	NC_051394.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0379	NC_051403.1	98	88	93	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0380	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0392	NC_051397.1	94	100	88	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0393	NC_051396.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0394	NC_051413.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0395	NC_051414.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0399	NC_051414.1	98	87	84	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0402	NC_051389.1	98	88	93	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0403	NC_051411.1	99	94	96	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0404	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0405	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0407	NC_051421.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0409	NC_051425.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0410	NC_051410.1	98	90	94	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0411	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0412	NC_051394.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0422	NC_051404.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0424	NC_051391.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0425	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0426	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0427	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0428	NC_051402.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0429	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0430	NC_051404.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0431	NC_051397.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0432	NC_051398.1	99	99	99	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0433	NC_051410.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0434	NC_051404.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0436	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0438	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0439	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_0440	NC_051419.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 34, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0441	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0442	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0443	NC_051400.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0444	NC_051394.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0445	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0446	NC_051403.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0447	NC_051423.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0448	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0449	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0450	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0452	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0453	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0455	NC_051387.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0456	NC_051410.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0457	NC_051399.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0458	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0459	NC_051421.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0461	NC_051391.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0463	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0465	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0466	NC_051399.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0467	NC_051395.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0468	NC_051422.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0469	NC_051407.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0470	NC_051405.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0471	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0472	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0473	NC_051390.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0475	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0478	NC_051400.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0480	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0481	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0482	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0483	NC_051399.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0484	NC_051403.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0485	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0487	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_0488	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0489	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0490	NC_051410.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0491	NC_051386.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0492	NC_051396.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0493	NC_051422.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0494	NC_051404.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0496	NC_051397.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0497	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0498	NC_051426.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0499	NC_051407.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0500	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0501	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0502	NC_051410.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0504	NC_051412.1	99	100	99	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0505	NC_051414.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0506	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0507	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0508	NC_051394.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0510	NC_051411.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0511	NC_051405.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0512	NC_051419.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 34, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0513	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0515	NC_051403.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0516	NC_051409.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0517	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0518	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0519	NC_051394.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0520	NC_051416.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0521	NC_051407.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0522	NC_051410.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0523	NC_051416.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0524	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0525	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0526	NC_051412.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0527	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0529	NC_051404.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0530	NC_051402.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_0531	NC_051415.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0532	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0533	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0534	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0535	NC_051412.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0536	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0537	NC_051386.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0538	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0539	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0540	NC_051393.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0541	NC_051422.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0542	NC_051403.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0543	NC_051406.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0544	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0545	NC_051424.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0546	NC_051416.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0547	NC_051403.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0549	NC_051391.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0550	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0552	NC_051397.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0553	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0554	NC_051403.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0555	NC_051402.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0556	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0557	NC_051422.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0558	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0559	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0560	NC_051403.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0561	NC_051399.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0562	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0563	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0564	NC_051411.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0565	NC_051412.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0566	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0567	NC_051394.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0569	NC_051419.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 34, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0570	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0571	NC_051427.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 42, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_0572	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0573	NC_051393.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0574	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0575	NC_051400.1	99	96	97	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0576	NC_051417.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0577	NC_051418.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0578	NC_051404.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0580	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0581	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0582	NC_051400.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0583	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0584	NC_051420.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 35, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0585	NC_051410.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0586	NC_051404.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0587	NC_051397.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0588	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0589	NC_051397.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0590	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0591	NC_051399.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0594	NC_051424.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0595	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0596	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0597	NC_051392.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0598	NC_051426.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0599	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0601	NC_051419.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 34, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0602	NC_051425.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0603	NC_051425.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0604	NC_051423.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0605	NC_051423.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0606	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0607	NC_051419.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 34, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0608	NC_051399.1	99	97	98	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0609	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0610	NC_051416.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0611	NC_051414.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0612	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0614	NC_051410.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_0616	NC_051410.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0618	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0619	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0620	NC_051415.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0621	NC_051392.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0622	NC_051421.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0623	NC_051424.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0624	NC_051406.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0625	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0626	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0627	NC_051400.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0629	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0630	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0631	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0632	NC_051392.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0633	NC_051422.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0646	NC_051415.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0651	NC_051390.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0655	NC_051398.1	96	99	95	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0668	NC_051397.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0673	NC_051404.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0686	NC_051415.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0697	NC_051387.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0698	NC_051397.1	98	87	90	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0700	NC_051425.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0705	NC_051418.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0724	NC_051406.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0728	NC_051405.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0733	NC_051417.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0758	NC_051423.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0770	NC_051390.1	98	91	95	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0778	NC_051416.1	98	88	93	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0783	NC_051415.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0798	NC_051387.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0801	NC_051390.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0807	NC_051395.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0826	NC_051397.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0827	NC_051393.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_0834	NC_051386.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0854	NC_051425.1	98	87	90	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0864	NC_051402.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0865	NC_051392.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0867	NC_051413.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0883	NC_051425.1	98	87	90	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0893	NC_051389.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0899	NC_051397.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0907	NC_051413.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0924	NC_051393.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0927	NC_051421.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0929	NC_051399.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0930	NC_051406.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0934	NC_051415.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0943	NC_051421.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0946	NC_051406.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0950	NC_051400.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0951	NC_051411.1	98	91	95	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0953	NC_051392.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0958	NC_051411.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0959	NC_051388.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0966	NC_051405.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0973	NC_051389.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0976	NC_051388.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0996	NC_051395.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0997	NC_051398.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1004	NC_051396.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1005	NC_051400.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1008	NC_051400.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1009	NC_051386.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1016	NC_051405.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1030	NC_051427.1	98	86	83	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 42, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1038	NC_051386.1	98	87	84	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1039	NC_051423.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1041	NC_051414.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1042	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1043	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1044	NC_051416.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_1045	NC_051386.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1046	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1048	NC_051416.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1049	NC_051421.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1050	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1051	NC_051410.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1052	NC_051422.1	97	100	99	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1053	NC_051424.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1054	NC_051408.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1055	NC_051399.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1056	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1057	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1058	NC_051420.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 35, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1060	NC_051395.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1061	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1063	NC_051423.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1064	NC_051390.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1065	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1066	NC_051395.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1067	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1068	NC_051386.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1069	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1070	NC_051391.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1072	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1073	NC_051400.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1074	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1075	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1076	NC_051416.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1077	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1078	NC_051398.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1079	NC_051390.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1080	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1082	NC_051410.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1083	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1084	NC_051428.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1086	NC_051411.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1088	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1090	NC_051423.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_1091	NC_051408.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1092	NC_051399.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1093	NC_051404.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1094	NC_051410.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1095	NC_051407.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1097	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1098	NC_051419.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 34, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1099	NC_051410.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1103	NC_051420.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 35, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1104	NC_051394.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1105	NC_051402.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1106	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1107	NC_051422.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1108	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1109	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1111	NC_051410.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1112	NC_051390.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1113	NC_051391.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1114	NC_051404.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1115	NC_051399.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1116	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1118	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1119	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1120	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1121	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1122	NC_051404.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1123	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1125	NC_051428.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1126	NC_051418.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1128	NC_051423.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1129	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1130	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1132	NC_051422.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1133	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1134	NC_051390.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1135	NC_051418.1	97	97	97	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1136	NC_051411.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1141	NC_051411.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_1146	NC_051393.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1147	NC_051392.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1156	NC_051398.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1167	NC_051395.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1170	NC_051403.1	98	90	93	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1195	NC_051395.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1200	NC_051426.1	96	96	93	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1208	NC_051398.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1210	NC_051422.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1214	NC_051417.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1216	NC_051425.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1222	NC_051408.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1224	NC_051415.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1232	NC_051401.1	97	91	85	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1255	NC_051395.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1266	NC_051419.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 34, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1291	NC_051387.1	98	87	84	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1300	NC_051411.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1302	NC_051428.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1309	NC_051417.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1310	NC_051400.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1312	NC_051410.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1316	NC_051404.1	98	88	91	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1326	NC_051397.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1327	NC_051424.1	97	100	99	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1328	NC_051397.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1329	NC_051410.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1330	NC_051409.1	99	97	98	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1333	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1334	NC_051409.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1335	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1336	NC_051399.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1337	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1338	NC_051390.1	97	100	99	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1339	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1340	NC_051402.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1341	NC_051422.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1342	NC_051396.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_1343	NC_051410.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1345	NC_051409.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1346	NC_051389.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1347	NC_051418.1	97	94	95	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1348	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1349	NC_051400.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1356	NC_051402.1	96	97	96	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1362	NC_051410.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1366	NC_051426.1	98	87	84	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1369	NC_051397.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1416	NC_051427.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 42, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1417	NC_051424.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1420	NC_051391.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1421	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1422	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1432	NC_051421.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1440	NC_051411.1	98	87	84	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1446	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1447	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1448	NC_051421.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1451	NC_051403.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1452	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1453	NC_051406.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1454	NC_051424.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1456	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1457	NC_051409.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1458	NC_051414.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1459	NC_051400.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1460	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1461	NC_051402.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1462	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1463	NC_051416.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1464	NC_051410.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1465	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1466	NC_051417.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1467	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1468	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1469	NC_051394.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_1470	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1471	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1474	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1475	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1476	NC_051403.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1477	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1478	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1479	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1483	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1484	NC_051386.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1485	NC_051400.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1486	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1487	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1488	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1490	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1491	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1492	NC_051421.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1493	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1494	NC_051424.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1495	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1498	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1499	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1500	NC_051422.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1501	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1503	NC_051402.1	99	94	96	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1525	NC_051408.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1527	NC_051424.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1538	NC_051400.1	99	96	97	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1541	NC_051393.1	98	88	91	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1550	NC_051403.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1554	NC_051392.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1565	NC_051412.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1570	NC_051389.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1576	NC_051389.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1579	NC_051397.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1590	NC_051399.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1605	NC_051427.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 42, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1607	NC_051410.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_1615	NC_051392.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1616	NC_051428.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1623	NC_051402.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1629	NC_051421.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1633	NC_051398.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1634	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1635	NC_051402.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1636	NC_051405.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1637	NC_051407.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1638	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1639	NC_051423.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1640	NC_051404.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1641	NC_051407.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1642	NC_051402.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1644	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1645	NC_051412.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1646	NC_051404.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1648	NC_051425.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1649	NC_051394.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1650	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1651	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1652	NC_051387.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1653	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1654	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1655	NC_051425.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1656	NC_051406.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1657	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1658	NC_051411.1	97	99	98	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1659	NC_051422.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1660	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1661	NC_051424.1	98	90	94	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1662	NC_051402.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1663	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1664	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1665	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1666	NC_051421.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1667	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1668	NC_051428.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_1669	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1671	NC_051409.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1672	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1673	NC_051421.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1674	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1675	NC_051401.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1676	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1678	NC_051394.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1679	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1680	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1681	NC_051415.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1683	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1686	NC_051386.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1687	NC_051426.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1688	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1690	NC_051401.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1691	NC_051400.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1692	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1693	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1694	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1695	NC_051416.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1696	NC_051397.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1697	NC_051399.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1698	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1699	NC_051399.1	97	97	97	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1700	NC_051422.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1701	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1702	NC_051423.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1703	NC_051405.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1704	NC_051392.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1705	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1706	NC_051393.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1714	NC_051424.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1717	NC_051408.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1721	NC_051409.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1728	NC_051405.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1737	NC_051389.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1743	NC_051399.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_1759	NC_051406.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1772	NC_051398.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1780	NC_051409.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1788	NC_051422.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1801	NC_051412.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1812	NC_051412.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1813	NC_051416.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1829	NC_051395.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1845	NC_051396.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1859	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1860	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1861	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1862	NC_051406.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1863	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1864	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1865	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1867	NC_051395.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1868	NC_051388.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1869	NC_051426.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1870	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1871	NC_051404.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1872	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1874	NC_051427.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 42, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1875	NC_051398.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1876	NC_051419.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 34, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1878	NC_051425.1	98	90	94	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1879	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1880	NC_051399.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1881	NC_051411.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1882	NC_051403.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1883	NC_051397.1	97	97	97	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1884	NC_051416.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1885	NC_051426.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1886	NC_051426.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1887	NC_051395.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1888	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1889	NC_051402.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1890	NC_051426.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_1891	NC_051410.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1892	NC_051416.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1893	NC_051391.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1894	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1895	NC_051400.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1896	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1897	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1898	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1899	NC_051411.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1900	NC_051416.1	97	96	96	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1902	NC_051421.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1903	NC_051401.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1904	NC_051419.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 34, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1906	NC_051400.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1907	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1908	NC_051398.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1909	NC_051409.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1910	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1911	NC_051390.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1912	NC_051406.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1913	NC_051396.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1914	NC_051427.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 42, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1915	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1916	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1917	NC_051420.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 35, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1918	NC_051402.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1919	NC_051414.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1920	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1921	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1924	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1925	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1927	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1928	NC_051393.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1931	NC_051408.1	97	90	85	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1935	NC_051410.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1944	NC_051410.1	98	87	90	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1949	NC_051392.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1958	NC_051421.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_1971	NC_051424.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1980	NC_051405.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1990	NC_051403.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1991	NC_051386.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1993	NC_051390.1	98	87	90	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1996	NC_051427.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 42, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_1998	NC_051391.1	97	90	85	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2008	NC_051404.1	97	97	97	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2010	NC_051387.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2012	NC_051410.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2016	NC_051401.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2049	NC_051421.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2053	NC_051391.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2059	NC_051403.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2060	NC_051420.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 35, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2061	NC_051389.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2063	NC_051416.1	98	90	94	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2064	NC_051403.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2065	NC_051390.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2066	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2068	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2069	NC_051427.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 42, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2071	NC_051394.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2073	NC_051387.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2074	NC_051405.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2077	NC_051414.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2078	NC_051393.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2079	NC_051419.1	97	100	99	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 34, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2081	NC_051414.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2082	NC_051394.1	97	100	99	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2084	NC_051415.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2085	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2086	NC_051409.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2088	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2089	NC_051401.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2090	NC_051397.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2091	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2092	NC_051415.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_2093	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2094	NC_051415.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2096	NC_051417.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2097	NC_051411.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2098	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2099	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2101	NC_051387.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2102	NC_051414.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2124	NC_051415.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2129	NC_051399.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2130	NC_051405.1	99	96	97	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2131	NC_051414.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2132	NC_051391.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2139	NC_051419.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 34, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2143	NC_051389.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2149	NC_051399.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2151	NC_051402.1	98	91	95	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2152	NC_051404.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2154	NC_051398.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2156	NC_051423.1	98	88	93	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2161	NC_051390.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2173	NC_051404.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2178	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2179	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2180	NC_051389.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2181	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2182	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2184	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2185	NC_051393.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2186	NC_051422.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2187	NC_051428.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2188	NC_051394.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2190	NC_051422.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2191	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2192	NC_051428.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2193	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2194	NC_051388.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2195	NC_051421.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_2196	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2197	NC_051398.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2198	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2199	NC_051392.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2200	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2201	NC_051425.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2202	NC_051407.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2210	NC_051393.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2220	NC_051410.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2221	NC_051397.1	97	94	96	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2224	NC_051390.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2235	NC_051392.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2237	NC_051394.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2238	NC_051386.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2241	NC_051387.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2247	NC_051396.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2248	NC_051393.1	98	87	90	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2249	NC_051391.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2251	NC_051400.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2254	NC_051411.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2258	NC_051407.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2259	NC_051395.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2260	NC_051427.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 42, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2261	NC_051428.1	98	93	96	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2262	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2268	NC_051404.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2269	NC_051404.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2274	NC_051400.1	98	86	83	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2282	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2284	NC_051415.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2285	NC_051426.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2286	NC_051415.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2293	NC_051413.1	98	90	93	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2294	NC_051386.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2295	NC_051428.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2297	NC_051414.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2298	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2299	NC_051398.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_2304	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2305	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2307	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2308	NC_051428.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2316	NC_051421.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2323	NC_051397.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2324	NC_051386.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2325	NC_051423.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2326	NC_051423.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2327	NC_051404.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2328	NC_051425.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2329	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2331	NC_051402.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2334	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2336	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2337	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2353	NC_051395.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2364	NC_051418.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2366	NC_051393.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2367	NC_051411.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2370	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2371	NC_051427.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 42, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2374	NC_051398.1	99	96	97	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2375	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2376	NC_051406.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2377	NC_051423.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2391	NC_051421.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2392	NC_051423.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2397	NC_051405.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2398	NC_051418.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2399	NC_051413.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2400	NC_051398.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2401	NC_051410.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2402	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2403	NC_051395.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2404	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2405	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2406	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_2407	NC_051409.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2408	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2409	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2410	NC_051406.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2411	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2412	NC_051400.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2418	NC_051416.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2425	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2426	NC_051400.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2427	NC_051397.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2428	NC_051390.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2429	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2430	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2431	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2434	NC_051415.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2436	NC_051417.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2438	NC_051400.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2440	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2441	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2442	NC_051396.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2443	NC_051421.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2444	NC_051418.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2445	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2446	NC_051397.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2447	NC_051401.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2448	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2449	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2450	NC_051424.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2454	NC_051395.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2457	NC_051402.1	98	90	93	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2460	NC_051421.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2463	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2464	NC_051410.1	99	99	99	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2474	NC_051389.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2475	NC_051426.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2476	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2477	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2478	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_2479	NC_051409.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2488	NC_051401.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2490	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2491	NC_051397.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2495	NC_051418.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2496	NC_051413.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2498	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2499	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2500	NC_051415.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2501	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2502	NC_051387.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2507	NC_051408.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2508	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2512	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2513	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2519	NC_051422.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2524	NC_051398.1	98	91	95	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2529	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2530	NC_051422.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2531	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2532	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2533	NC_051400.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2534	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2541	NC_051387.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2546	NC_051417.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2548	NC_051402.1	99	99	99	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2555	NC_051402.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2557	NC_051409.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2559	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2560	NC_051420.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 35, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2561	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2562	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2563	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2564	NC_051392.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2565	NC_051421.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2566	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2567	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2568	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_2583	NC_051393.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2595	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2596	NC_051425.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2598	NC_051421.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2604	NC_051428.1	98	90	94	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2607	NC_051397.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2608	NC_051422.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2616	NC_051388.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2620	NC_051389.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2625	NC_051395.1	93	100	88	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2627	NC_051427.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 42, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2629	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2630	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2631	NC_051422.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2632	NC_051392.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2634	NC_051421.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2635	NC_051404.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2636	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2637	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2638	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2639	NC_051400.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2640	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2641	NC_051403.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2642	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2648	NC_051418.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2655	NC_051428.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2657	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2658	NC_051418.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2659	NC_051428.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2660	NC_051399.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2661	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2662	NC_051405.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2663	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2664	NC_051419.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 34, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2667	NC_051390.1	94	99	88	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2669	NC_051421.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2670	NC_051409.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2673	NC_051423.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_2674	NC_051410.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2676	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2677	NC_051421.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2678	NC_051388.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2679	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2680	NC_051425.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2682	NC_051392.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2691	NC_051393.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2692	NC_051408.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2693	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2697	NC_051399.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2698	NC_051391.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2699	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2700	NC_051427.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 42, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2701	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2702	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2704	NC_051426.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2705	NC_051388.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2706	NC_051391.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2707	NC_051405.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2715	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2716	NC_051401.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2717	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2719	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2720	NC_051405.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2721	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2729	NC_051398.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2746	NC_051387.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2782	NC_051387.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2784	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2785	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2786	NC_051391.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2788	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2790	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2791	NC_051387.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2792	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2794	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_2795	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2796	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2802	NC_051417.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2806	NC_051386.1	98	88	91	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2815	NC_051409.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2827	NC_051407.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2829	NC_051413.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2865	NC_051418.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2884	NC_051428.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2892	NC_051418.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2897	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2898	NC_051422.1	100	94	97	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2899	NC_051396.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2900	NC_051400.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2901	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2902	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2903	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2904	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2905	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2906	NC_051409.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 24, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2907	NC_051398.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2908	NC_051401.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2909	NC_051417.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2910	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2912	NC_051418.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2913	NC_051392.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2914	NC_051406.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2915	NC_051414.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2916	NC_051397.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2917	NC_051391.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2918	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2919	NC_051423.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2920	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2921	NC_051425.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 40, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2940	NC_051410.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2951	NC_051406.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2973	NC_051426.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2979	NC_051411.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 26, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_2999	NC_051398.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3001	NC_051406.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3003	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3005	NC_051424.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3006	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3007	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3009	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3010	NC_051415.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3011	NC_051403.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3013	NC_051426.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3016	NC_051428.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3017	NC_051414.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3018	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3020	NC_051418.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3021	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3022	NC_051415.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3030	NC_051400.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3031	NC_051386.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3038	NC_051386.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3062	NC_051424.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 39, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3068	NC_051418.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3080	NC_051417.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3090	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3092	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3093	NC_051392.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3094	NC_051421.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 36, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3095	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3096	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3097	NC_051392.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3099	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3100	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3102	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3103	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3104	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3105	NC_051403.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3106	NC_051390.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3107	NC_051396.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3108	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_3118	NC_051386.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3123	NC_051387.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3136	NC_051396.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3139	NC_051403.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3146	NC_051393.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3154	NC_051394.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3155	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3156	NC_051410.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 25, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3157	NC_051387.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3158	NC_051418.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3159	NC_051403.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3160	NC_051420.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 35, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3162	NC_051391.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3163	NC_051413.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 28, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3175	NC_051401.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3180	NC_051388.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3186	NC_051390.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3191	NC_051387.1	98	87	84	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3199	NC_051417.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3200	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3201	NC_051395.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 10, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3203	NC_051387.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3204	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3206	NC_051418.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 33, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3207	NC_051389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 4, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3208	NC_051391.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3210	NC_051403.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 18, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3211	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3212	NC_051423.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3215	NC_051388.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3219	NC_051426.1	98	93	96	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 41, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3220	NC_051407.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3221	NC_051393.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3229	NC_051415.1	98	90	94	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 30, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3230	NC_051416.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3232	NC_051423.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3233	NC_051407.1	98	93	96	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3235	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_3236	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3237	NC_051394.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3239	NC_051398.1	99	99	99	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3245	NC_051422.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 37, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3250	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3251	NC_051390.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 5, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3252	NC_051419.1	100	97	99	2.69E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 34, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3253	NC_051400.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 15, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3254	NC_051423.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 38, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3258	NC_051416.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3259	NC_051397.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 12, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3264	NC_051392.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 7, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3265	NC_051406.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3267	NC_051406.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 21, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3273	NC_051396.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 11, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3278	NC_051405.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3284	NC_051391.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 6, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3303	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3347	NC_051399.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3349	NC_051407.1	100	93	96	1.25E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3366	NC_051405.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 chromosome 20, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3367	NC_051417.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 32, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3381	NC_051404.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 19, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3393	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3395	NC_051394.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 9, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3398	NC_051386.1	98	88	91	9.75E-22	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3401	NC_051427.1	99	96	97	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 42, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3403	NC_051386.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3404	NC_051408.1	100	91	96	4.51E-25	*Penaeus monodon* isolate SGIC_2016 chromosome 23, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3405	NC_051398.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 13, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3411	NC_051402.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 17, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3412	NC_051386.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 1, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3413	NC_051428.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 43, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3414	NC_051399.1	100	96	98	9.68E-27	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3424	NC_051407.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 22, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3427	NC_051401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 16, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3430	NC_051416.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 31, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3432	NC_051412.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 27, NSTDA_Pmon_1, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_3433	NC_051388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 3, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0022	NW_023639857.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_11128, whole genome shotgun sequence
2021PMGT_0048	NW_023639480.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_10789, whole genome shotgun sequence
2021PMGT_0063	NW_023659678.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_4812, whole genome shotgun sequence
2021PMGT_0230	NW_023664112.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_8807, whole genome shotgun sequence
2021PMGT_0360	NW_023660942.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_5951, whole genome shotgun sequence
2021PMGT_0451	NW_023660421.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_5482, whole genome shotgun sequence
2021PMGT_0579	NW_023653161.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_23115, whole genome shotgun sequence
2021PMGT_0592	NW_023644389.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_15211, whole genome shotgun sequence
2021PMGT_0593	NW_023640934.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_121, whole genome shotgun sequence
2021PMGT_1085	NW_023644668.1	99	97	98	1.25E-25	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_15463, whole genome shotgun sequence
2021PMGT_1344	NW_023649379.1	100	99	99	7.49E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_19707, whole genome shotgun sequence
2021PMGT_1355	NW_023638616.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_1001, whole genome shotgun sequence
2021PMGT_1473	NW_023663563.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_8312, whole genome shotgun sequence
2021PMGT_1496	NW_023658109.1	96	100	97	2.71E-22	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_339, whole genome shotgun sequence
2021PMGT_1497	NW_023643587.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_1449, whole genome shotgun sequence
2021PMGT_1643	NW_023652278.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_2232, whole genome shotgun sequence
2021PMGT_1677	NW_023661085.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_6080, whole genome shotgun sequence
2021PMGT_1763	NW_023658674.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_3903, whole genome shotgun sequence
2021PMGT_1873	NW_023665070.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_9670, whole genome shotgun sequence
2021PMGT_1905	NW_023641388.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_12508, whole genome shotgun sequence
2021PMGT_2080	NW_023643031.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_13989, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_1940	NW_023649640.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_19942, whole genome shotgun sequence
2021PMGT_1923	NW_023638928.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_10291, whole genome shotgun sequence
2021PMGT_2022	NW_023659013.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_4211, whole genome shotgun sequence
2021PMGT_2067	NW_023663324.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_8098, whole genome shotgun sequence
2021PMGT_2100	NW_023650812.1	98	93	96	5.83E-24	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_210, whole genome shotgun sequence
2021PMGT_2373	NW_023643957.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_14822, whole genome shotgun sequence
2021PMGT_2349	NW_023643287.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_14219, whole genome shotgun sequence
2021PMGT_2368	NW_023651059.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_21221, whole genome shotgun sequence
2021PMGT_2451	NW_023657885.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_3186, whole genome shotgun sequence
2021PMGT_2521	NW_023660507.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_556, whole genome shotgun sequence
2021PMGT_2510	NW_023643665.1	100	83	89	9.75E-22	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_1456, whole genome shotgun sequence
2021PMGT_2535	NW_023662662.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_7500, whole genome shotgun sequence
2021PMGT_2601	NW_023665035.1	100	87	93	2.10E-23	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_9639, whole genome shotgun sequence
2021PMGT_2683	NW_023643762.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_14647, whole genome shotgun sequence
2021PMGT_2684	NW_023648614.1	94	100	88	3.51E-21	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_19018, whole genome shotgun sequence
2021PMGT_3406	NW_023658668.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_3898, whole genome shotgun sequence
2021PMGT_2780	NW_023648482.1	100	81	82	3.51E-21	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_189, whole genome shotgun sequence
2021PMGT_2718	NW_023641401.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_1252, whole genome shotgun sequence
2021PMGT_2789	NW_023657798.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_3107, whole genome shotgun sequence
2021PMGT_3202	NW_023659815.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_4936, whole genome shotgun sequence
2021PMGT_3015	NW_023649226.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_1957, whole genome shotgun sequence

Appendix 5.2 (cont.)

SNP_ID	Contig name	% Pairwise Identity	Query coverage	Grade	E Value	Description
2021PMGT_3045	NW_023660034.1	100	86	93	7.54E-23	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_5133, whole genome shotgun sequence
2021PMGT_3225	NW_023659775.1	100	84	91	2.71E-22	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_490, whole genome shotgun sequence
2021PMGT_3145	NW_023652900.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_22880, whole genome shotgun sequence
2021PMGT_3205	NW_023639629.1	99	100	99	3.48E-26	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_10922, whole genome shotgun sequence
	NW_023655363.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_251, whole genome shotgun sequence
2021PMGT_3198	NW_023655617.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_25328, whole genome shotgun sequence
	NW_023654808.1	100	88	94	5.83E-24	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_246, whole genome shotgun sequence
2021PMGT_1450	NW_023658207.1	100	90	95	1.62E-24	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_3479, whole genome shotgun sequence
	NW_023658168.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_3443, whole genome shotgun sequence
2021PMGT_3008	NW_023660306.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_5379, whole genome shotgun sequence
	NW_023647809.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_18293, whole genome shotgun sequence
2021PMGT_3012	NW_023660312.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_5384, whole genome shotgun sequence
	NW_023659907.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_5019, whole genome shotgun sequence
2021PMGT_3098	NW_023661758.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_6687, whole genome shotgun sequence
	NC_051414.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 29, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_0020	NW_023660507.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_556, whole genome shotgun sequence
	NC_051387.1	99	99	99	3.48E-26	*Penaeus monodon* isolate SGIC_2016 chromosome 2, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_2072	NW_023649640.1	99	99	99	3.48E-26	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_19942, whole genome shotgun sequence
	NC_051399.1	97	90	85	3.51E-21	*Penaeus monodon* isolate SGIC_2016 chromosome 14, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3014	NW_023659095.1	98	90	94	7.54E-23	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_4286,
	NC_051393.1	100	100	100	2.08E-28	*Penaeus monodon* isolate SGIC_2016 chromosome 8, NSTDA_Pmon_1, whole genome shotgun sequence
2021PMGT_3276	NW_023640922.1	99	100	99	9.68E-27	*Penaeus monodon* isolate SGIC_2016 unplaced genomic scaffold, NSTDA_Pmon_1 PmonScaffold_12089, whole genome shotgun sequence

Appendix 6.1 Published journal article reprints

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

- **Vu, N. T. T**., Zenger, K. R., Guppy, J. L., Sellars, M. J., Silva, C. N. S., Kjeldsen, S. R., & Jerry, D. R. (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. doi:10.1186/s12864-020-07084-x.

- **Vu, N. T. T**., Zenger, K. R., Silva, C. N. S., Guppy, J. L., & Jerry, D. R. (2021). Population structure, genetic connectivity, and signatures of local adaptation of the giant black tiger shrimp (*Penaeus monodon*) throughout the Indo–Pacific Region. Genome biology and evolution. doi:10.1093/gbe/evab214.