





# A chromosome-level assembly of the black tiger shrimp (*Penaeus monodon*) genome facilitates the identification of growth-associated genes

Tanaporn Uengwetwanit PhD<sup>1</sup>  | Wirulda Pootakham PhD<sup>2</sup>  |  
Intawat Nookaew Ph.D.<sup>3</sup>  | Chutima Sonthirod<sup>2</sup> | Pacharaporn Angthong PhD<sup>1</sup> |  
Kanchana Sittikankaew<sup>1</sup> | Wanilada Rungrassamee PhD<sup>1</sup> | Sopacha Arayamethakorn<sup>1</sup> |  
Thidathip Wongsurawat PhD<sup>3,4</sup> | Piroon Jenjaroenpun PhD<sup>3,4</sup> | Duangjai Sangsrakru<sup>2</sup> |  
Rungnapa Leelatanawit PhD<sup>1</sup> | Jutatip Khudet<sup>5</sup> | Jasper J. Koehorst<sup>6</sup> | Peter J. Schaap<sup>6</sup> |  
Vitor Martins dos Santos<sup>6</sup> | Frédéric Tangy<sup>7</sup> | Nitsara Karoonuthaisiri PhD<sup>1</sup> 

<sup>1</sup>National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Pathum Thani, 12120, Thailand

<sup>2</sup>National Omics Center, National Science and Technology Development Agency, Pathum Thani, Thailand

<sup>3</sup>Department of Biomedical Informatics, College of Medicine, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA

<sup>4</sup>Division of Bioinformatics and Data Management for Research, Department of Research and Development, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand

<sup>5</sup>Shrimp Genetic Improvement Center, Integrative Aquaculture Biotechnology Research Group, Surat Thani, Thailand

<sup>6</sup>Laboratory of Systems and Synthetic Biology, Department of Agrotechnology and Food Sciences, Wageningen University and Research, Wageningen, The Netherlands

<sup>7</sup>Viral Genomics and Vaccination Unit, UMR3569 CNRS, Virology Department, Institut Pasteur, Paris, France

## Correspondence

Nitsara Karoonuthaisiri, National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Pathum Thani, 12120, Thailand.  
Email: nitsara@alumni.stanford.edu; nitsara.kar@biotec.or.th

## Funding information

National Center for Genetic Engineering and Biotechnology, Grant/Award Number: P1651718; European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant, Grant/Award Number: 734486; TDG Food and Feed program, National Science and Technology Development Agency, Grant/Award Number: P1950419; National Institute of General Medical Sciences of the National Institutes of Health, Grant/Award Number: P20GM125503

## Abstract

To salvage marine ecosystems from fishery overexploitation, sustainable and efficient aquaculture must be emphasized. The knowledge obtained from available genome sequence of marine organisms has accelerated marine aquaculture in many cases. The black tiger shrimp (*Penaeus monodon*) is one of the most prominent cultured penaeid shrimps (Crustacean) with an average annual global production of half a million tons in the last decade. However, its currently available genome assemblies lack the contiguity and completeness required for accurate genome annotation due to the highly repetitive nature of the genome and technical difficulty in extracting high-quality, high-molecular weight DNA. Here, we report the first chromosome-level whole-genome assembly of *P. monodon*. The combination of long-read Pacific Biosciences (PacBio) and long-range Chicago and Hi-C technologies enabled a successful assembly of this first high-quality genome sequence. The final assembly covered 2.39 Gb (92.3% of the estimated genome size) and contained 44 pseudomolecules, corresponding to

Tanaporn Uengwetwanit and Wirulda Pootakham contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. Molecular Ecology Resources published by John Wiley & Sons Ltd.

the haploid chromosome number. Repetitive elements occupied a substantial portion of the assembly (62.5%), the highest of the figures reported among crustacean species. The availability of this high-quality genome assembly enabled the identification of genes associated with rapid growth in the black tiger shrimp through the comparison of hepatopancreas transcriptome of slow-growing and fast-growing shrimps. The results highlighted several growth-associated genes. Our high-quality genome assembly provides an invaluable resource for genetic improvement and breeding penaeid shrimp in aquaculture. The availability of *P. monodon* genome enables analyses of ecological impact, environment adaptation and evolution, as well as the role of the genome to protect the ecological resources by promoting sustainable shrimp farming.

#### KEYWORDS

black tiger shrimp, growth-associated genes, Hi-C, PacBio, *Penaeus monodon*, reference genome, transcriptomics

## 1 | INTRODUCTION

Crustaceans have critical roles in many ecosystems and represent an important aquatic food source for humankind. They serve as a vital food web link between primary producers (algae and aquatic weeds) and higher trophic levels (Thorp & Rogers, 2011). Crustaceans are also considered bioindicators for environmental quality of ecosystems (Angel et al., 2010; Farkas et al., 2003). In addition, crustaceans are an important source of aquatic food protein to human (Bondad-Reantaso et al., 2012). Crustacean fisheries produced an average of over six million tons from 2015–2018, contributing over 7% of global fishery (FAO, 2020). However, overexploitation of crustacean fisheries has undeniably affected marine ecosystems in various aspects such as a reduction of fish abundance composition, genetics and other population parameters (growth, age, sex ratio, maturation, size structure, etc.), species (Agardy, 2000; Garcia et al., 2003; Gislason, 2003; Porobic et al., 2019).

Given the threat from fishery overexploitation to marine ecosystems, the need to feed a growing global population puts pressure on the further development of aquaculture. Consequently, aquaculture has become one of the world's fastest-growing food sectors with an annual growth rate of 5.8% since 2010 (FAO, 2018). Integration of genomic information provides a better understanding of marine biodiversity and supports the development of sustainable aquaculture, reducing our dependence on fishery and, in turn leading to ecological sustainability (Huete-Pérez & Quezada, 2013; Wenne et al., 2007). Genetic improvement of cultivated fish and crustaceans through marker-assisted selection and genome-wide selection is an ongoing process with a tremendous benefit for sustainable aquaculture development (Houston et al., 2020; Zenger et al., 2019). Nevertheless, only a few complete genomes of crustaceans have been reported to date.

Among edible crustaceans in aquaculture, the penaeid marine shrimp (family Penaeidae) are the predominately cultured group (Thornber et al., 2019), with an annual production exceeding 4.5

million tons (Anderson, 2019). In this group, the black tiger shrimp (*Penaeus monodon*) is one of the dominant cultured species, accounting for 9% of total crustacean production (FAO, 2018). While the penaeid shrimp industry has seen dramatic growth for the past few decades, aquaculture production of *P. monodon* proved to be unsustainable due to a lack of biological and genetic knowledge to achieve industrial desirable traits such as fast growth, disease resistance, reproductive maturation without reliance on wild brooders (Guppy et al., 2018). The lack of a high-quality reference genome sequence makes it more difficult to perform a large-scale SNP identification for marker-assisted breeding programs, to carry out genetic diversity studies and to elucidate the mechanisms underlying the regulation of biological processes of interest such as adaptive evolution, stress responses, growth and reproductive maturation.

While such a high-quality draft genome sequence has not been reported for *P. monodon*, over the past two decades, several attempts have been made to investigate the genetic architecture of this important species. BAC library construction (Wuthisuthimethavee et al., 2009), fosmid library end sequencing (Huang et al., 2011), molecular marker development (Brooker et al., 2000; Tassanakajon et al., 1997, 1998), linkage map construction (Baranski et al., 2014; Wilson et al., 2002), and transcriptome analysis (Huerlimann et al., 2018; Karoonuthaisiri et al., 2009; Leelatanawit et al., 2011; Lehnert et al., 1999; Pootakham et al., 2020; Sittikankaew et al., 2020; Supungul et al., 2002; Tassanakajon et al., 2006; Tong et al., 2002; Uengwetwanit et al., 2018) were successfully explored, but these attempts yielded rather limited genome information. Previous attempts to obtain genome sequence data of the black tiger shrimp relied primarily on short-read sequencing platforms. Because a large proportion of the black tiger shrimp genome is highly repetitive, the two publicly available (short-read based) draft genome assemblies are highly fragmented, with N50 contig lengths of merely 937 bp (Yuan et al., 2018) and 1,982 bp (Quyen et al., 2020). Even though these genome resources are useful in understanding the black tiger shrimp genetics, they lack the contiguity and completeness required

for accurate genome annotation, thorough comparative analyses and marker-assisted breeding programs.

Here, we combined a long-read sequencing technology and two long-range scaffolding techniques to obtain a high-quality, chromosome-scale genome assembly. First, the Pacific Biosciences (PacBio) sequencing platform was employed to generate the preliminary assembly. The PacBio sequencing technology enables contiguous assembly of repetitive regions containing transposable elements and tandem repeats, which are often omitted or highly fragmented in genomic sequences currently available in public databases. Subsequently, we applied the long-range Chicago (in vitro proximity ligation) and Hi-C (in vivo fixation of chromosomes) scaffolding techniques to further scaffold the preliminary assembly to achieve the first chromosome-scale genome assembly in *P. monodon*. The availability of this genome assembly will allow us to investigate the genetic diversity of black tiger shrimp population, gain a better understanding of their ability to adapt to captive farming systems and overcome the challenges from fishery overexploitation to preserve marine ecosystem.

To demonstrate the utilization of this dramatically improved genome assembly for breeding purposes, it was used to identify genes related to growth performance by comparing gene expression profiles in the hepatopancreas of the large black tiger shrimp and the small black tiger shrimp from the same culture condition.

## 2 | MATERIALS AND METHODS

### 2.1 | Sample collection and DNA extraction

Muscle tissue of a 5-month-old female *P. monodon* was collected from the Shrimp Genetic Improvement Center (SGIC, Surat Thani, Thailand), immediately frozen in liquid nitrogen, and stored at  $-80^{\circ}\text{C}$  until use. Frozen muscle tissue was pulverized in liquid nitrogen and genomic DNA was extracted using a Genomic Tip 100/G kit (Qiagen) as previously described (Angthong et al., 2020). DNA quantity was measured using a NanoDrop ND-8000 spectrophotometer and a Qubit dsDNA BR Assay kit (Invitrogen, USA) using Qubit fluorometer. The DNA quality and integrity were visualized under pulsed-field gel electrophoresis at 80 volts for 9 h in 0.5x KBB buffer (51 mM Tris, 28 mM TASP, 0.08 mM EDTA, pH 8.7) (Sage Science) containing SYBR Safe DNA gel staining (Invitrogen).

### 2.2 | PacBio and Illumina library preparation and sequencing

Whole-genome sequencing was performed using long-read PacBio RS II and SEQUEL (Pacific Biosciences outsourced to NovogenAIT). The 15-kb and 20-kb SMRTbell libraries were constructed for the PacBio RSII and SEQUEL systems, respectively.

For short read sequencing, the paired-end library with 150 bp was prepared and sequenced using Illumina HiSeq 2000 (Illumina

outsourced to Macrogen). Illumina reads (133x sequence coverage) were used for error correction of the PacBio assembly.

### 2.3 | Chicago library preparation and sequencing

A Chicago library was prepared as described previously (Putnam et al., 2016). Approximately 500 ng of high molecular weight genomic DNA (mean fragment length = 60 kbp) was reconstituted into chromatin in vitro and fixed with formaldehyde. Fixed chromatin was digested with DpnII, the 5' overhangs filled in with biotinylated nucleotides, and then free blunt ends were ligated. After ligation, crosslinks were reversed to remove protein from DNA. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then sheared to ~350 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The library was sequenced on an Illumina HiSeq X Ten to produce 444 million  $2 \times 150$  bp paired-end reads, which provided 51.43x sequence coverage.

### 2.4 | Dovetail Hi-C library preparation and sequencing

A Dovetail Hi-C library was prepared in a similar way as described (Lieberman-Aiden et al., 2009). Briefly, chromatin was fixed in place with formaldehyde in the nucleus, and then extracted fixed chromatin was digested with DpnII, the 5' overhangs filled in with biotinylated nucleotides, and then free blunt ends were ligated. After ligation, crosslinks were reversed to remove protein from DNA. The purified DNA was then processed as similar as aforementioned in Chicago library preparation. The library was sequenced on an Illumina HiSeq X Ten to produce 430 million  $2 \times 150$  bp paired-end reads, which provided 49.8x sequence coverage.

### 2.5 | Genome assembly and scaffolding

Sequencing reads from Illumina, PacBio, and Dovetail (Chicago and Hi-C reads) were used to assemble *P. monodon* genome. PacBio sequence data were used for de novo assembly, and Illumina sequence data were subsequently used for polishing to obtain high-quality contigs. Chicago and Hi-C reads were used to scaffold the draft genome. In brief, high-quality Illumina reads were prepared using TrimGalore (<https://github.com/FelixKrueger/TrimGalore>) based on the following criteria: (i) no "N" base, (ii) trimming of adaptor sequences and low-quality bases ( $Q < 20$ ), (iii) no trimmed reads  $< 100$  bp. To avoid mis-assembly due to repetitive sequences (Tørresen et al., 2019), PacBio SEQUEL subreads with repetitive sequences comprised over 85% of total sequences were filtered out. The GC content criteria ( $< 25\%$  and  $> 85\%$ ) were applied for filtering low complexity

DNA sequences before assembly. Noted that GC content of crustaceans ranges around 35%–41% (Gao et al., 2017; Yu et al., 2015; Zhao et al., 2012). Moreover, the reads matched mitochondria sequence (NC\_002184.1) were excluded from nucleus sequences and processed separately (Supporting Information method). Reads  $\geq 5,000$  bp were assembled using WTDBG2 (Hu et al., 2019). Illumina reads were then aligned to the assembled contigs by minimap2 (Li, 2018) and polished using wtpoa-cns mode in WTDBG2 (Hu et al., 2019).

Scaffolding of the genome assemblies was performed using HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al., 2016). The input de novo assembly, shotgun reads, Chicago library reads, and Dovetail Hi-C library reads were used as input data for HiRise. An iterative analysis was conducted. First, shotgun and Chicago library sequences were aligned to the draft input assembly using a modified SNAP read mapper (<http://snap.cs.berkeley.edu>). The separations of Chicago read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold. After aligning and scaffolding Chicago data, Dovetail Hi-C library sequences were aligned and scaffolded following the same method. After scaffolding, shotgun sequences were used to close gaps between contigs.

The *P. monodon* genome sequence was aligned to the Pacific white shrimp *Litopenaeus vannamei* (Zhang et al., 2019) using Mugsy v1.2.3 (Angiuoli & Salzberg, 2011). Alignments with a length  $< 1$  kb were filtered out. The output alignments between genomes were visualized using Circos v0.69.9 (Krzywinski et al., 2009). The density of repetitive sequences was calculated as the percentage of genomic regions covered by simple repeat sequences in a 500-kb window. Gene density and GC content was calculated by counting the number of genes and the percentage of GC bases in a 500-kb window, respectively. Syntenic blocks were identified by MCScanX with criteria at least 10 syntenic genes and a maximum of six intervening genes allowed.

## 2.6 | Repetitive element analysis

Species-specific repeat library was generated using RepeatModeler2 (Flynn et al., 2020) prior to the masking with RepeatMasker version open-4.0.9 (Smit et al., 2013–2015). Annotation of repeats was aligned to Repbase using RMBlast (Bao et al., 2015). All processes were performed using default parameters.

## 2.7 | RNA isolation, PacBio Iso-Seq and ONT library preparation, sequencing and analysis

A total of nine organs (gill, heart, hepatopancreas, intestine, ovary, testis, pleopods, stomach and thoracic ganglia) and hemocytes

were harvested from two male and two female 4-month-old juvenile black tiger shrimp (Prachuap Khiri Khan, Thailand). Total RNA was extracted using TRI REAGENT according to the manufacturer's instructions (Molecular Research Center, USA). Contaminated genomic DNA was removed by treatment with DNase I at 0.5 U/ $\mu$ g total RNA at 37°C for 30 min. The DNA-free RNA was subjected for sequencing analysis using PacBio Iso-Seq SEQUEL platform (Pacific Biosciences outsourced to NovogenAIT) and Oxford Nanopore technologies (ONT) platform. Sequences obtained from Iso-Seq were prepared as previously described (Pootakham et al., 2020). Only RNA extracted from ovary, testis, hepatopancreas and intestine were sequenced using ONT platform. ONT libraries were prepared as previously described (Jenjaroenpun et al., 2020).

## 2.8 | Gene prediction and gene annotation

Gene prediction and protein-coding sequence identification were performed using a combination of transcriptome-based prediction, homology-based prediction, and ab initio prediction methods using EvidenceModeler (Haas et al., 2008) to generate consensus gene prediction for training species-specific parameter in AUGUSTUS (Stanke et al., 2008). To locate intron and exon regions, transcriptome-based prediction methods combined information from PacBio Iso-seq and other available *P. monodon* transcriptome databases (PRJNA4214000, SRR1648423, SRR1648424, SRR2191764, SRR2643301, SRR2643302, SRR2643304, SRR2643305) (Supporting Information Table S1) to align against the genome sequence. For short-read transcripts, STAR (Dobin et al., 2013) was employed to align against the genome before spliced read information was generated according to a previously published protocol (Hoff & Stanke, 2019) using bam2wig script in AUGUSTUS. Iso-Seq raw reads containing both 5' and 3' adapters (derived from full-length transcripts) were identified, and the adapters and poly (A) sequences were trimmed. Cleaned consensus reads were then mapped using Genomic Mapping and Alignment Program (GMAP) (Wu & Watanabe, 2005). Expressed sequence tags (Tassanakajon et al., 2006) were aligned to the genome using BLAT (Kent, 2002) and converted to potential gene structures using blat2hints script of AUGUSTUS.

Protein sequences of *P. monodon* were mapped against proteins from closely related organisms including *Hyalella azteca*, and *L. vannamei*, using Exonerate version 2.2.0 (Slater & Birney, 2005). All gene models derived from these three methods were integrated by EvidenceModeler into a high confident nonredundant gene set, which was used to set species-specific parameters. Finally, AUGUSTUS was used to predict genes in the genome based on the extrinsic evidence. Functional annotations of the obtained gene set were conducted using Semantic Annotation Platform (SAPP) using the InterProScan module (Jones et al., 2014; Koehorst et al., 2018) and Blast2GO (Götz et al., 2008).

To find homeobox (*Hox*) gene cluster in *P. monodon*, Basic Local Alignment Search Tool (BLAST) was used to identify predicted

proteins containing homeobox domains against Hox proteins in a model organism *Drosophila melanogaster*. Hox gene clusters in *Daphnia magna*, *H. azteca*, *Eurytemora affinis* and *Tigriopus californicus* were identified based on homologue search. (BLAST) The copy number of genes in Toll pathway in *P. monodon* and *L. vannamei* (GCF\_003789085.1) were obtained from the annotated sequences.

## 2.9 | Phylogenetic analysis

Phylogenetic analysis was conducted based on the mitochondrial and the nuclear genomes. Mitochondrial-based phylogenetic tree was constructed by deduced amino acids of 13 conserved core protein-coding genes (*nad2*, *cox1*, *cox2*, *atp8*, *atp6*, *cox3*, *nad3*, *nad1*, *nad5*, *nad4*, *nad4l*, *nad6*, and *cytb*). Each gene was aligned separately based on corresponding amino acids using the MUSCLE algorithm (Edgar, 2004) in MEGA X (Kumar et al., 2018). Alignments were trimmed at the terminal ends to avoid unambiguous alignment and concatenated to construct the phylogenetic tree. The evolutionary history was inferred by using the Maximum Likelihood method and General Reversible Mitochondrial model (Adachi & Hasegawa, 1996). Initial trees for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model and then selecting the topology with superior log likelihood value. The analysis involved 25 species including *Anopheles quadrimaculatus* (NC\_000875.1), *Anopheles gambiae* (NC\_002084.1), *Apis mellifera* (MT188686.1), *Artemia franciscana* (NC\_001620.1), *Ceratitis capitata* (NC\_000857.1), *Daphnia pulex* (NC\_000844.1), *Daphnia magna* (NC\_026914), *Drosophila yakuba* (NC\_001322.1), *Drosophila melanogaster* (NC\_024511.2), *Eurytemora affinis* (NC\_046694.1), *Halocaridina rubra* (NC\_008413.1), *H. azteca* (NC\_039403.1), *Ixodes hexagonus* (NC\_002010.1), *L. vannamei* (NC\_009626.1), *Ligia oceanica* (NC\_008412.1), *Locusta migratoria* (NC\_014891.1), *Macrobrachium rosenbergii* (NC\_006880.1), *Panulirus japonicus* (NC\_004251.1), *Parhyale hawaiiensis* (NC\_039402.1), *P. monodon*, *Portunus trituberculatus* (NC\_005037.1), *Procambarus alleni* (NC\_028447.1), *Rhipicephalus sanguineus* (JX416325.1), *Speleonectes tulumensis* (NC\_005938.1) and *Tigriopus californicus* (NC\_008831.2).

A nuclear genome-based phylogenetic tree was constructed for *P. monodon* using the available crustacean genomes, namely *T. californicus* (GCA\_007210705.1), *D. magna* (GCF\_003990815.1) (Lee et al., 2019), *D. pulex* (GCA\_000187875.1) (Colbourne et al., 2011), *E. affinis* (GCF\_000591075.1) (Eyun et al., 2017), *P. hawaiiensis* (Kao et al., 2016), *L. vannamei* (GCF\_003789085.1) and *H. azteca* (GCF\_000764305.1). Moreover, two insect genomes (*A. gambiae* and *D. melanogaster*) were included in the phylogenetic tree. The reported protein sequences were used for orthologous gene identification (15,577 protein sequences of *T. californicus*, 23,583: *D. magna*, 30,611: *D. pulex*, 30,425: *E. affinis*, 28,666: *P. hawaiiensis*, 33,273: *L. vannamei* and 22,749: *H. Azteca*, 14,102: *A. gambiae* and 23,583: *D. melanogaster*). Single-copy orthologues were identified using OrthoFinder (Emms & Kelly, 2019) and used for phylogeny construction using BEAST (Bouckaert et al., 2019). Besides the nuclear

genome-based phylogeny, protein families among the crustacean genomes were compared. The orthologous protein family was identified by clustering together using MMseq2 (Steinegger & Söding, 2017) with standard parameters of 0.5 identity and 0.5 sequence coverage. The pan-core protein family matrix was used to construct a pan-core protein family tree, which captures species-specific ecological niches, following the method presented by Snipen and Ussery (2010). The common and specific protein family among the crustacean genomes was present in the UpSet plot (Lex et al., 2014).

## 2.10 | Transcriptome sequencing and analyses of the large shrimp and small shrimp groups

Shrimp experiment was conducted at Shrimp Genetic Improvement Center (SGIC, Thailand). Sibling stock of black tiger shrimp used in this study was specific-pathogen free (guaranteed to be free of white spot syndrome virus, yellow head virus, Taura syndrome virus, *Vibrio parahaemolyticus*, and *Enterocytozoon hepatopenaei*). The shrimp were reared in the same controlled environment: tanks size 800 m<sup>2</sup> (25 shrimp/m<sup>2</sup>) and fed at a rate of 3% of shrimp bodyweight per day. Five-month-old female shrimp (*N* = 140) were randomly selected for bodyweight measurement (Figure S1). Two sample groups were separated based on the lowest and highest shrimp weights, called the “small shrimp” group and the “large shrimp” group, respectively. An average bodyweight of the lowest bodyweight of the small shrimp group (*n* = 15) was 13.46 ± 0.52 g, while the average bodyweight of the highest of the large shrimp group (*n* = 15) was 36.27 ± 1.96 g (Table S2). All hepatopancreas samples from the small and large groups were immediately frozen in liquid nitrogen and stored at -80°C until use. To extract total RNA, individual frozen hepatopancreas tissues were pulverized in liquid nitrogen and subjected to a TRI-REAGENT extraction method and DNase treatment as previously described (Yuan et al., 2017).

## 2.11 | Data preprocessing

To obtain short-read RNA sequences, 30 libraries (*n* = 15 for each group) were constructed using the HiSeq Library Preparation kit (Illumina) and sequenced using Illumina NovaSeq 6000. Illumina sequencing (150 pair-end reads) was performed at Omics Drive, Singapore. The raw reads were quality-filtered (*Q* < 20, >50 bp) and adapter-trimmed using TrimGalore (<https://github.com/FelixKrueger/TrimGalore>).

## 2.12 | Comparison between “reference genome-based” and “de novo transcriptome assembly-based” RNA-seq mapping

Reference genome-based RNA-seq mapping was conducted by mapping Illumina reads to the genome using STAR (Dobin et al., 2013). Number of mapped reads per gene of each sample was counted

TABLE 1 Assembly statistics of the *P. monodon* genome

	PacBio	PacBio + Chicago	PacBio + Chicago + Hi-C
Number of contigs/scaffolds	70,380	30,179	26,877
Number of contigs/scaffolds >1 kb	70,373	30,171	26,869
Number of contigs/scaffolds >50 kb	11,084	3,725	869
Number of contigs/scaffolds >1 Mb	6	569	44
Number of contigs/scaffolds >5 Mb	-	27	44
Number of contigs/scaffolds >10 Mb	-	1	43
Total length (bases)	2,389,954,102	2,394,031,700	2,394,363,600
Longest contig/scaffold (bases)	1,387,722	12,098,070	65,869,259
Mean contig/scaffold size (bases)	33,958	79,328	89,086
Contig/scaffold N50 (Mb)	0.079	1.157	44.862
Contig/scaffold L50	6,786	509	23
% N in scaffolds	-	0.17	0.18
Genome annotation			
Number of predicted gene models			31,640
Number of protein-coding genes			30,038
Average gene length (bp)			1,428
Number of annotated genes:			
Refseq			30,038
Uniprot			10,068
GO			22,243
InterPro			20,615

using HTSeq-count (Anders et al., 2015). The genes that could be mapped >1 reads were used to compare with the de novo transcriptome assembly-based RNA-seq mapping. The preprocessed reads from all samples were subjected to the de novo transcriptome assembly using Trinity with default parameters (Grabherr et al., 2011). To avoid redundant sequences, the longest isoform for each gene identified by Trinity was kept and consequently clustered using CD-HIT (Li & Godzik, 2006) with criteria of 98% sequence identity and 80% minimal alignment coverage for the shorter sequence. Open reading frame (ORF) was predicted using TransDecoder 5.5.0 (Haas et al., 2013). ORFs of at least 70 amino acids were annotated by BLAST against the NCBI reference proteins at E-value  $<10^{-3}$  in Blast2GO (Götz et al., 2008). The quality of these two methods was assessed by the number of contigs, contig length distribution, and gene model representation via BLAST (Camacho et al., 2009).

### 2.13 | Differential gene expression analysis of small shrimp group and large shrimp group

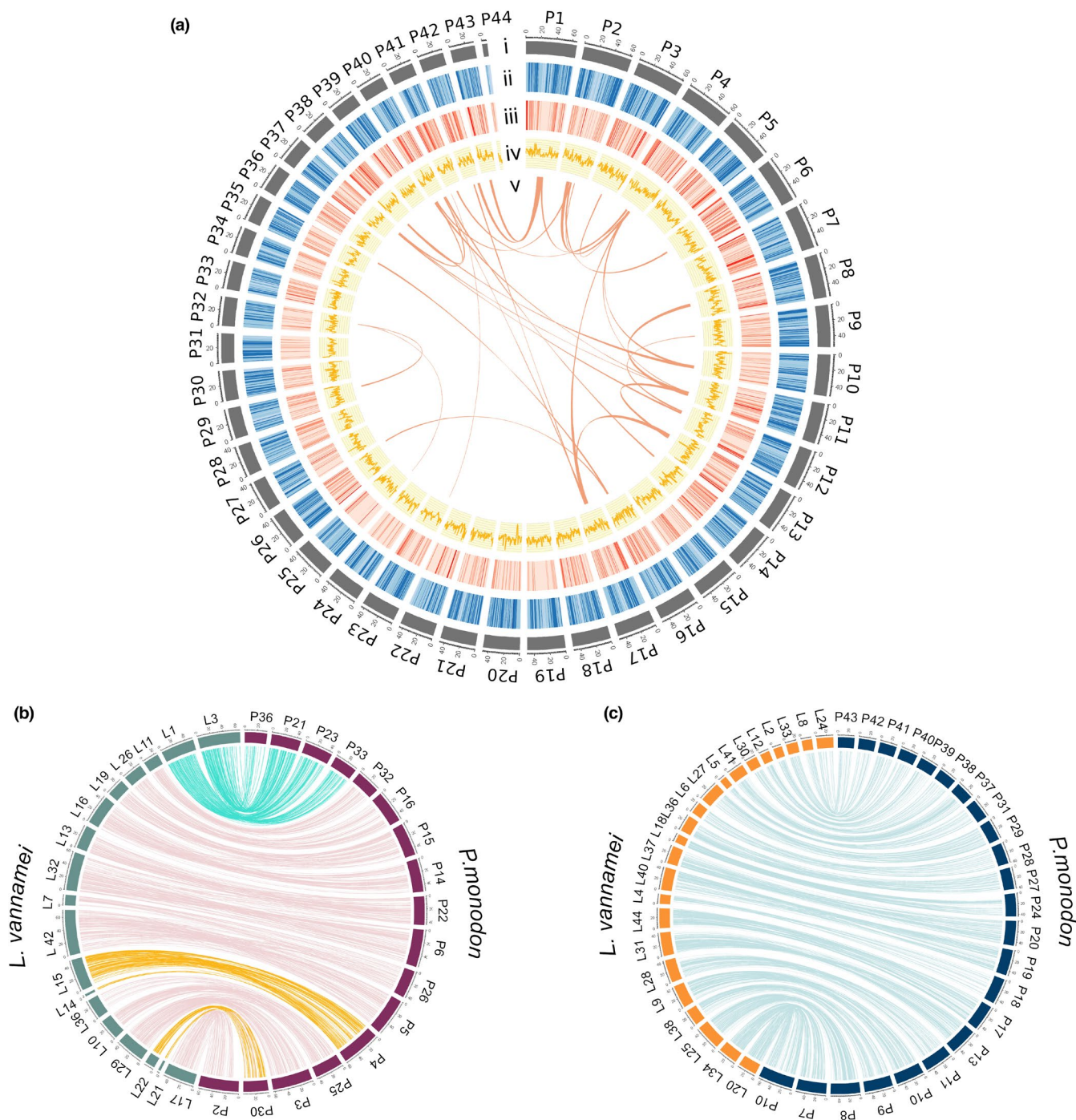
Based on the reference genome-based RNA-seq mapping, genes with low expression levels based on counts per million (CPM) values less than 1 in both groups were discarded from downstream analysis (Robinson et al., 2010). After that, differentially expressed genes (DEGs) between large and small shrimp were identified using DESeq2 (Love et al., 2014) using the following criteria: their expression level differences >2.0 change with Bonferroni adjusted *p*-value

of <0.05. Functional pathway analysis was carried using EggNOG (Huerta-Cepas et al., 2019) and KEGG mapper (Kanehisa et al., 2016).

## 3 | RESULTS

### 3.1 | Genome sequencing, assembly, and annotation

A whole-genome shotgun strategy was used to sequence and assemble a black tiger shrimp genome from PacBio long-read data. A total of 13,157,113 raw reads (178.94 Gb) representing 69.08× sequence coverage based on the estimated genome size of 2.59 Gb obtained from a previous report using the flow cytometry method (Swathi et al., 2018). De novo assembly of PacBio sequences yielded a preliminary assembly of 2.39 Gb (70,380 contigs) with a contig N50 of 79 kb and an L50 of 6786 contigs (Table 1). The draft genome was further assembled using the long-range Chicago (in vitro proximity ligation; 444 million read pairs) and Hi-C (in vivo fixation of chromosomes; 430 million read pairs) library data scaffolded with the HiRise software (Dovetail Genomics; Tables S2 and S3). The final assembly contained 44 pseudomolecules greater than 5 Mb in length (hereafter referred to as pseudochromosomes, numbered according to size; Figure 1a, track a and Figure S2), corresponding to the haploid chromosome number in *P. monodon* ( $1n = 44$ ,  $2n = 88$ ). The 44 pseudochromosomes covered 1,986,035,066 bases or 82.9% of the 2.39 Gb assembly with N content of 0.21%. Simple sequence



**FIGURE 1** *P. monodon* genome assembly and phylogenetic analysis. (a) Genomic landscape of 44 assembled *P. monodon* pseudochromosomes. (i) Physical map of *P. monodon* pseudochromosomes (Mb scale). (ii) Density of repetitive sequences represented by percentage of genomic regions covered by simple sequence repeats in 500-kb window. (iii) Gene density represented by number of genes in 500-kb window. (iv) GC content represented by percentage of G/C bases in 500-kb window. Syntenic blocks are depicted by connected lines. (v) Syntenic relationship of gene blocks among *P. monodon* pseudochromosomes. Syntenic blocks were identified by MCSanX with criteria at least ten syntenic genes and a maximum of six intervening genes allowed. (b) Diagrams showing collinearity between *P. monodon* and *L. vannamei* pseudochromosomes. Lines link the syntenic block with sequence coverage >1 kb and identity >90%. Most regions exhibit one-to-many relationship between *P. monodon* and *L. vannamei*. The yellow line represents the region on a single *P. monodon* pseudochromosome that exhibits synteny to regions on two *L. vannamei* pseudochromosomes, whereas the turquoise line represents syntenic regions on a single *L. vannamei* pseudochromosome and the two *P. monodon* pseudochromosomes. Black tiger shrimp pseudochromosomes are designated with "P" followed by pseudochromosome numbers, and Pacific white shrimp pseudochromosomes are designated with "L" followed by the pseudochromosome numbers. (c) Diagrams showing collinearity between *P. monodon* and *L. vannamei* where the syntenic relationship between pseudochromosomes is one-to-one

repeats (SSRs) occurred 22.2% per 500 Kb (Figure 1a, track b). On average, the gene density of *P. monodon* genome was 9 genes per 500 Kb (Figure 1a, track c) and GC content was 36.6% per 500 Kb (Figure 1a, track d).

To evaluate the quality of our de novo assembly, we aligned Illumina DNA short reads obtained from this work and the previous report (Yuan et al., 2018) to the assembled genome and found that approximately 93% of the DNA short reads could be mapped on our de novo assembly. We also aligned the publicly available RNA-seq reads (Huerlimann et al., 2018) and Iso-Seq reads (Pootakham et al., 2020) to the assembly and 90.22% and 98.77% of the RNA-seq and Iso-Seq reads were mapped to the assembly, respectively. To further assess the completeness of the genome assembly, we checked the gene content with the BUSCO software using the Eukaryota (odb9) database (Simão et al., 2015). Our gene prediction contained 94.72% of the highly conserved orthologues (85.15% complete, 9.57% partial, 5.28% missing) in the eukaryotic lineage. High mapping rates, comparable to the numbers observed in the *L. vannamei* genome, and a high percentage of identified highly conserved orthologues provided the evidence for a high-quality assembly obtained for the *P. monodon* genome.

The chromosome-level contiguity achieved in our assembly enabled a syntenic analysis between *P. monodon* and the Pacific white shrimp *L. vannamei*. Portions of conserved syntenic genes between the black tiger shrimp and Pacific white shrimp are illustrated in Figure 1b,c. The distribution of paralogous gene pairs revealed one-to-one synteny across 15 pseudochromosomes/linkage groups and one-to-two synteny in *P. monodon* pseudochromosomes 4 and 30 to *L. vannamei* pseudochromosomes 14/15 and 21/22, respectively (Figure 1b).

A combination of ab initio prediction, homology-based search, and transcript evidence obtained from both Iso-Seq and RNA-seq data was used for gene prediction. The genome annotations of *P. monodon* contained 31,640 predicted gene models, of which 30,038 were protein-coding genes (Table 1, Table S4). Of these protein-coding genes, 25,569 (85.02%) had evidence support from our RNA-seq and Iso-Seq data. In addition to the nuclear genome, a complete mitochondrial genome assembly was obtained with a size of 15,974 bp and 29.09% GC content. The mitochondria genome showed high sequence identity with the published *P. monodon* mitochondrial genomes (~98.8% identity in comparison to AF217843.1 and MN057663.1) (Figure S3).

### 3.2 | Comparative analyses between *P. monodon* and other crustacean genomes

The mitochondrial phylogenetic analysis of 25 arthropods based on 13 concatenated protein-coding genes showed that *P. monodon* was closely related to other species in order Decapoda. These decapods were the nearest neighbour to order Diptera, class Insecta and distantly apart from other crustaceans such as order Amphipoda, class Malacostraca (*H. azteca* and *P. haweiensis*) and order Cladocera, class Branchiopoda (*D. pulex*, *D. magna*; Figure 2a). In contrast to the mitochondrial-based phylogenetic tree, the nuclear-based

phylogenetic tree of single-copy orthologous genes of *P. monodon* genome and other publicly available crustacean genomes showed that *P. monodon* was clustered together with other crustacean species in class Malacostraca. Furthermore, the nearest neighbours of species at a higher taxonomic rank (Class level) presented in the mitochondrial-based phylogenetic tree and nuclear-based phylogenetic tree were different. The nuclear-based phylogenetic tree showed the nearest neighbours of insects (*A. gambiae* and *D. melanogaster*) were *Daphnia* spp. (order Cladocera, class Branchiopoda; Figure 2b). Nevertheless, species belonging to the same order, including *P. monodon* and *L. vannamei*, were clustered together in both phylogenetic trees.

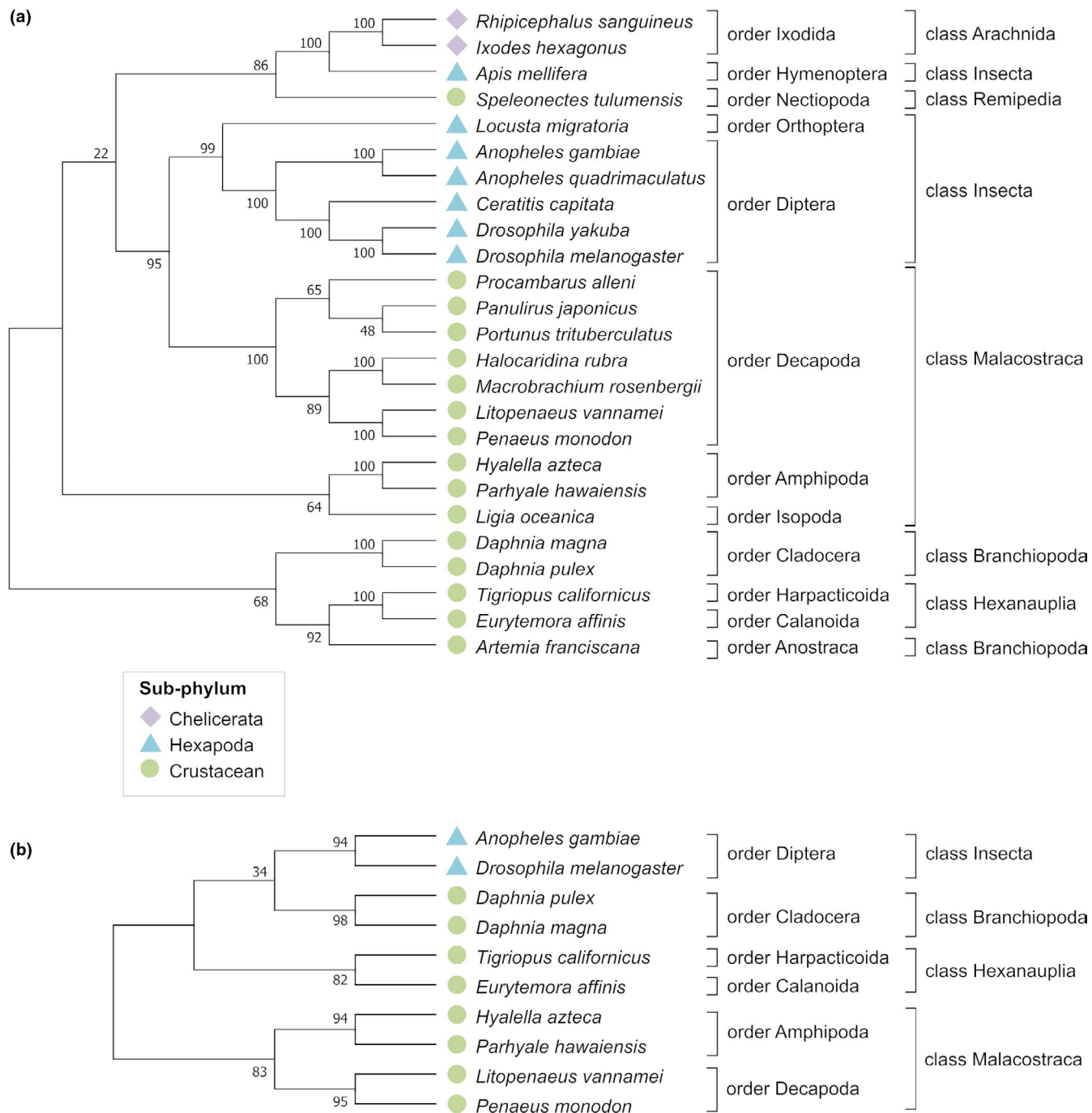
The identification of orthologous groups with other crustaceans showed 447 conserved protein families across the eight crustaceans and 363 common protein families shared among species in class Malacostraca (*P. monodon*, *L. vannamei*, *P. haweiensis* and *H. azteca*; Figure 3a). Among the examined crustacean genome sequences, *L. vannamei* has the highest number of shared orthologous protein families (5445 families) to *P. monodon*. *P. monodon* and *L. vannamei* were found to shared 493 and 84 orthologous protein families with *P. haweiensis* and *H. azteca*, respectively (Figure 3a).

Among the core gene families of the examined crustaceans, we investigated two highly conserved gene families, namely *Hox* and Toll pathway genes. *Hox* genes are involved in a wide range of critical developmental and physiological processes, ranging from embryonic development, innate immune homeostasis to whole-body regeneration (Chang & Lai, 2018). *Hox* genes are coregulated and often positioned relatively close to each other on the same chromosome (Akam, 1995). The *Hox* gene clusters have been shown to span a few kilobases (Ferrier & Akam, 1996) to several megabases (Baldwin-Brown et al., 2018). Our *P. monodon* assembly showed that all 10 *Hox* genes were clustered together on chromosome 37 (Figure 3b). Similarly, the other two species with chromosome-level genome assemblies (*T. californicus* and *D. magna*) also have *Hox* genes located on the same chromosome (Figure 3b). Some of the *Hox* genes in *L. vannamei* were found on the same linkage group (Zhang et al., 2019).

As the copy number variation of immune genes has been linked to pathogen tolerance in arthropods (Mitri et al., 2020), we compared the copy number of genes in the Toll pathway, an essential immune response to defend pathogen invasion, in *P. monodon* and *L. vannamei*. Most members of the Toll pathway had higher copy numbers in *L. vannamei* than in *P. monodon* with the exception of *MyD88* (Figure 3c). Interestingly, *Spaetzle*, *TRAF6*, *Dorsal* appeared to be single-copy genes in *P. monodon*.

### 3.3 | Repetitive elements in *P. monodon* genome

The total repeat content in the black tiger shrimp genome assembly was 62.5% (Figure 4a). We identified repeats comprising 572.87 Mb of SSRs (23.93%), 276.64 Mb of long interspersed nuclear elements (LINEs, 11.55%), 93.21 Mb of long terminal repeats (LTRs, 3.89%), 75.96 Mb of DNA elements (3.17%), 59.09 Mb of low complexity repeats (2.47%), 49.49 Mb of short interspersed nuclear elements

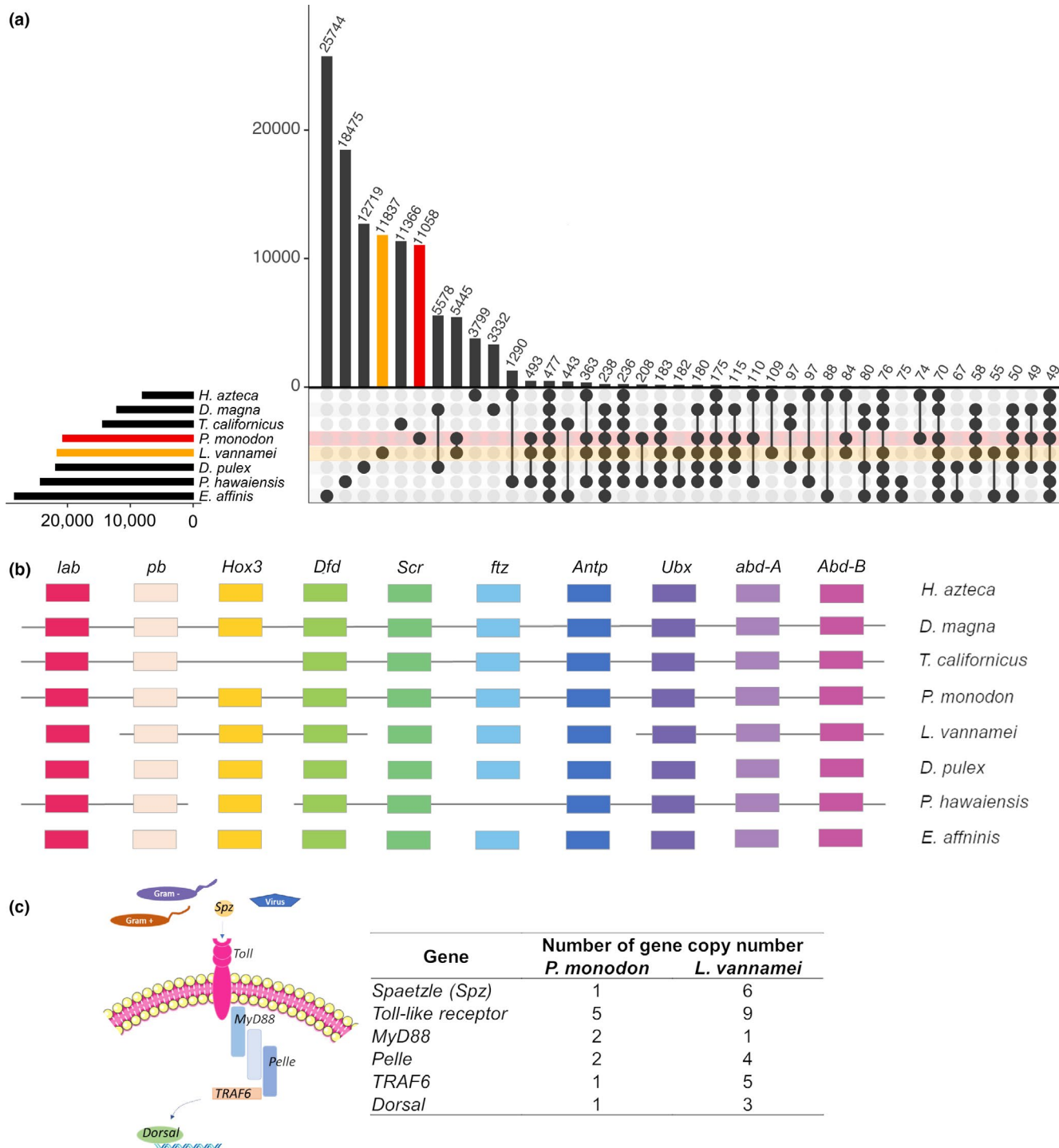


**FIGURE 2** Phylogenetic analysis of (a) mitochondrial protein coding genes and (b) single ortholog nucleus protein encoding genes. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The coloured symbols at the node denote the subphylum including Crustacean (green circles), Hexapoda (blue triangles) and Chelicerata (purple diamonds)

(SINEs, 2.07%), 2.18 Mb of small RNA (0.09%), and 368.13 Mb of unclassified repeat elements (15.37%). SSRs and LINEs are the two most abundant repeat categories in *P. monodon*, together accounting for 35.48% of the genome assembly. Moreover, LINE/I (4.65%), RTE-BovB (3.59%), and Penelope (0.96%) were the three major components of LINEs (Figure 4b). The major LINEs components found here were in accord with the previously reported *P. monodon* genome, but with different proportions (2.03%, 4.96%, and 0.82% for LINE-I, RTE-BovB, and Penelope, respectively) (Yuan et al., 2018).

### 3.4 | Comparison between “reference genome-based” and “de novo transcriptome assembly-based” RNA-seq mapping

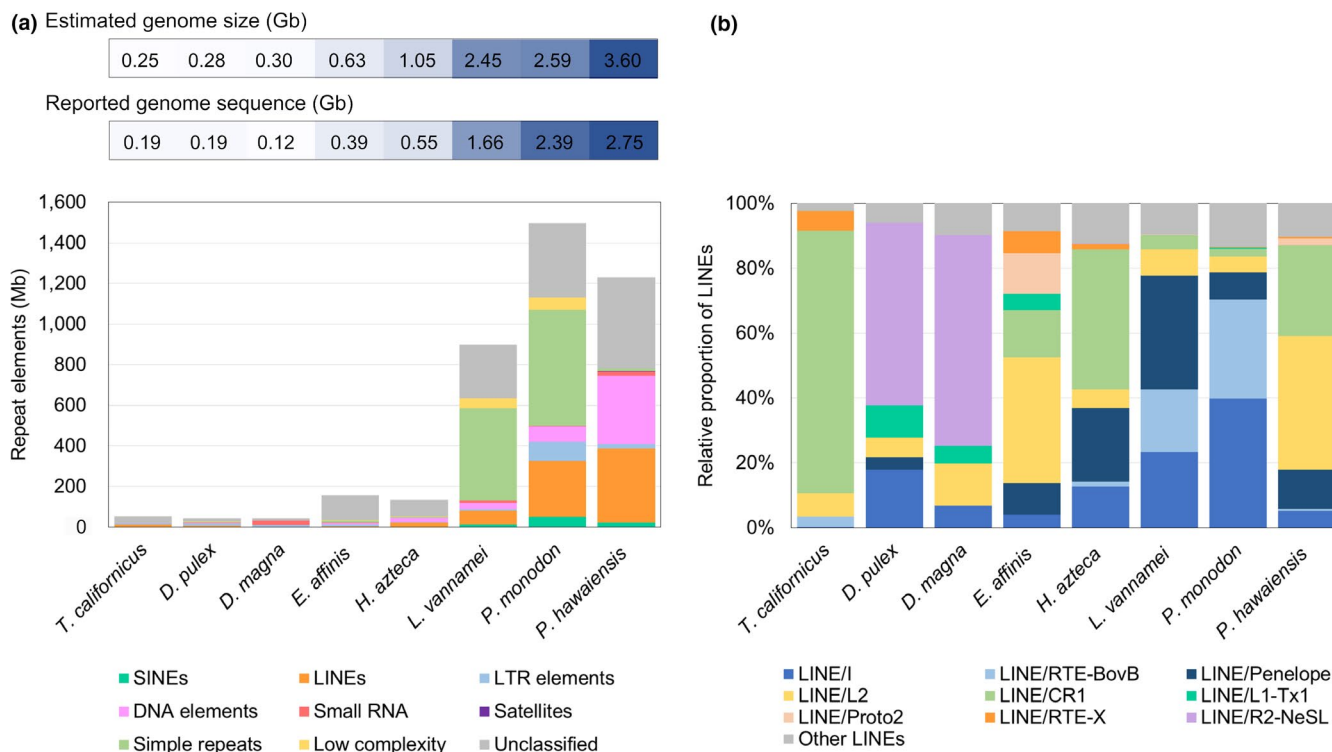
Transcriptome analysis is a useful approach for identifying differentially expressed genes (DEGs) among individuals with various traits of interest. Without a high-quality reference genome, gene expression studies rely on a de novo transcriptome assembly, which often contains a large number of fragmented transcript sequences with



**FIGURE 3** Comparative genomics. (a) UpSet plot represents pan-core protein family of crustacean genomes (*P. monodon*, *L. vannamei*, *P. hawaiiensis*, *E. affinis*, *H. azteca*, *T. californicus*, *D. pulex*, and *D. magna*). The vertical bars show the number of intersecting protein family among the crustacean genomes, denoted by the connected black circles below the histogram. The horizontal bars show the protein family size. (b) *Hox* gene clusters in the crustaceans. *Hox* gene cluster includes *labial* (*lab*), *proboscipedia* (*pb*), *Hox3/zen*, *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), *fushi tarazu* (*Ftz*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominal-A* (*Abd-A*), and *Abdominal-B* (*Abd-B*). The lines connecting genes indicate that the genes are in the same scaffold. (c) Copy number of genes involving the Toll pathway

no annotation. To evaluate the utility of this genome in transcriptomic analysis, we compared gene expression in the hepatopancreas between large shrimp and small shrimp using both reference genome-based and de novo transcriptome assembly-based RNA-seq

mapping. The reference genome-based RNA-seq mapping showed that raw reads were mapped to 19,032 predicted genes, with the N50 length of 1884 bp and the longest sequence of 27,633 bp. Over 87.37% of the 19,032 mapped genes were longer than 500 bp,



**FIGURE 4** Distribution of repeat elements in the crustacean genomes. (a) Repeat elements in *P. monodon* and other crustaceans including *T. californicus*, *D. pulex*, *D. magna*, *E. affinis*, *H. azteca*, *L. vannamei*, and *P. hawaiiensis*. (b) Relative percentage of each LINE type

roughly half (56.44%) were longer than 1,000 bp, and 95.92% could be annotated using BLAST (Figure S4A). De novo transcriptome assembly using the same set of RNA-seq reads yielded 60,846 non-redundant contigs with the N50 length of 930 bp and the longest sequence of 27,264 bp. Only 22.97% were longer than 500 bp, 12.57% were longer than 1000 bp, and 49.90% could be annotated using BLAST (Figure S4B).

### 3.5 | Identification of growth-associated genes

The genome assembly was used as a reference to identify genes related to growth performance by comparing gene expression profiles in the hepatopancreas of the large shrimp and the small shrimp from the same culture condition. The large shrimp grew at a faster rate and had twice the weight of the small shrimp under the same rearing condition at 5 months old. This analysis identified 383 genes exhibiting higher levels of expression in the large shrimp and 95 genes exhibiting higher levels of expression in the small shrimp ( $|\log_2 \text{fold-change}| > 1$  and  $p\text{-value} < 0.05$ ; Table S5). To further access gene interaction networks, KEGG pathway mapping was employed. The results showed that metabolic pathways (42 KEGG orthology gene IDs: KO), biosynthesis of secondary metabolites (15 KO), and lysosome (11 KO) were the top three pathways to which DEGs were categorized (Table S6). The remaining KEGG pathways consisted of few DEGs ( $< 5$  KO) in each pathway. The functions of DEGs were classified by Clusters of Orthologous Groups (COGs)

annotation into 23 categories (Supporting Information Figure S5). The top five highly enriched metabolic processes were (i) carbohydrate/ (ii) lipid/ (iii) amino acid transport and metabolisms, (iv) secondary metabolites biosynthesis, transport and catabolism, and (v) inorganic ion transport and metabolism (Figure 5a). For instance, DEGs involved in carbohydrate metabolism were *amylase (amy)*, *fructose-bisphosphate aldolase (aldo)*, *glyceraldehyde 3-phosphate dehydrogenase (gapdh)*, and *insulin-like growth factor receptor (insr)*. Genes involved in lipid metabolism were *nose resistant to fluoxetine (nrf)*, *lipase (lip)*, *glycerol-3-phosphate acyltransferase (gpat3)*, *acyl-CoA delta desaturase (scd)*, *long-chain-fatty-acid--CoA ligase (acsl)* and *elongation of very-long-chain fatty acids protein 4-like (elovl4)*. In addition, we further investigated the PI3 K-Akt signaling pathway in which four DEGs (*integrin beta isoform 1 (itgb1)*, *integrin alpha-4 like (itga4)*, *serine/threonine-protein kinase N isoform 1 (pkn1)* and *insr*) were found (Figure 5b). This pathway was reported to contribute to a variety of cellular functions such as transcription, cell survival and cell growth (Rauch et al., 2011).

Besides genes related to nutrient metabolisms, immune-related genes were also examined as they have been reported to affect growth performance. From the DEGs, we found several immune-related genes with higher levels of expression in large shrimp such as *heat shock 70 kDa (hsp70)*, *anti-lipopolysaccharide factor-like*, *laccase-4*, and *C-type lectin*. Immune-related genes with higher expression levels in small shrimp were such as *caspase-1-like*, *suppressor of cytokine signaling 2-like*, and *zinc finger protein OZF-like*.



**FIGURE 5** Expression of differentially expressed genes of small shrimps ( $n = 15$ ) and large shrimps ( $n = 15$ ). (a) The top five enriched COGs metabolism groups were carbohydrate transport and metabolism, secondary metabolites biosynthesis, transport and catabolism, inorganic ion transport and metabolism, lipid transport and metabolism and amino acid transport and metabolism. (b) Differentially expressed genes in the PI3 K-Akt pathway. The dagger (†) indicates genes that were further validated by quantitative real-time PCR

## 4 | DISCUSSION

In this study, we report a whole-genome sequence of *P. monodon* assembled from long-read PacBio data and scaffolded using long-range Chicago and Hi-C techniques. Our chromosome-scale assembly has shown tremendous improvement in contiguity and completeness compared to the previously reported *P. monodon* genomes (Quyen et al., 2020; Yuan et al., 2018). Based on a 10-Mb benchmark (Reference standard for genome biology, 2018), the assembly presented here is considered a high-quality reference genome as it has the N50 scaffold length of 44.86 Mb. It is also one of the highest-quality crustacean genomes currently available. Of the 45 crustacean genome sequences listed in the NCBI genome database, only three species (*Tigriopus japonicus* (PRJNA592403), *T. californicus* (PRJNA237968), and *Daphnia magna* (PRJNA612523)) have their genomes assembled at a chromosome level.

The present genome provides an invaluable resource for shrimp research. The availability of the chromosome-scale assembly allowed us to examine the syntenic relationship between the genomes of two economically important penaeid shrimp species: the black tiger shrimp and the Pacific white shrimp. The synteny analysis revealed a one-to-one relationship between most of the *P. monodon* and *L. vannamei* pseudochromosomes; however, a few pairs exhibited one-to-two synteny, suggesting that certain chromosomes derived from the common ancestor of *P. monodon* and *L. vannamei* were fragmented into two smaller pseudochromosomes in *L. vannamei* (for example, *P. monodon* pseudochromosome 4 displayed synteny to *L. vannamei* pseudochromosomes 14 and 15; Figure 1b) but remained as single pseudochromosome in *P. monodon*. Alternatively, there could have been scaffolding/assembly errors in either the genetic map-based assembly of the *L. vannamei* pseudochromosomes or our HiC-based *P. monodon* assembly.

When compared to other arthropod species, the *P. monodon* mitochondrial genome reveals an interesting evolutionary relationship. The phylogenetic tree showed a sister-group relationship of Insecta (Hexapoda) and Malacostraca (Crustacea), whereas nodes in other stem branches cannot be resolved as bootstrap values were lower than 70%. The observation that Malacostraca (Crustacea) was more closely related to Insecta (Hexapoda) than Branchiopoda (Crustacea) agrees with several previously reported molecular and morphological phylogenies (Legg et al., 2013; Tamone & Harrison, 2015; Wilson et al., 2000). However, the results contradicted morphological classification and other studies (Cook et al., 2005; Podsiadlowski & Bartolomeaus, 2005). In a molecular phylogenetic analysis, different genes may yield contradictory topologies for the same diversity group (Betancur et al., 2014). The contradictory results might be due to the differences

in data matrix and methodology (Schwentner et al., 2017; Timm & Bracken-Grissom, 2015). For example, as opposed to our study that used 13 protein-coding genes, when 12 protein-coding genes (excluding *atp8*) were used in the phylogenetic analysis, sister-taxa of Malacostraca and Branchiopoda was obtained (Cook et al., 2005). Similarly, when a large data set of 100 mitochondrial genomes of Pancrustacea was used in the analysis, a mitochondrial phylogenetic tree revealed the cluster between the Remipedia and the Cirripedia (Maxillopoda) (Carapelli et al., 2007). Together, these examples indicate the influence of analytical methods and the data size on the derived phylogenetic relationship.

Although mitogenomic data are commonly used in phylogenetics, it might not always be suitable for evolutionary study in higher taxonomy. Mitochondrial genome evolves at a faster rate than nuclear genome, thus the saturation of substitution rates in mitochondrial genome renders it less suitable for studying deep divergences (Burger et al., 2003; Lin & Danforth, 2004). Moreover, adaptive introgression events that were observed in nature among close relatives and recombination could not be captured by mitochondrial markers (Ballard & Whitlock, 2004; Spinks & Shaffer, 2009). Given the limitation of using mitochondrial sequences for evolutionary study, the availability of completely sequenced genomes should allow more accurate phylogenetic construction than using few mitochondrial marker genes in the phylogenetic trees.

The availability of genome sequences allows the identification of single copy orthologues, which have been suggested as more accurate markers for phylogenetic analysis (Teasdale et al., 2016; Washburn et al., 2017). The phylogenetic tree constructed based on single-copy orthologues from eight crustaceans and two selected insects with high-quality genome sequences confirmed that *L. vannamei* and *P. monodon* were sister taxa, agreeing with the mitochondrial-based phylogenetic tree and consistent with the highest number of shared orthologous protein families. On the other hand, the relationships among higher taxonomy levels observed in the nuclear-based phylogenetic tree mostly contradicted the results of the mitochondrial-based phylogenetic tree. For example, the nuclear-based phylogenetic tree showed class Insecta was closely related to *Daphnia* spp. (class Branchiopoda) whereas class Insecta was found closely related to class Malacostraca in the mitochondrial-based phylogenetic tree. Given that it is widely accepted that insects are closely related to crustaceans as both belong to the Pancrustacea (Tamone & Harrison, 2015; Wolff et al., 2017), the nuclear-based phylogenetic tree using information from the high-quality genome sequence agrees with the widely accepted taxonomy classification. Indeed, there are growing evidence that nuclear-based phylogenetic tree is more reliable to represent a lineage relationship than the mitochondrial-based counterpart (Springer et al., 2001; Wahlberg &

Wheat, 2008). Thus, having a high-quality genome assembly like in our case will undoubtedly facilitate accurate evolutionary study.

The availability of genome sequences enables the comparison of orthologous protein families among the crustaceans. While the examined crustaceans share only 477 protein families, *P. monodon* and *L. vannamei* shared as many as 5445 orthologous protein families confirming their close relationship. Among the orthologous proteins shared across the organisms examined in this study, we focused on Hox proteins, which are the key class of proteins that have been used to underpin the evolutionary of animal body plans and morphologies (Lin et al., 2004; Pavlopoulos et al., 2009). There are eight *Hox* genes in segment specification that are conserved during the diversification of arthropoda including *labial* (*lab*), *proboscipedia* (*pb*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*), and *Abdominal-B* (*Abd-B*). The other two *Hox* genes including *Hox3* and *fushi tarazu* (*ftz*) are commonly lost or duplicated and functionally diverged in some animals (Gibson, 2000). In this study, we found the loss of *Hox3* in *T. californicus* and *ftz* in *P. hawaiiensis* whereas both genes were present in *P. monodon*. The *pb* gene was previously reported to be absent in the *P. monodon* genome while being present in *L. vannamei* and *Marsupenaeus japonicus* (Yuan et al., 2017, 2018). Given the availability of the high-quality genome sequence, we could however identify the *pb* gene in *P. monodon*. Furthermore, we found that all the *Hox* genes are located on the same pseudochromosome (#37). Physical linkage of the *Hox* genes in other crustaceans could not be determined as only two genomes have been assembled at a chromosome level to date (*T. californicus* and *D. magna*).

To further demonstrate the usefulness of the genome sequence, we examined genes involved in the Toll pathway, one of the most important immune-related pathways in defence against bacterial and viral infection in both *L. vannamei* and *P. monodon*. Given differences in disease susceptibility between the two penaeid shrimp, we hypothesized that the copy number of immune-related genes might influence the susceptibility as it has been reported that the variation of gene copy number can be related to underlying biological differences (Katju & Bergthorsson, 2013). The copy numbers of genes involved in the Toll pathway were higher in *L. vannamei* than *P. monodon*, except for *MyD88* (Figure 3c). Besides the fact that gene duplication could be beneficial by increasing the expression of immune genes (Palmer & Jiggins, 2015), neofunctionalization of the duplicated gene might occur and subsequently affect pathogen resistance as reported in *Anopheles* spp. (Mitri et al., 2020). Specially, *A. stephensi* with a single copy of leucine-rich repeat immune factor (*Apl1*) gene and *A. gambiae* with three copies of *Apl1* showed the expanded *Apl1* genes contribute to tolerance of *Plasmodium* infection (Mitri et al., 2020). Thus, these higher copy numbers of immune-related genes in *L. vannamei* might enhance its defence against pathogens, potentially contributing to higher degrees of disease resistance in *L. vannamei* (Briggs et al., 2004; Rungrassamee et al., 2016). However, shrimp immunity also depends on other molecular pathways such as the JAK/STAT pathway, prophenoloxidase activating system, and antimicrobial proteins (Chayaburakul et al., 2005;

Raja et al., 2015). These pathways can now be further investigated using our genome assembly.

Long-read sequences from PacBio allow a contiguous assembly that spans repeat regions containing transposable elements and tandem repeats. We found that *P. monodon* has the highest repeat abundance (62.5%) when compared to five available genome sequences of crustacean species: *L. vannamei* (53.9%), *P. hawaiiensis* (44.7%), *E. affinis* (40.2%), *H. azteca* (24.3%) and *D. pulex* (22.0%). Moreover, the percentage of repeat elements observed in this assembly was substantially higher than those reported in previous studies using *P. monodon* fosmid libraries (51.8%) (Huang et al., 2011) and draft genome (46.8%) (Yuan et al., 2018). Our result revealed that *P. monodon* genome contains highest proportion of SSRs along with *L. vannamei* (23.93%) (Zhang et al., 2019). These high repeat content and long repetitive sequences undoubtedly have hindered genome assembly in the previous attempts (Quyen et al., 2020; Yuan et al., 2018). With the current chromosome-scale assembly, an in-depth investigation of repeat elements is possible. Even though the biological function of repeats has not been well studied in shrimp, they are associated with important functions in insects such as adaptations (Christmas et al., 2019; Gilbert et al., 2020), sex chromosome differentiation (Rosolen et al., 2018) and morphology (Bouchebti & Arganda, 2020). In crustaceans, the proliferation of repeat elements has been reported to cause the expansion in genome sizes (Alfsnes et al., 2017). From the comparison of repeat contents among the examined crustacean species, we observed that higher repeat contents are often associated with larger genomes (Figure 4b). Although there has not been any study on the implication of their larger genome sizes, we speculate that an increase in genome size might in part be a consequence of the expansion of repeat elements that might contribute to fitness-promoting in responding to their habitats and growth (Talla et al., 2017).

With this high-quality reference genome assembly, we were able to obtain an improved gene set with better contiguity and annotation rate. Of the predicted genes in the genome, 95.01% could be functionally annotated. The results suggested that our genome assembly could serve as a high-quality reference for facilitating transcriptomic studies in the black tiger shrimp over the de novo transcriptome assembly-based RNA-seq mapping. The reference genome-based RNA-seq mapping was carried out for small and large shrimp groups to unravel genes related to growth performance.

Growth is undoubtedly an important factor for profitable shrimp production (Benzie et al., 2001; Cheng & Chen, 1990). Domesticated shrimp do not mature well with declining growth rates over generations (Jackson & Wang, 1998). Hence, a wild black tiger shrimp broodstock is still necessary. Notwithstanding its importance, knowledge of genes controlling shrimp's growth remained limited partly due to the lack of a high-quality reference genome. Thus, up to now most growth-related genes identified in *P. monodon* were taken from *L. vannamei* (Gao et al., 2015, 2017; Santos et al., 2018). In this study, reference-based transcriptome analysis was used to elucidate genes associated with growth in hepatopancreas.

Hepatopancreas has an important function in nutrient digestion and absorption, energy storage for tissue growth, and energy consumption during the growth stages (Liong Ong & Johnston, 2006; Wang et al., 2014). Furthermore, hepatopancreas is involved in a synthesis of vitellogenin and several sex steroid hormones (Kung et al., 2004). The comparison of gene expression in hepatopancreas, therefore, could reveal genes related to growth and pave the way to manage shrimp farming, which consequently reduces natural exploitation.

The comparison of hepatopancreas transcriptomes of small and large shrimp revealed that DEGs were mainly involved in nutrient metabolism, which is in concordance with the hepatopancreas functions in feed utilization and energy storage. Growth is a complex regulatory process in which we could observe that DEGs were presented in several pathways. We further investigated DEGs related to metabolisms of carbohydrate and lipid as they are the main nutrients reported to enhance growth in shrimp (Coelho et al., 2019; González-Félix et al., 2002; Hu et al., 2019; Olmos et al., 2011). In agreement with the prior finding that the enhancement of digestive enzyme activities improves growth performance of shrimp (Anand et al., 2013; Gómez & Shen, 2008), genes with higher expression levels in the large shrimp group were found to be involved in secondary metabolite biosynthesis and nutrient metabolism. Moreover, the higher expression level of carbohydrate metabolism genes such as *aldo*, *gapdh* and *insr* in the large shrimp might suggest that the large shrimp have a more effective digestion and nutrient utilization than the small shrimp.

Furthermore, DEGs involving fatty acid metabolism, which is necessary for biological structure and the function of cell membranes, were also observed. Previously, it was believed that penaeid shrimp could not de novo synthesize polyunsaturated fatty acids (PUFAs) such as linoleic acid (LOA, 18:2n-6) and alpha-linolenic acid (LNA, 18:3n-3), and could not elongate PUFA to highly unsaturated acids (HUFAs) (Wouters et al., 2001). However, genes involving in fatty acid biosynthesis and elongation have been reported in *L. vannamei* (Chen et al., 2015). Similarly, we could find enzyme genes involved in HUFA biosynthesis such as *acyl-CoA delta desaturase (scd)* and *elongation of very-long-chain fatty acids protein (elovl)* in our *P. monodon* genome sequence (Figure 5). From the transcriptomic result, *acsl* and *scd* were highly expressed in the large shrimp group. *Acsl* has a role in conversion of long-chain fatty acids to the active form acyl-CoAs, which contribute to enhance fatty acid uptake and its metabolism in downstream processes (Mashek et al., 2007). *Scd* involves in converting LOA and LNA to HUFAs, which subsequently contribute to membrane phospholipid biosynthesis (Mansilla et al., 2008). The higher expression levels of the genes regulated fatty acid metabolism in shrimp might not only be important knowledge for shrimp biology but also provide a crucial guide for feed formulations to promote shrimp growth.

Among the DEGs, we also found several genes involved in insulin signaling pathway with significantly higher expression levels in the large shrimp group such as *trf*, *pkn* and *insr* (Figure S6). *Trf* is an insulin-like growth factor that stimulates both proliferation and differentiation in a cell line (Kiepe et al., 2005). *Pkn1*, *itga4*, and *insr*

belonged to the PI3 K-Akt pathway, which is involved in growth regulation by controlling glucose balance through cross-talking with the insulin signaling pathway (Shi & He, 2016). Although an association between the PI3 K-Akt pathway and shrimp growth performance has not been well investigated in *P. monodon*, a link between the PI3 K-Akt growth factors and cellular survival has been reported in *Drosophila* (Schmitt et al., 2015) and *Daphnia* (Zhou et al., 2020). The insulin signaling pathway plays a key role in carbohydrate metabolism and growth regulation and its function are probably conserved in both vertebrate and invertebrate (Boucher et al., 2010; Herran et al., 2018). Thus, with their potential growth-related functions and higher expression levels in the large shrimp, the genes associated with PI3 K-Akt pathway present as possible marker candidates to facilitate black tiger shrimp genomic breeding programs.

In addition, immune-related genes were found among DEGs from the comparison between the large and small shrimp. Immune-related genes have previously been reported to not only play important roles in pathogen defence, but also in growth of shrimp (Jung et al., 2013). For example, *C-type lectin* and *hsp70* were expressed at a higher level in the large shrimp groups. *C-type lectin*, a known immune-related gene in shrimp, was induced by probiotics and subsequently improved the growth and survival of *L. vannamei* (Chai et al., 2016). *Hsp70* is one of the most conserved genes responding to various stresses such as temperature changes and pathogen invasion by helping the cells to recover from damaged proteins and in turn increasing the survival probability and life span (Valenzuela-Castillo et al., 2019). For example, the grass thrip, *Anaphothrips obscurus*, which have three wing morphs including macropters (winged), koelopters (intermediate), and brachypters (wingless) showed the brachypters were more tolerant to thermal stress than macropters, which corresponded to the significant up-regulation of *hsp70* in brachypters in comparison to macropters (Guo & Feng, 2018). In *Daphnia*, the decline of *hsp70* expression resulted in a shorter life span (Schumpert et al., 2014). Thus, our finding of higher expression of stress gene *hsp70* in the large shrimp might suggest that better ability to cope with stress in the large shrimp may contribute to better growth.

In conclusion, we have successfully overcome the technical challenges in obtaining the first high-quality chromosome-scale genome assembly of the economically important *P. monodon*. The investigation of genome characteristics and functional features yields a fundamental resource for deciphering the molecular mechanisms underlying the growth performance. The availability of this reference genome enables several downstream biological and industrial applications that would otherwise be difficult. This reference genome will benefit not only the *P. monodon* research community but also other researchers working on related shrimp and crustacean species, allowing for in-depth ecological research.

## ACKNOWLEDGEMENTS

The authors acknowledge NSTDA Supercomputer Center (ThaiSC) for providing HPC resources to conduct data analysis. We thank Ms Somjai Wongtripop, Mr Kaidtisak Kaewlok, and Shrimp Genetic

Improvement Center (SGIC) members, as well as Ms Sudtida Phuengwas and Ms Nuchjaree Pisamai (BIOTEC) for the shrimp sample collection. This project was funded by National Center for Genetic Engineering and Biotechnology (BIOTEC Platform No. P1651718), National Science and Technology Development Agency (NSTDA TDG Food and Feed program, No. P1950419) and partially supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 734486 (SAFE-Aqua). Visiting Professor Program 2019 was awarded to Dr Intawat Nookaew by NSTDA, Thailand. National Institute of General Medical Sciences of the National Institutes of Health (P20GM125503) awarded to Dr Intawat Nookaew. We acknowledge Zulema Dominguez (University of Arkansas for Medical Sciences, USA) for technical assistance on the comparative genome. We are grateful to Professor Morakot Tanticharoen (NSTDA), Dr Kanyawim Kirtikara (BIOTEC), Dr Wonnop Visessanguan (BIOTEC) and Associate Professor Skorn Koonawootrittriron (Kasetsart University, Thailand) for their mentorship on the shrimp research programmes.

#### AUTHOR CONTRIBUTIONS

Pacharaporn Anghthong, Kanchana Sittikankaew, Sopacha Arayamethakorn, Rungnapa Leelatanawit and Jasper J. Koehorst collected samples. Pacharaporn Anghthong, Kanchana Sittikankaew, Sopacha Arayamethakorn, Thidathip Wongsurawat, Rungnapa Leelatanawit, Duangjai Sangsrakru extracted DNA and RNA for sequencing. Tanaporn Uengwetwanit, Wirulda Pootakham, and Chutima Sonthirod carried out genome assembly and assessment. Tanaporn Uengwetwanit, Jasper J. Koehorst, Peter Schaap and Vitor Martins dos Santos performed gene prediction and annotation. Tanaporn Uengwetwanit, Wirulda Pootakham, and Chutima Sonthirod performed repeat analysis. Intawat Nookaew, Thidathip Wongsurawat, and Piroon Jenjaroenpun performed ONT sequencing and analysis. Intawat Nookaew and Piroon Jenjaroenpun carried out comparative genome analysis. Tanaporn Uengwetwanit, Sopacha Arayamethakorn, Wanilada Rungrasamee and Nitsara Karoonuthaisiri performed differential expression analysis. Wirulda Pootakham, Intawat Nookaew, Frédéric Tangy and Nitsara Karoonuthaisiri conceived this project. Tanaporn Uengwetwanit, Wirulda Pootakham, Intawat Nookaew, and Nitsara Karoonuthaisiri wrote the manuscript. All authors contributed to the final manuscript editing.

#### DATA AVAILABILITY STATEMENT

Genomic sequences, Chicago and Hi-C data used for generating the *P. monodon* genome assembly were deposited in NCBI under the accession number: PRJNA611030, SRX9895096 and SRX9895095 respectively. The genome sequences and annotation can be accessed from the *P. monodon* website (<http://www.biotec.or.th/pmonodon/index.php>). The raw RNA-seq data used for transcriptome analysis in this study were deposited in NCBI under accession number: PRJNA680027.

#### ORCID

Tanaporn Uengwetwanit  <https://orcid.org/0000-0003-4710-2613>

Wirulda Pootakham  <https://orcid.org/0000-0001-6721-6453>

Intawat Nookaew  <https://orcid.org/0000-0001-8901-1088>

Nitsara Karoonuthaisiri  <https://orcid.org/0000-0003-0838-8399>

#### REFERENCES

- Adachi, J., & Hasegawa, M. (1996). Model of amino acid substitution in proteins encoded by mitochondrial DNA. *Journal of Molecular Evolution*, 42(4), 459–468. <https://doi.org/10.1007/BF02498640>
- Agardy, T. (2000). Effects of fisheries on marine ecosystems: a conservationist's perspective. *ICES Journal of Marine Science*, 57(3), 761–765. <https://doi.org/10.1006/jmsc.2000.0721>
- Akam, M. (1995). Hox Genes and the evolution of diverse body plans. *Philosophical Transactions: Biological Sciences*, 349(1329), 313–319.
- Alfsnes, K., Leinaas, H. P., & Hessen, D. O. (2017). Genome size in arthropods; different roles of phylogeny, habitat and life history in insects and crustaceans. *Ecology and Evolution*, 7(15), 5939–5947. <https://doi.org/10.1002/ece3.3163>
- Anand, P. S. S., Kohli, M. P. S., Roy, S. D., Sundaray, J. K., Kumar, S., Sinha, A., Pailane, G. H., & Sukhame, M. (2013). Effect of dietary supplementation of periphyton on growth performance and digestive enzyme activities in *Penaeus monodon*. *Aquaculture*, 392–395, 59–68. <https://doi.org/10.1016/j.aquaculture.2013.01.029>
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Anderson, J. L. (2019). GOAL 2019: Global shrimp production review. <https://www.aquaculturealliance.org/advocate/goal-2019-global-shrimp-production-review/>. Accessed July 14, 2020.
- Angel, C., Francisco, C.-A., & Patricio De Los, R. (2010). Amphipoda and Decapoda as potential bioindicators of water quality in an urban stream (38°S, Temuco, Chile). *Crustaceana*, 83(8), 897–902. <https://doi.org/10.1163/001121610X502948>
- Angiuoli, S. V., & Salzberg, S. L. (2011). Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics (Oxford, England)*, 27(3), 334–342. <https://doi.org/10.1093/bioinformatics/btq665>
- Anghthong, P., Uengwetwanit, T., Pootakham, W., Sittikankaew, K., Sonthirod, C., Sangsrakru, D., Yoocha, T., Nookaew, I., Wongsurawat, T., Jenjaroenpun, P., Rungrasamee, W., & Karoonuthaisiri, N. (2020). Optimization of high molecular weight DNA extraction methods in shrimp for a long-read sequencing platform. *PeerJ*, 8, e10340. <https://doi.org/10.7717/peerj.10340>
- Baldwin-Brown, J. G., Weeks, S. C., & Long, A. D. (2018). A new standard for crustacean genomes: The highly contiguous, annotated genome assembly of the clam shrimp *Eulimnadia texana* reveals HOX gene order and identifies the sex chromosome. *Genome Biology and Evolution*, 10(1), 143–156. <https://doi.org/10.1093/gbe/evx280>
- Ballard, J. W., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, 13(4), 729–744. <https://doi.org/10.1046/j.1365-294x.2003.02063.x>
- Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, 6(1), 11. <https://doi.org/10.1186/s13100-015-0041-9>
- Baranski, M., Gopikrishna, G., Robinson, N. A., Katneni, V. K., Shekhar, M. S., Shanmugakarthik, J., Jothivel, S., Gopal, C., Ravichandran, P., Kent, M., Arnyasi, M., & Ponniah, A. G. (2014). The development of a high density linkage map for black tiger shrimp (*Penaeus monodon*) based on cSNPs. *PLoS One*, 9(1), e85413. <https://doi.org/10.1371/journal.pone.0085413>

- Benzie, J. A. H., Kenway, M., & Ballment, E. (2001). Growth of *Penaeus monodon* x *Penaeus esculentus* tiger prawn hybrids relative to the parental species. *Aquaculture*, 193(3–4), 227–237. [https://doi.org/10.1016/S0044-8486\(00\)00487-7](https://doi.org/10.1016/S0044-8486(00)00487-7)
- Betancur, R. R., Naylor, G. J., & Ortí, G. (2014). Conserved genes, sampling error, and phylogenomic inference. *Systematic Biology*, 63(2), 257–262. <https://doi.org/10.1093/sysbio/syt073>
- Bondad-Reantaso, M. G., Subasinghe, R. P., Josupeit, H., Cai, J., & Zhou, X. (2012). The role of crustacean fisheries and aquaculture in global food security: past, present and future. *Journal of Invertebrate Pathology*, 110(2), 158–165. <https://doi.org/10.1016/j.jip.2012.03.010>
- Bouchebti, S., & Arganda, S. (2020). Insect lifestyle and evolution of brain morphology. *Current Opinion in Insect Science*, 42, 90–96. <https://doi.org/10.1016/j.cois.2020.09.012>
- Boucher, P., Ditlecadet, D., Dubé, C., & Dufresne, F. (2010). Unusual duplication of the insulin-like receptor in the crustacean *Daphnia pulex*. *BMC Evolutionary Biology*, 10, 305. <https://doi.org/10.1186/1471-2148-10-305>
- Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., D Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M. A., ... Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 15(4), e1006650. <https://doi.org/10.1371/journal.pcbi.1006650>
- Briggs, M., Funge-smith, S., Subasinghe, R., & Phillips, M. (2004). Introductions and movement of *Penaeus vannamei* and *Penaeus stylirostris* in Asia and the Pacific. <http://www.fao.org/3/ad505e/ad505e00.htm>. Accessed July 17, 2020
- Brooker, A. L., Benzie, J. A. H., 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. <https://doi.org/10.1007/s002270050017>
- Burger, G., Gray, M. W., & Lang, B. F. (2003). Mitochondrial genomes: anything goes. *Trends in Genetics*, 19(12), 709–716. <https://doi.org/10.1016/j.tig.2003.10.012>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1), 421. <https://doi.org/10.1186/1471-2105-10-421>
- Carapelli, A., Liò, P., Nardi, F., van der Wath, E., & Frati, F. (2007). Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. *BMC Evolutionary Biology*, 7(2), S8. <https://doi.org/10.1186/1471-2148-7-S2-S8>
- Chai, P. C., Song, X. L., Chen, G. F., Xu, H., & Huang, J. (2016). Dietary supplementation of probiotic *Bacillus* PC465 isolated from the gut of *Fenneropenaeus chinensis* improves the health status and resistance of *Litopenaeus vannamei* against white spot syndrome virus. *Fish and Shellfish Immunology*, 54, 602–611. <https://doi.org/10.1016/j.fsi.2016.05.011>
- Chang, W. H., & Lai, A. G. (2018). A TALE of shrimps: Genome-wide survey of homeobox genes in 120 species from diverse crustacean taxa. *F1000Research*, 7, 71. <https://doi.org/10.12688/f1000research.13636.1>
- Chayaburakul, K., Lightner, D. V., Sriurairattana, S., Nelson, K. T., & Withyachumrannkul, B. (2005). Different responses to infectious hypodermal and hematopoietic necrosis virus (IHHNV) in *Penaeus monodon* and *P. vannamei*. *Diseases of Aquatic Organisms*, 67(3), 191–200. <https://doi.org/10.3354/dao067191>
- Chen, K., Li, E., Li, T., Chang, X. U., Wang, X., Lin, H., Qin, J. G., & Chen, L. (2015). Transcriptome and molecular pathway analysis of the hepatopancreas in the Pacific White Shrimp *Litopenaeus vannamei* under Chronic Low-Salinity Stress. *PLoS One*, 10(7), e0131503. <https://doi.org/10.1371/journal.pone.0131503>
- Cheng, C. S., & Chen, L.-C. (1990). Growth characteristics and relationships among body length, body weight and tail weight of *Penaeus monodon* from a culture environment in Taiwan. *Aquaculture*, 91(3), 253–263. [https://doi.org/10.1016/0044-8486\(90\)90192-P](https://doi.org/10.1016/0044-8486(90)90192-P)
- Christmas, M. J., Wallberg, A., Bunikis, I., Olsson, A., Wallerman, O., & Webster, M. T. (2019). Chromosomal inversions associated with environmental adaptation in honeybees. *Molecular Ecology*, 28(6), 1358–1374. <https://doi.org/10.1111/mec.14944>
- Coelho, R. T. I., Yasumaru, F. A., Passos, M. J. A. C. R., Gomes, V., & Lemos, D. (2019). Energy budgets for juvenile Pacific whiteleg shrimp *Litopenaeus vannamei* fed different diets. *Brazilian Journal of Oceanography*, 67, e19243. <https://doi.org/10.1590/s1679-87592019024306701>
- Colbourne, J. K., Pfrender, M. E., Gilbert, D., Kelley Thomas, W., Tucker, A., Oakley, T. H., Tokishita, S., Aerts, A., Arnold, G. J., Basu, M. K., Bauer, D. J., Cáceres, C. E., Carmel, L., Casola, C., Choi, J.-H., Detter, J. C., Dong, Q., Dusheyko, S., Eads, B. D., & Boore, J. L. (2011). The ecoresponsive genome of *Daphnia pulex*. *Science*, 331(6017), 555–561. <https://doi.org/10.1126/science.1197761>
- Cook, C. E., Yue, Q., & Akam, M. (2005). Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. *Proceedings Biological Sciences*, 272(1569), 1295–1304. <https://doi.org/10.1098/rspb.2004.3042>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5(1), 113. <https://doi.org/10.1186/1471-2105-5-113>
- Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20(1), 238. <https://doi.org/10.1186/s13059-019-1832-y>
- Eyun, S.-I., Soh, H. Y., Posavi, M., Munro, J. B., Hughes, D. S. T., Murali, S. C., Jiaxin, Q. U., Dugan, S., Lee, S. L., Chao, H., Dinh, H., Han, Y. I., Doddapaneni, H. V., Worley, K. C., Muzny, D. M., Park, E.-O., Silva, J. C., Gibbs, R. A., Richards, S., & Lee, C. E. (2017). Evolutionary history of chemosensory-related gene families across the Arthropoda. *Molecular Biology and Evolution*, 34(8), 1838–1862. <https://doi.org/10.1093/molbev/msx147>
- FAO (2018). The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals.
- FAO (2020). The State of World Fisheries and Aquaculture 2020. Sustainability in action.
- Farkas, A., Salánki, J., & Varanka, I. (2003). Crustaceans as biological indicators of heavy metal pollution in Lake Balaton (Hungary). *Hydrobiologia*, 506(1), 359–364. <https://doi.org/10.1023/B:HYDR.0000008615.86330.1d>
- Ferrier, D. E. K., & Akam, M. (1996). Organization of the Hox gene cluster in the grasshopper, *Schistocerca gregaria*. *Proceedings of the National Academy of Sciences*, 93(23), 13024. <https://doi.org/10.1073/pnas.93.23.13024>
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2: automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, 117(17), 9451–9457. <https://doi.org/10.1073/pnas.1921046117>
- Gao, Y., Wei, J., Yuan, J., Zhang, X., Li, F., & Xiang, J. (2017). Transcriptome analysis on the exoskeleton formation in early developmental stages and reconstruction scenario in growth-moulting in *Litopenaeus vannamei*. *Scientific Reports*, 7(1), 1098. <https://doi.org/10.1038/s41598-017-01220-6>
- Gao, Y., Zhang, X., Wei, J., Sun, X., Yuan, J., Li, F., & Xiang, J. (2015). Whole transcriptome analysis provides insights into molecular mechanisms for molting in *Litopenaeus vannamei*. *PLoS One*, 10(12), e0144350. <https://doi.org/10.1371/journal.pone.0144350>

- Garcia, S. M. Z. A., Aliaume, C., Do Chi, T., & Lasserre, G. (2003). Fisheries impact on the ecosystem. In S. M. Garcia (Ed.), *The ecosystem approach to fisheries. Issues, terminology, principles, institutional foundations, implementation and outlook* (Vol. 443, pp. 71). FAO Fisheries Department.
- Gibson, G. (2000). Evolution: Hox genes and the cellared wine principle. *Current Biology*, 10(12), R452–R455. [https://doi.org/10.1016/S0960-9822\(00\)00531-5](https://doi.org/10.1016/S0960-9822(00)00531-5)
- Gilbert, C., Peccoud, J., & Cordaux, R. (2020). Transposable elements and the evolution of insects. *Annual Review of Entomology*, 66, 355–372. <https://doi.org/10.1146/annurev-ento-070720-074650>
- Gislason, H. (2003). The effect of fishing on non-target species and ecosystem structure and function. In Sinclair, M., & Valdimarsson, G. (Eds.), *Responsible fisheries in the marine ecosystem* (pp. 255–274). FAO, CABI publ.
- Gómez, R. G. D., & Shen, M. A. (2008). Influence of probiotics on the growth and digestive enzyme activity of white Pacific shrimp (*Litopenaeus vannamei*). *Journal of Ocean University of China*, 7(2), 215–218. <https://doi.org/10.1007/s11802-008-0215-x>
- González-Félix, M. L., Gatlin, D. M., Lawrence, A. L., & Perez-Velazquez, M. (2002). Effect of dietary phospholipid on essential fatty acid requirements and tissue lipid composition of *Litopenaeus vannamei* juveniles. *Aquaculture*, 207(1–2), 151–167. [https://doi.org/10.1016/S0044-8486\(01\)00797-9](https://doi.org/10.1016/S0044-8486(01)00797-9)
- Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, 36(10), 3420–3435. <https://doi.org/10.1093/nar/gkn176>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644–652. <https://doi.org/10.1038/nbt.1883>
- Guo, X.-J., & Feng, J.-N. (2018). Comparisons of expression levels of heat shock proteins (hsp70 and hsp90) from *Anaphothrips obscurus* (Thysanoptera: Thripidae) in polymorphic adults exposed to different heat shock treatments. *Journal of Insect Science* (Online), 18(3), 15. <https://doi.org/10.1093/jisesa/iey059>
- Guppy, J. L., Jones, D. B., Jerry, D. R., Wade, N. M., Raadsma, H. W., Huerlimann, R., & Zenger, K. R. (2018). The state of "Omics" research for farmed penaeids: advances in research and impediments to industry utilization. *Frontiers in Genetics*, 9, 282. <https://doi.org/10.3389/fgene.2018.00282>
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B. O., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., ... Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), 1494–1512. <https://doi.org/10.1038/nprot.2013.084>
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., Owen White, C., Buell, R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. *Genome Biology*, 9(1), R7. <https://doi.org/10.1186/gb-2008-9-1-r7>
- Herran, B., Bertaux, J., & Grève, P. (2018). Divergent evolution and clade-specific duplications of the Insulin-like Receptor in malacostracan crustaceans. *General and Comparative Endocrinology*, 268, 34–39. <https://doi.org/10.1016/j.ygcen.2018.07.013>
- Hoff, K. J., & Stanke, M. (2019). Predicting genes in single genomes with AUGUSTUS. *Current Protocols in Bioinformatics*, 65(1), e57. <https://doi.org/10.1002/cpbi.57>
- Houston, R. D., Bean, T. P., Macqueen, D. J., Gundappa, M. K., Jin, Y. H., Jenkins, T. L., Sarah Louise, C., Selly, S. A. M., Martin, J. R., Stevens, E. M., Santos, A. D., & Robledo, D. (2020). Harnessing genomics to fast-track genetic improvement in aquaculture. *Nature Reviews Genetics*, 21(7), 389–409. <https://doi.org/10.1038/s41576-020-0227-y>
- Hu, X., Yang, H.-L., Yan, Y.-Y., Zhang, C.-X., Ye, J.-D., Lu, K.-L., Hu, L.-H., Zhang, J.-J., Ruan, L., & Yun-Zhang, S. (2019). Effects of fructooligosaccharide on growth, immunity and intestinal microbiota of shrimp (*Litopenaeus vannamei*) fed diets with fish meal partially replaced by soybean meal. *Aquaculture Nutrition*, 25(1), 194–204. <https://doi.org/10.1111/anu.12843>
- Huang, S.-W., Lin, Y.-Y., You, E.-M., Liu, T.-T., Shu, H.-Y., Keh-Ming, W. U., Tsai, S.-F., Lo, C.-F., Kou, G.-H., Ma, G.-C., Chen, M., Dongying, W. U., Aoki, T., Hirono, I., & Hon-Tsen, Y. U. (2011). Fosmid library end sequencing reveals a rarely known genome structure of marine shrimp *Penaeus monodon*. *BMC Genomics*, 12(1), 242. <https://doi.org/10.1186/1471-2164-12-242>
- Huerlimann, R., Wade, N. M., Gordon, L., Montenegro, J. D., Goodall, J., McWilliam, S., Tinning, M., Siemering, K., Giardina, E., Donovan, D., Sellars, M. J., Cowley, J. A., Condon, K., Coman, G. J., Khatkar, M. S., Raadsma, H. W., Maes, G. E., Zenger, K. R., & Jerry, D. R. (2018). De novo assembly, characterization, functional annotation and expression patterns of the black tiger shrimp (*Penaeus monodon*) transcriptome. *Scientific Reports*, 8(1), 13553. <https://doi.org/10.1038/s41598-018-31148-4>
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1), D309–D314. <https://doi.org/10.1093/nar/gky1085>
- Huete-Pérez, J. A., & Quezada, F. (2013). Genomic approaches in marine biodiversity and aquaculture. *Biological Research*, 46(4), 353–361. <https://doi.org/10.4067/s0716-97602013000400007>
- Jackson, C. J., & 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. <https://doi.org/10.1111/j.1365-2109.1998.tb01358.x>
- Jenjaroenpun, P., Wongsurawat, T., Wadley, T. D., Wassenaar, T. M., Liu, J., Dai, Q., Wanchai, V., Akel, N. S., Jamshidi-Parsian, A., Franco, A. T., Boysen, G., Jennings, M. L., Ussery, D. W., He, C., & Nookaew, I. (2020). Decoding the epitranscriptional landscape from native RNA sequences. *Nucleic Acids Research*, 49(2), e7. <https://doi.org/10.1093/nar/gkaa620>
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)*, 30(9), 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
- Jung, H., Lyons, R. E., Hurwood, D. A., & Mather, P. B. (2013). Genes and growth performance in crustacean species: a review of relevant genomic studies in crustaceans and other taxa. *Reviews in Aquaculture*, 5(2), 77–110. <https://doi.org/10.1111/raq.12005>
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1), D457–D462. <https://doi.org/10.1093/nar/gkv1070>
- Kao, D., Lai, A. G., Stamatakis, E., Rosic, S., Konstantinides, N., Jarvis, E., Di Donfrancesco, A., Pouchkina-Stancheva, N., Sémon, M., Grillo, M., Bruce, H., Kumar, S., Siwanowicz, I., Le, A., Lemire, A., Eisen, M. B., Extavour, C., Browne, W. E., Wolff, C., ... Aboobaker, A. (2016). The genome of the crustacean *Parhyale hawaiiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. *eLife*, 5, e20062. <https://doi.org/10.7554/eLife.20062>

- Karoonuthaisiri, N., Sittikankeaw, K., Preechaphol, R., Kalachikov, S., Wongsurawat, T., Uawisetwathana, U., Russo, J. J., Jingyue, J. U., Klinbunga, S., & Kirtikara, K. (2009). ReproArray(GTS): a cDNA microarray for identification of reproduction-related genes in the giant tiger shrimp *Penaeus monodon* and characterization of a novel nuclear autoantigenic sperm protein (NASP) gene. *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*, 4(2), 90–99. <https://doi.org/10.1016/j.cbd.2008.11.003>
- Katju, V., & Bergthorsson, U. (2013). Copy-number changes in evolution: rates, fitness effects and adaptive significance. *Frontiers in Genetics*, 4, 273. <https://doi.org/10.3389/fgene.2013.00273>
- Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Research*, 12(4), 656–664. <https://doi.org/10.1101/gr.229202>
- Kiepe, D., Ciarmatori, S., Hoeflich, A., Wolf, E., & Tönshoff, B. (2005). Insulin-like Growth Factor (IGF)-I stimulates cell proliferation and induces IGF binding protein (IGFBP)-3 and IGFBP-5 gene expression in cultured growth plate chondrocytes via distinct signaling pathways. *Endocrinology*, 146(7), 3096–3104. <https://doi.org/10.1210/en.2005-0324>
- Koehorst, J. J., van Dam, J. C. J., Saccenti, E., Martins Dos Santos, V. A. P., Suarez-Diez, M., & Schaap, P. J. (2018). SAPP: functional genome annotation and analysis through a semantic framework using FAIR principles. *Bioinformatics (Oxford, England)*, 34(8), 1401–1403. <https://doi.org/10.1093/bioinformatics/btx767>
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Research*, 19(9), 1639–1645. <https://doi.org/10.1101/gr.092759.109>
- 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. <https://doi.org/10.1093/molbev/msy096>
- Kung, S. Y., Chan, S. M., Hui, J. H., Tsang, W. S., Mak, A., & He, J. G. (2004). Vitellogenesis in the sand shrimp, *metapenaeus ensis*: the contribution from the hepatopancreas-specific vitellogenin gene (MeVg2). *Biology of Reproduction*, 71(3), 863–870. <https://doi.org/10.1095/biolreprod.103.022905>
- Lee, B.-Y., Choi, B.-S., Kim, M.-S., Park, J. C., Jeong, C.-B., Han, J., & Lee, J.-S. (2019). The genome of the freshwater water flea *Daphnia magna*: A potential use for freshwater molecular ecotoxicology. *Aquatic Toxicology*, 210, 69–84. <https://doi.org/10.1016/j.aquatox.2019.02.009>
- Leelatanawit, R., Uawisetwathana, U., Klinbunga, S., & Karoonuthaisiri, N. (2011). A cDNA microarray, UniShrimpChip, for identification of genes relevant to testicular development in the black tiger shrimp (*Penaeus monodon*). *BMC Molecular Biology*, 12, 15. <https://doi.org/10.1186/1471-2199-12-15>
- Legg, D. A., Sutton, M. D., & Edgecombe, G. D. (2013). Arthropod fossil data increase congruence of morphological and molecular phylogenies. *Nature Communications*, 4(1), 2485. <https://doi.org/10.1038/ncomms3485>
- Lehnert, S. A., Wilson, K. J., Byrne, K., & Moore, S. S. (1999). Tissue-specific expressed sequence tags from the black tiger shrimp *Penaeus monodon*. *Marine Biotechnology*, 1(5), 465–476.
- Lex, A., Gehlenborg, N., Strobel, H., Vuilleumot, R., & Pfister, H. (2014). UpSet: visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1983–1992. <https://doi.org/10.1109/TVCG.2014.2346248>
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)*, 34(18), 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)*, 22(13), 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bradley Bernstein, M. A., Bender, M. G., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, 326(5950), 289–293. <https://doi.org/10.1126/science.1181369>
- Lin, C. P., & Danforth, B. N. (2004). How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. *Molecular Phylogenetics and Evolution*, 30(3), 686–702. [https://doi.org/10.1016/s1055-7903\(03\)00241-0](https://doi.org/10.1016/s1055-7903(03)00241-0)
- Lin, H., Zhijuan, C., Mingyu, X., Shengguo, L., & Lu, W. (2004). Hox genes and study of Hox genes in crustacean. *Chinese Journal of Oceanology and Limnology*, 22(4), 392–398. <https://doi.org/10.1007/BF02843634>
- Liong Ong, B., & Johnston, D. (2006). Influence of feeding on hepatopancreas structure and digestive enzyme activities in *Penaeus monodon*. *Journal of Shellfish Research*, 25(1), 113–121.
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Mansilla, M. C., Banchio, C. E., & de Mendoza, D. (2008). Signalling pathways controlling fatty acid desaturation. In P. J. Quinn, & X. Wang (Eds.), *Lipids in health and disease* (pp. 71–99). Springer. [https://doi.org/10.1007/978-1-4020-8831-5\\_3](https://doi.org/10.1007/978-1-4020-8831-5_3)
- Mashek, D. G., Li, L. O., & Coleman, R. A. (2007). Long-chain acyl-CoA synthetases and fatty acid channeling. *Future Lipidology*, 2(4), 465–476. <https://doi.org/10.2217/17460875.2.4.465>
- Mitri, C., Bischoff, E., Eglmeier, K., Holm, I., Dieme, C., Brito-Fravallo, E., Raz, A., Zakeri, S., Nejad, M. I. K., Djadid, N. D., Vernick, K. D., & Riehle, M. M. (2020). Gene copy number and function of the APL1 immune factor changed during *Anopheles* evolution. *Parasites & Vectors*, 13(1), 18. <https://doi.org/10.1186/s13071-019-3868-y>
- Olmos, J., Ochoa, L., Paniagua-Michel, J., & Contreras, R. (2011). Functional feed assessment on *Litopenaeus vannamei* using 100% fish meal replacement by soybean meal, high levels of complex carbohydrates and *Bacillus* probiotic strains. *Marine Drugs*, 9(6), 1119–1132. <https://doi.org/10.3390/md9061119>
- Palmer, W. J., & Jiggins, F. M. (2015). Comparative genomics reveals the origins and diversity of arthropod immune systems. *Molecular Biology and Evolution*, 32(8), 2111–2129. <https://doi.org/10.1093/molbev/msv093>
- Pavlopoulos, A., Kontarakis, Z., Liubicich, D. M., Serano, J. M., Akam, M., Patel, N. H., & Averof, M. (2009). Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean. *Proceedings of the National Academy of Sciences*, 106(33), 13897. <https://doi.org/10.1073/pnas.0902804106>
- Podsiadlowski, L., & Bartolomaeus, T. (2005). Organization of the mitochondrial genome of mantis shrimp *Pseudosquilla ciliata* (Crustacea: Stomatopoda). *Marine Biotechnology*, 7(6), 618–624. <https://doi.org/10.1007/s10126-005-0017-8>
- Pootakham, W., Uengwetwanit, T., Sonthirod, C., Sittikankeaw, 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. <https://doi.org/10.3389/fmars.2020.00172>
- Porobic, J., Fulton, E. A., Parada, C., Frusher, S., Ernst, B., & Manríquez, P. (2019). The impact of fishing on a highly vulnerable ecosystem, the case of Juan Fernández Ridge ecosystem. *PLoS One*, 14(2), e0212485. <https://doi.org/10.1371/journal.pone.0212485>
- Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., Troll, C. J., Fields, A., Hartley, P. D., Sugnet, C. W., Haussler, D., Rokhsar, D. S., & Green, R. E. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research*, 26(3), 342–350. <https://doi.org/10.1101/gr.193474.115>

- Raja, K., Rahman, M. M., Rajkumar, M., Gopalakrishnan, A., & Vijayakumar, R. (2015). Effect of ingestion and waterborne routes under different shrimp densities on white spot syndrome virus susceptibility in three commercially important penaeid shrimps. *Aquaculture Reports*, 2, 120–125. <https://doi.org/10.1016/j.aqrep.2015.09.005>.
- Rauch, J., Volinsky, N., Romano, D., & Kolch, W. (2011). The secret life of kinases: functions beyond catalysis. *Cell Communication and Signaling*, 9(1), 23. <https://doi.org/10.1186/1478-811X-9-23>
- Reference Standard for Genome Biology (2018). A reference standard for genome biology. *Nature Biotechnology*, 36(12), 1121. <https://doi.org/10.1038/nbt.4318>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rosolen, L. A. M., Vicari, M. R., & Almeida, M. C. (2018). Accumulation of transposable elements in autosomes and giant sex chromosomes of *Omophoita* (Chrysomelidae: Alticinae). *Cytogenetic and Genome Research*, 156(4), 215–222. <https://doi.org/10.1159/000495199>
- Rungrasamee, W., Klanchui, A., Maibunkaew, S., & Karoonuthaisiri, N. (2016). Bacterial dynamics in intestines of the black tiger shrimp and the Pacific white shrimp during *Vibrio harveyi* exposure. *Journal of Invertebrate Pathology*, 133, 12–19. <https://doi.org/10.1016/j.jip.2015.11.004>
- Santos, C. A., Andrade, S. C. S., Teixeira, A. K., Farias, F., Kurkjian, K., Guerrelhas, A. C., Rocha, J. L., Galetti Jr, P. M., & Freitas, P. D. (2018). *Litopenaeus vannamei* transcriptome profile of populations evaluated for growth performance and exposed to white spot syndrome virus (WSSV). *Frontiers in Genetics*, 9, 120. <https://doi.org/10.3389/fgene.2018.00120>
- Schmitt, S., Ugrankar, R., Greene, S. E., Prajapati, M., & Lehmann, M. (2015). *Drosophila*: Lipin interacts with insulin and TOR signaling pathways in the control of growth and lipid metabolism. *Journal of Cell Science*, 128(23), 4395. <https://doi.org/10.1242/jcs.173740>
- Schumpert, C., Handy, I., Dudycha, J. L., & Patel, R. C. (2014). Relationship between heat shock protein 70 expression and life span in *Daphnia*. *Mechanisms of Ageing and Development*, 139, 1–10. <https://doi.org/10.1016/j.mad.2014.04.001>
- Schwentner, M., Combosch, D. J., Pakes Nelson, J., & Giribet, G. (2017). A Phylogenomic solution to the origin of insects by resolving crustacean-hexapod relationships. *Current Biology*, 27(12), 1818–1824.e1815. <https://doi.org/10.1016/j.cub.2017.05.040>
- Shi, Y., & He, M.-X. (2016). PflRR interacts with HrlGF-I and activates the MAP-kinase and PI3-kinase signaling pathways to regulate glycometabolism in *Pinctada fucata*. *Scientific Reports*, 6(1), 22063. <https://doi.org/10.1038/srep22063>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics (Oxford, England)*, 31, 3210–3212.
- Sittikankaew, K., Pootakham, W., Sonthirod, C., Sangsrakru, D., Yoocha, T., Khudet, J., Nookeaw, I., Uawisetwathana, U., Rungrasamee, W., & Karoonuthaisiri, N. (2020). Transcriptome analyses reveal the synergistic effects of feeding and eyestalk ablation on ovarian maturation in black tiger shrimp. *Scientific Reports*, 10(1), 3239. <https://doi.org/10.1038/s41598-020-60192-2>
- Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinform*, 6(1), 31. <https://doi.org/10.1186/1471-2105-6-31>
- Smit, A., Hubley, R., & Green, P. (2013–2015). RepeatMasker Open-4.0. Available from <http://www.repeatmasker.org>
- Snipen, L., & Ussery, D. W. (2010). Standard operating procedure for computing pangenome trees. *Standards in Genomic Sciences*, 2(1), 135–141. <https://doi.org/10.4056/sigs.38923>
- Spinks, P. Q., & Shaffer, H. B. (2009). Conflicting mitochondrial and nuclear phylogenies for the widely disjunct Emys (Testudines: Emydidae) species complex, and what they tell us about biogeography and hybridization. *Systematic Biology*, 58(1), 1–20. <https://doi.org/10.1093/sysbio/syp005>
- Springer, M. S., DeBry, R. W., Douady, C., Amrine, H. M., Madsen, O., de Jong, W. W., & Stanhope, M. J. (2001). Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Molecular Biology and Evolution*, 18(2), 132–143. <https://doi.org/10.1093/oxfordjournals.molbev.a003787>
- Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics (Oxford, England)*, 24(5), 637–644. <https://doi.org/10.1093/bioinformatics/btn013>
- Steinberger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11), 1026–1028. <https://doi.org/10.1038/nbt.3988>
- Supungul, P., Klinbunga, S., Pichyangkura, R., Jitrapakdee, S., Hirono, I., Aoki, T., & Tassanakajon, A. (2002). Identification of immune-related genes in hemocytes of black tiger shrimp (*Penaeus monodon*). *Marine Biotechnology*, 4(5), 487–494. <https://doi.org/10.1007/s10126-002-0043-8>
- Swathi, A., Shekhar, M. S., Katneni, V. K., & Vijayan, K. K. (2018). Genome size estimation of brackishwater fishes and penaeid shrimps by flow cytometry. *Molecular Biology Reports*, 45(5), 951–960. <https://doi.org/10.1007/s11033-018-4243-3>
- Talla, V., Suh, A., Kalsoom, F., Dinca, V., Vila, R., Friberg, M., Wiklund, C., & Backström, N. (2017). Rapid increase in genome size as a consequence of transposable element hyperactivity in wood-white (Leptidea) butterflies. *Genome Biology and Evolution*, 9(10), 2491–2505. <https://doi.org/10.1093/gbe/evx163>
- Tamone, S. L., & Harrison, J. F. (2015). Linking insects with crustacea: physiology of the pancrustacea: an introduction to the symposium. *Integrative and Comparative Biology*, 55(5), 765–770. <https://doi.org/10.1093/icb/icv093>
- Tassanakajon, A., Klinbunga, S., Paunglarp, N., Rimphanitchayakit, V., Udomkit, A., Jitrapakdee, S., Sritunyalucksana, K., Phongdara, A., Pongsomboon, S., Supungul, P., Tang, S., Kuphanumart, K., Pichyangkura, R., & Lursinsap, C. (2006). *Penaeus monodon* gene discovery project: the generation of an EST collection and establishment of a database. *Gene*, 384, 104–112. <https://doi.org/10.1016/j.gene.2006.07.012>
- Tassanakajon, A., Pongsomboon, S., Jarayabhand, P., Klinbunga, S., & Boonsaeng, V. V. (1998). Genetic structure in wild populations of black tiger shrimp (*Penaeus monodon*) using randomly amplified polymorphic DNA analysis. *Journal of Marine Biotechnology*, 6(4), 249–254.(9852621).
- Tassanakajon, A., Pongsomboon, S., Rimphanitchayakit, V., Jarayabhand, P., & Boonsaeng, V. (1997). Random amplified polymorphic DNA (RAPD) markers for determination of genetic variation in wild populations of the black tiger prawn (*Penaeus monodon*) in Thailand. *Molecular Marine Biology and Biotechnology*, 6(2), 110–115. (9200837).
- Teasdale, L. C., Köhler, F., Murray, K. D., O'Hara, T., & Moussalli, A. (2016). Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon capture. *Molecular Ecology Resources*, 16(5), 1107–1123. <https://doi.org/10.1111/1755-0998.12552>
- Thornber, K., Verner-Jeffreys, D., Hinchliffe, S., Rahman, M. M., Bass, D., & Tyler, C. R. (2019). Evaluating antimicrobial resistance in the global shrimp industry. *Reviews in Aquaculture*, 12(2), 966–986. <https://doi.org/10.1111/raq.12367>
- Thorp, J. H., & Rogers, D. C. (2011). Chapter 13 - Introduction to Freshwater Invertebrates in the Phylum Arthropoda. In J. H. Thorp, & D. C. Rogers (Eds.), *Field guide to freshwater invertebrates of North America* (pp. 109–119). Academic Press. <https://doi.org/10.1016/B978-0-12-381426-5.00013-2>

- Timm, L., & Bracken-Grissom, H. D. (2015). The forest for the trees: evaluating molecular phylogenies with an emphasis on higher-level Decapoda. *Journal of Crustacean Biology*, 35(5), 577–592. <https://doi.org/10.1163/1937240X-00002371>
- Tong, J., Lehnert, S. A., Byrne, K., Kwan, H. S., & Chu, K. H. (2002). Development of polymorphic EST markers in *Penaeus monodon*: applications in penaeid genetics. *Aquaculture*, 208(1–2), 69–79. [https://doi.org/10.1016/S0044-8486\(01\)00706-2](https://doi.org/10.1016/S0044-8486(01)00706-2)
- Tørresen, O. K., Star, B., Mier, P., Andrade-Navarro, M. A., Bateman, A., Jarnot, P., Gruca, A., Grynberg, M., Kajava, A. V., Promponas, V. J., Anisimova, M., Jakobsen, K. S., & Linke, D. (2019). Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Research*, 47(21), 10994–11006. <https://doi.org/10.1093/nar/gkz841>
- Uengwetwanit, T., Ponza, P., Sangsakru, D., Wichadakul, D., Ingsriswang, S., Leelatanawit, R., Klinbunga, S., Tangphatsornruang, S., & 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. <https://doi.org/10.1016/j.margen.2017.08.007>
- Valenzuela-Castillo, A., Sánchez-Paz, A., Castro-Longoria, R., López-Torres, M. A., & Grijalva-Chon, J. M. (2019). Hsp70 function and polymorphism, its implications for mollusk aquaculture: a review. *Latin American Journal of Aquatic Research*, 47, 224–231.
- Quyen, D. V., Gan, H. M., Lee, Y. P., Nguyen, D. D., Nguyen, T. H., Tran, X. T., Nguyen, V. S., Khang, D. D., & Austin, C. M. (2020). Improved genomic resources for the black tiger prawn (*Penaeus monodon*). *Marine Genomics*, 52, 100751. <https://doi.org/10.1016/j.margen.2020.100751>
- Wahlberg, N., & Wheat, C. W. (2008). Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of *Lepidoptera*. *Systematic Biology*, 57(2), 231–242. <https://doi.org/10.1080/10635150802033006>
- Wang, W., Wu, X., Liu, Z., Zheng, H., & Cheng, Y. (2014). Insights into hepatopancreatic functions for nutrition metabolism and ovarian development in the crab *Portunus trituberculatus*: gene discovery in the comparative transcriptome of different hepatopancreas stages. *PLoS One*, 9(1), e84921. <https://doi.org/10.1371/journal.pone.0084921>
- Washburn, J. D., Schnable, J. C., Conant, G. C., Brutnell, T. P., Shao, Y., Zhang, Y., Ludwig, M., Gerrit Davidse, J., & Pires, C. (2017). Genome-guided phylo-transcriptomic methods and the nuclear phylogenetic tree of the Paniceae grasses. *Scientific Reports*, 7(1), 13528. <https://doi.org/10.1038/s41598-017-13236-z>
- Wenne, R., Boudry, P., Hemmer-Hansen, J., Lubieniecki, K. P., Was, A., & Kause, A. (2007). What role for genomics in fisheries management and aquaculture? *Aquatic Living Resources*, 20(3), 241–255. <https://doi.org/10.1051/alr:2007037>
- Wilson, K., Cahill, V., Ballment, E., & Benzie, J. (2000). The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: Are malacostracan crustaceans more closely related to Insects than to branchiopods? *Molecular Biology and Evolution*, 17(6), 863–874. <https://doi.org/10.1093/oxfordjournals.molbev.a026366>
- Wilson, K., Li, Y., Whan, V., Brutnell, T. P., Shao, Y., Zhang, Y., Ludwig, M., Gerrit Davidse, J., & Pires, C. (2002). Genetic mapping of the black tiger shrimp *Penaeus monodon* with amplified fragment length polymorphism. *Aquaculture*, 204(3–4), 297–309. [https://doi.org/10.1016/S0044-8486\(01\)00842-0](https://doi.org/10.1016/S0044-8486(01)00842-0)
- Wolff, G. H., Thoen, H. H., Marshall, J., Sayre, M. E., & Strausfeld, N. J. (2017). An insect-like mushroom body in a crustacean brain. *eLife*, 6, e29889. <https://doi.org/10.7554/eLife.29889>
- Wouters, R., Lavens, P., Nieto, J., & Sorgeloos, P. (2001). Penaeid shrimp broodstock nutrition: an updated review on research and development. *Aquaculture*, 202(1), 1–21. [https://doi.org/10.1016/S0044-8486\(01\)00570-1](https://doi.org/10.1016/S0044-8486(01)00570-1)
- Wu, T. D., & Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, 21(9), 1859–1875. <https://doi.org/10.1093/bioinformatics/bti310>
- Wuthisuthimethavee, S., Aoki, T., Hirono, I., & Tassanakajon, A. (2009). Preliminary study of BAC library construction in black tiger shrimp, *Penaeus monodon*. *Walailak Journal of Science and Technology*, 6(1), 37–48.
- Yu, Y., Zhang, X., Yuan, J., Li, F., Chen, X., Zhao, Y., Huang, L., Zheng, H., & Xiang, J. (2015). Genome survey and high-density genetic map construction provide genomic and genetic resources for the Pacific white shrimp *Litopenaeus vannamei*. *Scientific Reports*, 5(1), 15612. <https://doi.org/10.1038/srep15612>
- Yuan, J., Gao, Y., Zhang, X., Wei, J., Liu, C., Li, F., & Xiang, J. (2017). Genome sequences of marine shrimp *Exopalaemon carinicauda* Holthuis provide insights into genome size evolution of Caridea. *Marine Drugs*, 15(7), 213. <https://doi.org/10.3390/md15070213>
- Yuan, J., Zhang, X., Liu, C., Yu, Y., Wei, J., Li, F., & Xiang, J. (2018). Genomic resources and comparative analyses of two economical penaeid shrimp species, *Marsupenaeus japonicus* and *Penaeus monodon*. *Marine Genomics*, 39, 22–25. <https://doi.org/10.1016/j.margen.2017.12.006>
- 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. <https://doi.org/10.3389/fgene.2018.00693>
- Zhang, X., Yuan, J., Sun, Y., Li, S., Gao, Y. I., Yang, Y. U., Liu, C., Wang, Q., Lv, X., Zhang, X., Ma, K. Y., Wang, X., Lin, W., Wang, L., Zhu, X., Zhang, C., Zhang, J., Jin, S., Kuijie, Y. U., ... Xiang, J. (2019). Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nature Communications*, 10(1), 356. <https://doi.org/10.1038/s41467-018-08197-4>
- Zhao, C., Zhang, X., Liu, C., Huan, P., Li, F., Xiang, J., & Huang, C. (2012). BAC end sequencing of Pacific white shrimp *Litopenaeus vannamei*: a glimpse into the genome of Penaeid shrimp. *Chinese Journal of Oceanology and Limnology*, 30(3), 456–470. <https://doi.org/10.1007/s00343-012-1159-y>
- Zhou, X., Cheng, S., Zhao, C., Lin, C., Wang, D., & Zhao, Y. (2020). Expression and functional analysis of the Akt gene from *Daphnia pulex*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 248–249, 110462. <https://doi.org/10.1016/j.cbpb.2020.110462>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Uengwetwanit T, Pootakham W, Nookaew I, et al. A chromosome-level assembly of the black tiger shrimp (*Penaeus monodon*) genome facilitates the identification of growth-associated genes. *Mol Ecol Resour*. 2021;00:1–21. <https://doi.org/10.1111/1755-0998.13357>