



The State of “Omics” Research for Farmed Penaeids: Advances in Research and Impediments to Industry Utilization

Jarrold L. Guppy^{1,2*}, David B. Jones^{1,2}, Dean R. Jerry^{1,2}, Nicholas M. Wade^{1,3}, Herman W. Raadsma^{1,4}, Roger Huerlimann^{1,2} and Kyall R. Zenger^{1,2}

¹ Australian Research Council Industrial Transformation Research Hub for Advanced Prawn Breeding, James Cook University, Townsville, QLD, Australia, ² College of Science and Engineering and Centre for Sustainable Tropical Fisheries and Aquaculture, James Cook University, Townsville, QLD, Australia, ³ Aquaculture Program, CSIRO Agriculture & Food, Queensland Bioscience Precinct, St Lucia, QLD, Australia, ⁴ Faculty of Science, Sydney School of Veterinary Science, The University of Sydney, Camden, NSW, Australia

OPEN ACCESS

Edited by:

Ross Houston,
University of Edinburgh,
United Kingdom

Reviewed by:

George E. Liu,
Agricultural Research Service (USDA),
United States
Jesús Fernández,
Instituto Nacional de Investigación y
Tecnología Agraria y Alimentaria (INIA),
Spain

*Correspondence:

Jarrold L. Guppy
jarrod.guppy@my.jcu.edu.au

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 14 May 2018

Accepted: 09 July 2018

Published: 03 August 2018

Citation:

Guppy JL, Jones DB, Jerry DR,
Wade NM, Raadsma HW,
Huerlimann R and Zenger KR (2018)
The State of “Omics” Research for
Farmed Penaeids: Advances in
Research and Impediments to
Industry Utilization.
Front. Genet. 9:282.
doi: 10.3389/fgene.2018.00282

Elucidating the underlying genetic drivers of production traits in agricultural and aquaculture species is critical to efforts to maximize farming efficiency. “Omics” based methods (i.e., transcriptomics, genomics, proteomics, and metabolomics) are increasingly being applied to gain unprecedented insight into the biology of many aquaculture species. While the culture of penaeid shrimp has increased markedly, the industry continues to be impeded in many regards by disease, reproductive dysfunction, and a poor understanding of production traits. Extensive effort has been, and continues to be, applied to develop critical genomic resources for many commercially important penaeids. However, the industry application of these genomic resources, and the translation of the knowledge derived from “omics” studies has not yet been completely realized. Integration between the multiple “omics” resources now available (i.e., genome assemblies, transcriptomes, linkage maps, optical maps, and proteomes) will prove critical to unlocking the full utility of these otherwise independently developed and isolated resources. Furthermore, emerging “omics” based techniques are now available to address longstanding issues with completing keystone genome assemblies (e.g., through long-read sequencing), and can provide cost-effective industrial scale genotyping tools (e.g., through low density SNP chips and genotype-by-sequencing) to undertake advanced selective breeding programs (i.e., genomic selection) and powerful genome-wide association studies. In particular, this review highlights the status, utility and suggested path forward for continued development, and improved use of “omics” resources in penaeid aquaculture.

Keywords: penaeid shrimps, omics, advanced breeding, aquaculture, genome assembly, linkage mapping, functional genomics, molecular markers

BACKGROUND

The application of “omics” based technologies has provided unprecedented insight into the genetic and functional biology of livestock and crop species, and when integrated within selective breeding programs, “omics” techniques have facilitated significant and continued improvements in productivity (Dekkers, 2012; Pérez-de-Castro et al., 2012). In particular, “omics” approaches, including genomics, transcriptomics, proteomics, and metabolomics, have been applied widely to elucidate the molecular basis of performance traits (e.g., growth) and overcome poorly understood biological impediments that prevent efficient production (e.g., disease, reproductive failure, and undesired carcass composition) (Rothschild and Plastow, 2008; Taylor et al., 2016). Harnessing genomic information in Holstein Friesians (*Bos taurus*), for example, has led to a doubling of the annual rate of improvement in milk production (García-Ruiz et al., 2016; Taylor et al., 2016), while in layer (egg producing) chickens (*Gallus gallus domesticus*), 16 production traits (i.e., egg size, egg color, and age of sexual maturity) have all seen significant improvements through the application of genomic breeding methodologies (Wolc et al., 2015; Meuwissen et al., 2016). Numerous genetic tests, built upon genomic and transcriptomic research, are now also available for use for commercial applications (e.g., average daily gain, intramuscular fat marbling, meat tenderness, congenital defect screening) in livestock species (Dekkers, 2004). Through these “omics” based tests it has been possible for many industries to vastly improve the identification and selection of superior individuals for breeding programs (Dekkers, 2012; Pérez-de-Castro et al., 2012; Taylor et al., 2016).

The transformational effects that “omics” research has had on the production of plant and livestock industries is widely acknowledged (see reviews Agrawl and Narayan, 2015; Van Emon, 2015; Taylor et al., 2016). As a result, similar efforts to utilize “omics” techniques to improve farmed aquatic species is anticipated to achieve significant improvements to aquaculture production as well (Gjedrem and Rye, 2016). While most terrestrial food production species have already undergone decades of traditional selection, most aquaculture species have only recently been domesticated, with few having been improved through targeted selection programs. In stark contrast to livestock industries, less than 10% of aquaculture production is derived from improved lines (Gjedrem et al., 2012). While the relatively young nature of aquaculture and the scale of the industry has constrained “omics” resource development, this also means the latent natural variation of many species and the genetic potential for increased productivity remains unharnessed (Mackay et al., 2009).

Of the aquatic taxa cultured for food worldwide, crustaceans are the fourth most produced aquaculture commodity by weight per year (6.9 million tons), behind finfish (49.8 million tons), marine plants (27.3 million tons), and bivalves (19 million tons) (FAO, 2016). Yet by economic value, crustaceans are the second highest farmed seafood commodity with an annual production value exceeding US\$ 35 billion (i.e., ~\$5,072 per ton), only surpassed by finfish with an annual economic value of US\$ 60 billion (i.e., ~\$1,205 per ton) (FAO, 2016). While crustaceans

are diverse in species number, penaeid marine shrimp (Family; *Penaeidae*) are the predominately cultured group, with an annual production exceeding 4.8 million tons (FAO, 2017). Of the species farmed, Pacific white-leg shrimp (*Litopenaeus vannamei*) and black tiger shrimp (*Penaeus monodon*) are the primary species cultured, accounting for 80% (~3.9 million tons) and 15% (~0.7 million tons) of total penaeid shrimp production, respectively. The remaining 5% (~0.25 million tons) of production are cumulatively provided by kuruma shrimp (*Marsupenaeus japonicus*), Chinese shrimp (*Fenneropenaeus chinensis*), banana shrimp (*Fenneropenaeus merguensis*), Indian shrimp (*Fenneropenaeus indicus*), and blue shrimp (*Litopenaeus stylirostris*).

The global penaeid aquaculture industry has demonstrated remarkable growth, increasing in production from a meager 71 tons in the early 1980's, to 4.8 million tons in 2015 (FAO, 2017). Despite the industry's impressive growth, it has faced substantial challenges that still often constrain production. Decades of unsustainable reliance on wild sourced broodstock (for *P. monodon* in particular) (Benzie, 2009), a poor understanding of the biological and genetic basis of production traits (Dunham, 2011), a limited ability to maintain pedigree traceability and a number of devastating global disease outbreaks (Benzie, 2009; Lio-Po and Leñaño, 2016), have all impeded industry development. Nevertheless, penaeid shrimp aquaculture is perfectly placed to take advantage of the rapidly emerging suite of “omics” techniques that now provide the capacity to explore and address many of the key impediments faced by the industry (Liu, 2007; Gjedrem and Rye, 2016). As such, in a similar manner seen for agriculture, the field of penaeid “omics” is now rapidly evolving, fuelled by industry demand for high quality stock, decreased disease risk and improved productivity (Debnath et al., 2016). Furthermore, improvements in breeding techniques and a swell in industry critical mass through increasing centralization of breeding programs (Benzie, 2009) have coalesced to drive the development of critical “omics” resources for each of the commercially important penaeid shrimp species (Abdelrahman et al., 2017).

This review focuses on the “omics” tools, in particular genomic, transcriptomic, and proteomic resources, that have been developed for penaeids. Specifically, this includes a review of the state of genome sequencing and mapping (linkage, physical, and optical mapping), genomic marker development, functional genomics (transcriptomics and proteomics), and trait mapping studies for penaeid shrimp. This review also provides rationale for the implementation of emerging research methods (e.g., long-read sequencing, high-throughput optical mapping and genotype-by-sequencing), along with suggested approaches to improve the accessibility and application of “omics” tools within the global penaeid aquaculture industry.

SEQUENCING, ASSEMBLING, AND MAPPING COMPLEX GENOMES

Whole-Genome Sequencing

Comprehensive and complete genomic assemblies are often regarded as the keystone “omics” resource for a species due

to the significant nucleotide position information they convey, and the ability to leverage genomic data throughout a large variety of further applications. For example, genome assemblies provide a reference base for functional transcriptomic studies, can aid in the positioning of genetic markers, and allow for the examination and characterization of genomic regions of commercial or biological interest. While genomic assemblies are available for a number of aquaculture species [e.g., *Oncorhynchus mykiss* (Berthelot et al., 2014); *Oreochromis niloticus* (Conte et al., 2017); *Lates calcarifer* (Vij et al., 2016); *Ictalurus punctatus* (Liu et al., 2016); *Salmo salar* (Lien et al., 2016)], the comparatively large size (~1.8 Gbp *L. vannamei*, ~2.2 Gbp *P. monodon*) and the highly repetitive nature of penaeid genomes have presented a significant challenge to their assembly (Huang et al., 2011; Baranski et al., 2014; Yu et al., 2015). Likewise, the assembly of penaeid genomes has been further impeded by their large chromosome number ($n = 44$; **Table 1**) and higher levels of genomic heterozygosity (Abdelrahman et al., 2017) than genomes assemblies derived from inbred domesticated lines of livestock species such as *G. gallus* (International Chicken Genome Sequencing, 2004) or model organisms [e.g., *Mus musculus* (Mouse Genome Sequencing, 2002)]. Despite a number of long standing efforts by numerous research groups (Abdelrahman et al., 2017), no comprehensive genome assembly is available for a penaeid shrimp.

While genome sequencing efforts have seen dramatic improvement through the development of high-throughput sequencing, the task of resolving, and assembling the many repetitive regions within the penaeid genome (~80%; Abdelrahman et al., 2017) remains particularly challenging. Short read second-generation sequencing methods (i.e., Illumina HiSeq) involve the fragmentation of the genome into millions of 50–300 bp fragments, through which most genome positional based information is lost prior to sequencing. For non-complex genomes, these short sequences can be reconstructed iteratively into near complete chromosomal sections (contigs or scaffolds) by identifying and overlaying sequences that partially overlap each other. However, when short read sequencing methods are applied to highly repetitive regions within the genome, 50–300 bp fragments are of insufficient length to span the repeat regions, resulting in a break down in the iterative tiling approach. This difficulty in building contiguous sequences is exacerbated by diploid genomes with high levels of heterozygosity [e.g., *Ciona savignyi* (Small et al., 2007)] as each multiple assembly paths (each representing one chromosomal copy) cannot be resolved and are either reported as two unique sequences or erroneously joined adjacently to each other (Henson et al., 2012; Prysycz and Gabaldón, 2016). This effect is evident in the previous short-read assembly by Yu et al. (2015) for *L. vannamei*, in which a *de-novo* assembly of 80–100 bp fragments (at ~41x coverage) resulted in highly fragmented assembly with 4,336,336 scaffolds and maximum scaffold size of only 38,588 bp (**Table 1**; Yu et al., 2015). Likewise, early *P. monodon* assemblies of 250 bp fragments (~68x coverage) have resulted in 1,168,065–3,064,940 contigs (averaging 1,510 bp in length), and with a maximum contig size of 21,136–45,736 bp depending on the bioinformatics approach used for assembly (**Table 1**; Montenegro et al., 2018).

Recently, a number of genome assemblies, including those of the Peruvian scallop (*Argopecten purpuratus*; Li et al., 2018) and for the agricultural crop *Capsicum annuum* (bell-peppers; Hulse-Kemp et al., 2018) have included sequencing data from the novel 10X Chromium platform (10X Genomics) in an effort to decrease ambiguities during genome reconstruction. The preparation of 10X Chromium libraries involves individual barcoding of high-molecular weight DNA fragments (up to 1 million unique barcodes), prior to traditional short-read sequencing. Finally, the 10X Genomics assembler, Supernova, generates an initial global De Bruijn graph before utilizing the individual barcoding information to confirm that reads forming each contig are physically linked, and where possible, barcoding information is used to resolve uncertainties (branches, bubbles, and gaps) along each assembled contig (Weisenfeld et al., 2017). However, as Supernova in the first instance still relies upon an iterative sequence-read overlapping approach, difficulties still persistent in assembly of non-model species, or those with complex genomes. For example, early observations from the generation of a 10X Chromium assembly for *P. monodon* has seen Supernova repeatedly fail to reach completion (Montenegro et al., 2018), and suggests this approach may not be successful in resolving the extremely high repetitiveness of penaeid genomes. However, future improved versions of assembly algorithms may allow increase utility of this method for complex non-model species. Likewise, while the additional genome sequencing approach, chromosome conformation capture (Hi-C or ChIP-seq), has been used to improve genome assemblies for goat (*Capra aegarus hircus*) (Bickhart et al., 2017) and maize (*Zea mays*) (Jiao et al., 2017), but requires access to a draft genome which already has a relatively large average size (100kb; Oddes et al., 2018) for effective short read mapping and further improvement of genome scaffolding (Forcato et al., 2017).

As with many species, the traditional genome assembly methods that were initially employed for penaeids have required subsequent amendment to overcome the current difficulties of producing a comprehensive genome assembly. In particular, third-generation sequencing technologies (i.e., Pacific Biosciences Sequel, Nanopore MinION) which generate much longer sequencing reads (mean length over 10 kbp), are increasingly being utilized to improve the accuracy and completeness of genome assemblies in many plant and livestock species [e.g., pineapple, *Ananas comosus* (Ming et al., 2015); rice, *Oryza sativa* (Brozynska et al., 2016); and goat, *C. a. hircus* (Bickhart et al., 2017)]. Even more so, recent improvements in the sequencing chemistry, rate of throughput and accuracy of base calling have allowed the more cost effective generation of large volumes of long read data (e.g., 30–50x coverage, with an average ~10 kb read length; Phillippy, 2017). As such, aquaculture species are now looking toward long read sequencing technologies (LRSTs) to resolve the assembly of repetitive and complex genomes (i.e., Nile tilapia, barramundi, channel catfish, Atlantic salmon) (Liu et al., 2016; Vij et al., 2016; Conte et al., 2017). LRSTs have most notably helped to achieve a near chromosomal-scale assembly of the large Atlantic salmon genome, which contains complex remnants of three separate whole-genome duplication events (Lien et al., 2016).

TABLE 1 | Summary of the genome size and chromosome number of commercially important penaeids and other major aquaculture species, along with the status of current genome assemblies.

| Species | Global production in 2015 (tons) (FAO, 2017) | Haplloid Chromosome number | Genome size (Gb) | Sequencing used | Assembly size (Gb) | Number of scaffold (S) or Contigs (C) | Longest scaffold/Contig (Mb) | Scaffold N50/Contig N50 (Mb) | References |
|---------------------------------|--|----------------------------|------------------|---|--------------------|---------------------------------------|------------------------------|------------------------------|--|
| <i>Litopenaeus vannamei</i> | 3,879,786 | 44 | 1.78 | 828 GB Illumina PE* >70GB PacBio | 2.6 | 6007 S | Unknown | 0.66 | Yu et al., 2015; Xiang, 2016 |
| <i>Penaeus monodon</i> | 713,318 | 44 | 2.17 | >140 GB Illumina PE* | 1.5-3.7 | 1,168,065-3,064,940C | 0.0021-0.0045 | 0.0009-0.0019 | Baranski et al., 2014; Montenegro et al., 2018 |
| <i>Marsupenaeus japonicus</i> | 47,726 | 44 | 2.4 | 79.75 GB Illumina PE* | 1.36 | 3,244,404 S | 0.0015 | 0.0013 | Lu et al., 2017 |
| <i>Fenneropenaeus chinensis</i> | 44,812 | 44 | 2.1 | 530 GB Illumina PE* >70 GB PacBio | Unknown | Unknown | Unknown | 0.15 | Kong and Gao, 2005; Xiang, 2016 |
| <i>Oncorhynchus mykiss</i> | 2,381,576 | 30 | 1.9 - 2.18 | 45 GB Roche 454 128 GB Illumina PE+SE* | 1.9 | 2,018 S | 5.4 | 0.384 | Berthelot et al., 2014 |
| <i>Oreochromis niloticus</i> | 3,930,579 | 23 | 1.08 | >44 GB PacBio | 0.927 | 2,566 S | 37 | 3.11 | Conte et al., 2017 |
| <i>Lates calcarifer</i> | 76,498 | 24 | 0.587 | 53.6 GB Illumina PE >60 GB PacBio | 0.670 | 24 S | 30.7 | 25.85 | Vij et al., 2016 |
| <i>Ictalurus punctatus</i> | 411,790 | 29 | 1.0 | 88 GB Illumina PE 5.4 GB PacBio | 0.783 | 9,974 S | 22.6 | 7.73 | Liu et al., 2016 |
| <i>Salmo salar</i> | 761,766 | 29 | 2.97 | 13 GB Sanger 326 GB Illumina PE 267 GB Illumina MP• 57 GB PacBio | 3.06 | 9,447 S | >0.038 | 2.97 | Lien et al., 2016 |

N.B. No genome assembly information is currently available for *Fenneropenaeus merguensis*, *Fenneropenaeus indicus* or *Litopenaeus stylirostris*.

*PE, Paired End Reads; ^ SE, Single end; • MP, mate pair.

Recently published details of the ongoing effort to complete the *L. vannamei* have noted the contig N50 (minimum contig length to cover 50% of the genome) has been reported (Abdelrahman et al., 2017) to have increased from 547 to 660,000 bp and the total number of scaffolds decreased from 4,336,336 to 6,007 through the incorporation of Pac Bio sequencing reads (Yu et al., 2015; Abdelrahman et al., 2017). Long read sequencing approaches are currently also being employed in assembly of the *P. monodon* genome with the aim to achieve similar improvements in genome assembly quality as seen in *L. vannamei* (Montenegro et al., 2018). Details have been released on a current effort to assemble the *F. chinensis* genome (Xiang, 2016); however, beyond the scaffold N50 (154.2kb) and sequence data generation (530 GB short read, 70 GB long read), the final assembly quality metrics are currently unavailable to the public. Likewise as both *L. vannamei* and *P. monodon* genomes are yet to be published, the details of their final assembly quality (genome completeness, scaffold size, average genome coverage), given the technology now available, still remain to be seen. It is likely the most successful assemblies will rely upon a “hybrid assembly” approach, utilizing short-read sequencing to correct the high error rate base-calling seen in long read sequencing, prior to assembly of “corrected” long reads (Sedlazeck et al., 2018). In construction of a *P. monodon* genome, difficulties have been encountered with obtaining sufficient high molecular weight DNA from shrimp tissues while avoiding contamination of DNA with polysaccharides that can interfere with long-read sequencing devices (Montenegro et al., 2018), further development of shrimp specific high molecular weight DNA extraction protocols are required.

While genomic-survey sequencing has been conducted in Kuruma shrimp (*M. japonicus*) at low coverage (33.23x short read genome coverage; Lu et al., 2017), genome assemblies for the Kuruma shrimp, or other commercially important penaeid species such as the Indian shrimp (*F. indicus*) are still yet to be undertaken. Additional penaeid assemblies are likely to follow once the *P. monodon*, *L. vannamei*, or *F. chinensis* genomes are available to be used for reference-guided genome assembly. Release of fully annotated and integrated genome assemblies in publicly accessible formats (e.g., as genome browsers in NCBI) should be a primary focus of the penaeid “omics” research community.

Linkage Mapping of Genetic Markers

Linkage maps are highly versatile genomic resources which provide a wealth of genomic information and facilitate the examination of the underlying genetic architecture of commercially and biologically important traits. As such, the development of linkage maps for penaeids (Table 2) has recently seen significant research effort with a number of maps now available for *P. monodon* (Wilson et al., 2002; Wuthisuthimethavee et al., 2005; Staelens et al., 2008; Baranski et al., 2014), *L. vannamei* (Pérez et al., 2004; Alcivar-Warren et al., 2007; Zhang et al., 2007; Du et al., 2010; Andriantahina et al., 2013; Gonçalves et al., 2014; Yu et al., 2015; Jones et al., 2017a), *M. japonicus* (Li et al., 2003; Lu et al., 2016b), and *F. chinensis* (Li Z. et al., 2006; Sun et al., 2008; Tian et al.,

2008; Liu et al., 2010; Wang et al., 2012; Zhang et al., 2013). Linkage map construction involves the genetic analysis of family groups (both parents and progeny), to allow the identification of the recombination pattern of polymorphic markers, and in turn, the calculation of a relative genomic position for each marker. When markers are positioned accurately within discrete linkage groups (with each group indicative of a chromosome), the genomic map provides a robust method to validate and error correct draft genome assemblies (Fierst, 2015), and also provides critical tools for the study of trait and genome architecture (as discussed in section Dissecting and Exploiting the Genetic Variation Underlying Phenotypes).

Early versions of linkage maps for penaeid shrimps contained a limited number of genetic markers (Table 2) due to the inherent difficulties in isolating informative polymorphic markers (as discussed in section Development and Applications of Polymorphic Markers). More recently, however, the linkage maps available for penaeids have improved dramatically, as it has now become possible to generate and simultaneously map the position of thousands of polymorphic SNP markers (Baranski et al., 2014; Yu et al., 2015; Lu et al., 2016b; Jones et al., 2017a). Linkage maps are now available that include between 3,959 and 9,298 markers and cover all 44 chromosomes of the penaeid genome (Baranski et al., 2014; Yu et al., 2015; Lu et al., 2016b; Jones et al., 2017a). When compared to earlier maps (Wilson et al., 2002; Li et al., 2003; Pérez et al., 2004; Wuthisuthimethavee et al., 2005; Li Z. et al., 2006; Maneeruttanarunroj et al., 2006; Alcivar-Warren et al., 2007; Zhang et al., 2007; Staelens et al., 2008; Andriantahina et al., 2013; Gonçalves et al., 2014) that contained 27 to 451 markers, distributed across between 8 and 51 linkage groups, the increase in density of polymorphic markers recently achieved has dramatically increased the applicability of these resources (i.e., assisting genomic assembly, examining the genetic architecture of traits, and undertaking comparative mapping).

As such, the construction of linkage maps has provided a number of interesting insights into the genomic structure of penaeids. For example, Baranski et al. (2014) constructed maps for *P. monodon* in which the female-specific map was substantially shorter than the male-specific map (2,917 vs. 4,059 cM). Alternatively, in *L. vannamei*, Pérez et al. (2004) and Zhang et al. (2007) both obtained longer maps for females than males (4,134 vs. 3,221 cM, and 2,771 vs. 2,116 cM, respectively), indicating that there may be higher recombination in males. In the absence of obvious karyotype size and number differences between male and female shrimp (You et al., 2010), the observed differences in map lengths suggests that sex-biased recombination occurs in penaeids, but may occur in a species dependent manner (Baranski et al., 2014). Further detailed map construction is required to validate and understand this difference in recombination between sexes, along with identification of genomic recombination hotspots. Patterns of recombination, specifically in areas of the genome containing genes of commercial importance, may affect the accuracy of and impact derived from genome-informed selection programs (Meuwissen et al., 2001; Habier et al., 2007; Goddard, 2009). Construction of linkage disequilibrium unit (LDU) or haplotype

TABLE 2 | Linkage maps available for penaeid shrimp, organized by species and marker used.

| Species * | Marker type | Individuals used | Families used | Linkage groups covered (n/44) | | Markers mapped | | Total map length (cM) | | Average inter-marker distance (cM) | | Genome coverage | | Software used | Reference | | |
|---------------------|------------------|------------------|-----------------|-------------------------------|--------|----------------|-------|-----------------------|----------|------------------------------------|--------|-----------------|--------|---------------|-----------|---------------------------------|----------------------------------|
| | | | | Male | Female | Sex av | Male | Female | Sex av | Male | Female | Sex av | Male | | | Female | Sex av |
| <i>L. vannamei</i> | SNPs | 205● | 1 | 44 | 4,201 | 4,396 | 6,146 | 6,143.95 | 5,657.42 | 4,271.43 | 1.46 | 1.29 | 97.49% | 97.71% | 98.39% | HighMap | Yu et al., 2015 |
| <i>L. vannamei</i> | SNPs | 631△ | 49 ^b | 44 | NA | NA | 4,817 | 4,522.30 | 4,530.60 | 4,552.50 | NA | NA | NA | NA | 98.12% | Carthagene V1.3 | Jones et al., 2017a |
| <i>L. vannamei</i> | SNPs | 144△ | 3 | 45 | 413 | 418 | 418 | 2,130 | 2,071 | 2,262.3 | NA | NA | 37.25% | 48.27% | NA | Crimap | Du et al., 2010 |
| <i>L. vannamei</i> | AFLP | 42● | 1 | 51 ♂, 47♀ | 182 | 212 | 394 | 2,116 | 2,771 | NA | 15.6 | 17.1 | 59% | 62% | NA | MapManagerQTX MapMaker v 3.0 | Pérez et al., 2004 |
| <i>L. vannamei</i> | AFLP | 192△ | 2 | 18 | NA | NA | 157 | NA | NA | NA | NA | NA | NA | NA | NA | GOMOL | Gonçalves et al., 2014 |
| <i>L. vannamei</i> | AFLP, MSat | 94● | 1 | 45 | 267 | 319 | NA | 3,220.9 | 4,134.4 | NA | 14.5 | 15.1 | 69% | 75% | NA | MapMaker v 3.0 | Zhang et al., 2007 |
| <i>L. vannamei</i> | AFLP, MSat | 43△ | 1 | 49 | NA | NA | 451 | NA | NA | 3,13.9 | NA | NA | NA | NA | 72.63% | Mapmanager QTXb20 | Andriantahina et al., 2013 |
| <i>P. monodon</i> | SNPs | 1,024● | 7 | 44 | 3,486 | 3,592 | 3,959 | 2,917 | 4,060 | NA | 0.9 | 1.2 | 91% | 82% | NA | TMAP | Baranski et al., 2014 |
| <i>P. monodon</i> | AFLP | 126△ | 3 | 20 | NA | NA | 116 | 1,412 | NA | NA | 22 | NA | NA | NA | NA | MapMaker | Wilson et al., 2002 |
| <i>P. monodon</i> | AFLP | 344● | 3 | 44♂, 43♀ 21 sex ave. | 757 | 494 | 70 | 2378 | 2362 | NA | 2.1 | 2.8 | NA | NA | NA | Joinmap 3.0 | Staelens et al., 2008 |
| <i>P. monodon</i> | MSat, EST, SCAR | 76● | 1 | 9 | NA | NA | 27 | NA | NA | 103.6 | NA | NA | NA | NA | NA | Joinmap 2.0 | Wuthisuthimethavee et al., 2005 |
| <i>P. monodon</i> | MSat, EST | 42● | 1 | 47♂, 36♀ | 157 | 111 | NA | 1,101.0 | 891.4 | NA | 7.0 | 8.0 | NA | NA | NA | Mapmaker | Maneevitnanarungroj et al., 2006 |
| <i>M. japonicus</i> | SNPs | 150● | 1 | 41 | 3,927 | 5,849 | 9,298 | 3326.19 | 3,127.23 | 3,610.90 | 0.847 | 0.535 | NA | NA | 99.06% | JoinMap 4.0 | Lu et al., 2016b |
| <i>M. japonicus</i> | AFLP | 102● | 1 | 43♂, 31♀ | 217 | 125 | NA | 1780 | 1026 | NA | NA | NA | NA | NA | NA | MapMaker v 3.0 | Li et al., 2003 |
| <i>F. chinensis</i> | SNPs | 100● | 1 | 21 | 115 | 119 | 180 | 879.7 | 876.2 | 899.3 | 9.4 | 8.9 | 51.94% | 53.77% | NA | JoinMap 3.0 | Zhang et al., 2013 |
| <i>F. chinensis</i> | AFLP | 100● | 1 | 36♂, 35♀ | 194 | 197 | NA | 1,737.3 | 2,191.1 | NA | 11.0 | 13.5 | 68.1% | 69.6% | NA | MapMaker 3.0 | Li Z. et al., 2006 |
| <i>F. chinensis</i> | AFLP | 110△ | 1 | 35♂, 28♀ 44 sex ave. | 144 | 103 | 216 | 1617 | 1090 | 1772.1 | 16.36 | 14.53 | 44.41% | 47.61% | 63.17% | MapMaker v 3.0 | Tian et al., 2008 |
| <i>F. chinensis</i> | MSat, RAPD | 82● | 1 | 10 ♂, 8 ♀ | 46 | 49 | NA | 1,144.6 | 1,173 | NA | 12.05 | 11.28 | 62.01% | 59.36% | NA | Mapmaker v 3.0 | Sun et al., 2008 |
| <i>F. chinensis</i> | AFLP, MSat, RAPD | 100● | 1 | 47 | NA | NA | 354 | NA | NA | 4,580.5 | NA | NA | NA | NA | 75.8% | Joinmap 3.0 | Liu et al., 2010 |

* *L. vannamei* (Litopenaeus vannamei), *P. monodon* (Penaeus monodon), *M. japonicus* (Marsupenaeus japonicus), *F. chinensis* (Fenneropenaeus chinensis); ** 2.67 cM when 0 cM inter-marker distances are removed; SNPs, Single Nucleotide Polymorphisms; MSat, Microsatellite or Simple sequence repeats; EST, Expressed Sequence Tag; SCAR, Sequence Characterized Amplified Region; RAPD, Random Amplified Polymorphic DNA; AFLP, Amplified Fragment Length Polymorphism; △ Phase known; ● Phase unknown; ♀ Half-sib families; ♂ male map; ♀ female map.

block maps have been required to obtain an understanding of recombination across each chromosome of the genome of livestock and crop species (Habier et al., 2007; Amaral et al., 2008). However, given previous constraints generating dense linkage maps, no LDU or haplotype information is currently available for penaeid shrimp, but as advanced selection programs begin to be implemented in the industry generating haplotype maps should be a significant focus of future studies. Alternatively, haplotype information has been generated through either long or short-read (10x Genomics or Illumina) whole genome resequencing studies (multiple individuals at 5–10x coverage per individual) in many species including cattle (Daetwyler et al., 2014). However as sequence derived haplotype maps rely upon the availability of error-free genome assemblies, this approach is not currently feasible for penaeids.

To date, comprehensive comparative genomics studies examining genome synteny/divergence, chromosomal evolution, and structuring between penaeid species has been limited. Excluding the *L. vannamei* map published by Yu et al. (2015), integration of linkage maps with the fragmented genome sequence data has yet to be completed. Furthermore, novel marker sets have been developed and utilized for most maps available, with few maps [i.e., Maneeruttanarungroj et al. (2006) building upon Wilson et al. (2002), and Jones et al. (2017a) building upon Du et al. (2010), and Ciobanu et al. (2010)] having integrated previous mapping resources into the updated maps. Jones et al. (2017a) has undertaken early comparisons of genomic synteny between *P. monodon* and *L. vannamei* [along with other *L. vannamei* maps (Du et al., 2010; Yu et al., 2015)], predominately observing conservation of chromosomal organization. However, due to only a small number of common markers (275 SNPs) between both maps (Baranski et al., 2014; Jones et al., 2017a), it was not possible to thoroughly investigate chromosomal rearrangement, gene order, or differences in recombination (e.g., sex biased, genomic hotspots). While a substantial amount of research effort has been applied to date on developing maps (Table 2), further integration between linkage maps would provide an opportunity to rapidly increase map density for each species. A number of previous maps are not easily accessible, despite, in some cases, providing access points at the time of publication (e.g., Wilson et al., 2002; Maneeruttanarungroj et al., 2006). Ensuring long-term access to mapping resources is critical to maximize the use of previous resources and also enable detailed comparative genomic studies to be undertaken.

To extend the utility of linkage maps and provide the ability to understand the fine scale genomic drivers of commercial traits (e.g., genome-wide association, multifactorial association) it is necessary to increase marker density contained in maps for many species. Maps available for *P. monodon* (Baranski et al., 2014) and *L. vannamei* (Yu et al., 2015), have average inter-marker distances between 0.9 and 0.7 cM, respectively, across different map iterations. While this is a significant advance from earlier published maps, 1 cM equates to an estimated physical genome distance of ~400–600 Kb for penaeids [*P. monodon* 395 Kb/cM (Baranski et al., 2014); *L. vannamei*; 598.89 Kb/cM; (Yu et al., 2015), *M. japonicus* 657.89 Kb/cM (Lu et al., 2016b)],

and presents a significant challenge when looking to characterize potential genes or genomic regions underlying findings of trait association studies. The linkage map available for *M. japonicus* (Lu et al., 2016b) achieved the highest map density with an average inter-marker distance of 0.39 cM, representing an estimated physical distance between markers of ~230–257 Kb (Yu et al., 2015; Lu et al., 2016b). However it should be noted, only 41 linkage groups were reported in this map, rather than 44 observed in other penaeid species (Table 2).

To increase the density and decrease the interval between markers, a number of strategies can be applied. Further genotyping of families, and individuals in each family will provide additional observations of informative meiotic recombination events or integrate unplaced (i.e., orphaned) markers into existing maps (Fierst, 2015). Current maps have either had relatively few individuals from each family, yet many families included [e.g., 8–33 individuals per family from 49 families (Jones et al., 2017a)], or include many individuals from a smaller number of families [e.g., 100 progeny from one family (Lu et al., 2016b)], with both approaches restricting the number of informative meiotic events available for placement of each marker on the map (Jones et al., 2017a).

Likewise, incorporating additional genotyped individuals will assist in teasing apart the positioning and order of “binned” markers (those co-segregating to a single map location) (Fierst, 2015). In the *L. vannamei* linkage map by Jones et al. (2017a) 4,817 markers were mapped to 1,752 unique locations (bins), with an average of 2.75 markers per location. It is likely maps by Baranski et al. (2014), Yu et al. (2015), and Lu et al. (2016b) have observed the same effect, however, it is uncommon that the proportion of “binned” markers is discussed, with Jones et al. (2017a) being the only publication to provide details of both average (0.97 cM) and “non-zero” average inter-marker distance (2.67 cM). The discrepancy between total mapped markers and unique locations should be clearly reported with all linkage maps, as determining distribution of unique locations is important to integration of genome sequence information (Fierst, 2015), understanding of the power and resolution available to study commercial traits, as well as in reducing bias (from uneven marker distribution) in the calculation of genomic estimates of breeding value (Mathew et al., 2018). Utilizing increasingly cost effective genotyping strategies (e.g., genotype by sequencing; see section Dissecting and Exploiting the Genetic Variation Underlying Phenotypes) will facilitate a large number of individuals/families to be genotyped and aid efforts to achieve fine grain marker placement.

The anchoring of long-read sequences produced through ongoing genome assemblies can be used to tease apart the location of markers placed in unordered bins (Fierst, 2015). However, to integrate linkage maps with genome assemblies, markers are required to be sufficiently close together to place multiple markers on each genomic scaffold. It has been indicated that linkage maps have been integrated within genomic assembly strategies for *L. vannamei* (Yu et al., 2015; Xiang, 2016), however, an integrated resource has yet to be released in an accessible format. Once completed, these integrated mapping resources will provide an intermediate point for comparison between otherwise

disparate maps generated through different genotyping assays or with different marker types, and will provide a vital resource for comparative genomic studies and quantitative trait mapping.

Physical and Optical Genome Mapping

While linkage mapping and LRSTs promise to provide a significant increase in the quality of genome assemblies for penaeids, it is still difficult to produce a chromosome level assembly through the addition of these methods alone (Fierst, 2015; Howe and Wood, 2015). The generation of physical map information provides an additional approach for genome scaffolding alongside linkage maps and LRSTs (Fierst, 2015). Bacterial artificial chromosome (BAC) based physical mapping has been utilized previously in the genome assemblies of *Drosophila* (Myers et al., 2000), domestic cattle (*Bos taurus*) (Zimin et al., 2009), and barramundi (*Lates calcarifer*) (Vij et al., 2016), to aid in anchoring short scaffolds and improve the overall assembly quality and accuracy. BACs require the ligation of long fragments of DNA (200 kbp) into bacterial plasmids, after which colonies could be individually screened. Fragments are then reisolated and sequenced individually by either long or short read sequencing, with the overall aim to reduce the complexity of the sequence reassembly process. However, due to the high cost and difficulty in isolating sufficient unique BAC fragments, there is only two partial BAC-end libraries available for *L. vannamei* (Zhang X. et al., 2010; Zhao et al., 2012), one partial Fosmid library for *P. monodon* (random physical sheared ~40 kb library; Huang et al., 2011) and one small insert BAC library for Kuruma shrimp (Koyama et al., 2010). While Yu et al. (2015) have mapped the available BAC-end library to the *L. vannamei* linkage map, providing a valuable resource for fine trait mapping in QTL studies (See section Quantitative Trait Loci Mapping), it is unclear to what degree this data has aided assembly quality (Yu et al., 2015; Xiang, 2016).

In the study of many other species (Howe and Wood, 2015; Vij et al., 2016; Bickhart et al., 2017), the physical mapping information that BACs provide has instead been replaced by modern methods of fluorescent *in situ* hybridisation (FISH) and optical mapping. Through these methods, gene-specific fluorescent probes are developed to enable the direct visualization of a gene's location on a chromosome. By utilizing a combination of multiple probes it is possible to measure the physical distances between genes, validating gene order and distances established from linkage maps and genome assemblies. For example, physical maps can be utilized to identify and correct regions of whole genome assemblies which have ambiguously assembled regions (i.e., repeat regions, unknown gap lengths, chimeric scaffolds), or have included misassemblies (i.e., false joins, inversions, repeat collapsing) (Fierst, 2015; Howe and Wood, 2015). In several cases, due to the length of optical mapping fragments, it is possible to reconstruct reference sequences of entire chromosomes (Howe and Wood, 2015). Yet like BACs, the development of FISH-based optical maps has been limited within aquaculture species, with no optical maps available for penaeids. This is a result of their initial high cost, low throughput, and relatively limited data generation. Currently, only proof of concept studies have been completed with FISH

methods on penaeid species, with a single chromosome-specific FISH probe developed for *F. chinensis* (Huan et al., 2010) and four TAACC repeat specific FISH probes developed for *L. vannamei* (Alcivar-Warren et al., 2006).

While the high cost of initial manual optical mapping techniques has prevented their widespread use, optical mapping techniques are being increasingly refined to assist with the assembly of complex and repetitive genomes. The automation of optical mapping technology has led to the development of a number of commercial platforms, such as Irys (BioNano Genomics) and Opgen (Argus) which can deliver high-throughput physical map data at reduced costs (Howe and Wood, 2015; Sharp and Udall, 2016). High-throughput optical mapping provides a physical fingerprint of each DNA molecule, through the fluorescent labeling of multiple restriction sites or "nick sites" along large linear DNA fragments (20 Kbp to 3 Mbp in length; Sharp and Udall, 2016). With the inherent difficulties observed in constructing a complete chromosomal-scale genome assembly for penaeids, integrating physical mapping information into the assembly process will prove critical to resolving repetitive regions and producing a successful final draft assembly. Given the diverse range of genomic resources (i.e., linkage maps, optical mapping, short, and long range sequencing) now available to address complex assemblies, a completed penaeid genome assembly is likely to be achieved quite soon.

DEVELOPMENT AND APPLICATIONS OF POLYMORPHIC MARKERS

The Replacement of Traditional Genomic Markers

A large amount of research effort over the past three decades has been dedicated to developing a diverse range of traditional genomic markers (i.e., allozymes, microsatellites, AFLP, RFLP markers) for many penaeid species. Their development has increasingly provided the tools to assess wild source populations, manage family lines, and undertake cursory broad-brush assessment of the heritability and genetic architecture of traits. However, many of these markers exhibit caveats which have been reviewed extensively by Benzie (1998, 2009) and Wang et al. (2004). Ultimately, due to high development costs for each marker, the increasing complexity of breeding programs (Jerry et al., 2004; Sellars et al., 2014) and an inability to tease apart complex polygenic production traits, these markers have rapidly fallen out of favor in penaeid research programs and have seen a reduced application within industry. Instead the traditional marker sets described above, are being replaced by powerful and low cost modern marker panels, overcoming the many shortfalls of traditional marker sets.

Modern Genomic Markers and Solid State Arrays for Breeding Programs

Single nucleotide polymorphisms (SNPs) are the most abundant polymorphic markers contained within genomes and are fundamental to modern genomic studies. Large genome sequencing and resequencing projects have identified over 9

million SNPs in the cattle (*B. taurus*) genome (Xu Y. et al., 2017) and 7 million SNPs in chickens (*G. gallus*) (Rubin et al., 2010). Within aquaculture species, over 9.7 million SNPs were identified in Atlantic salmon (*Salmo salar*) through whole genome sequencing of 20 individuals (Yáñez et al., 2016). Similarly in the channel catfish (*Ictalurus punctatus*) and barramundi (*Lates calcarifer*), 8.6 and 5.6 million SNPs were identified through genome re-sequencing of 1,213 and 61 individuals, respectively (Vij et al., 2016; Zeng et al., 2017). In addition to SNP discovery, commercial genotyping arrays containing 50,000 to >500,000 markers have been developed as tools for the study and management of many primary production species, including cattle [BovineSNP50 assay (Matukumalli et al., 2009)], sheep [*Ovis aries*; OvineSNP50 Bead Chip (McEwan and Consortium, 2009); OvineSNP600 Bead Chip (Anderson, 2014)], and wheat [*Triticum* spp.; Axiom HD (820K) wheat genotyping array (Winfield et al., 2015)], as well as aquaculture species, including catfish [250K array (Liu et al., 2014), 690K array (Zeng et al., 2017)] and Atlantic salmon [~130K array (Houston et al., 2014)]. While a number of genome sequencing programs are underway for penaeids species (Yu et al., 2014, 2015; Xiang, 2016), they have yet to develop genotyping arrays of greater than 10,000 markers.

The development and use of high density genotyping arrays for penaeids has been impeded by a number of factors. Most critically, the fundamental genomic resources (high quality genomic assemblies and genomic re-sequencing data) that underpin the development of these high density arrays are not currently as refined as they need to be for penaeid species. With the exception of recent work by Yu et al. (2015), and Lu et al. (2016b) which utilized genotype-by-sequencing approaches (section Genotype by Sequencing as a Low Cost Approach) to produce 114,829 and 28,981 SNPs, studies have largely aimed to generate SNP markers through whole transcriptome studies (Baranski et al., 2014; Yu et al., 2014; Jones et al., 2017a). While this is an effective approach for marker discovery, it also requires costly additional genotyping of samples across the discovered SNPs to validate their performance. To date, only two studies have produced validated SNP genotyping arrays; Baranski et al. (Baranski et al., 2014) for *P. monodon*, and Jones et al. (Jones et al., 2017a) for *L. vannamei* (now sold commercially as the Infinium ShrimpLD-24 v1.0 Bead Chip). These arrays present a significant advancement for these two penaeid industries, containing 6,000 and 6,400 SNPs respectively. As both arrays are based on type-1 SNPs (i.e., genic, rather than intergenic), many of the SNPs have been annotated with putative genes (62 and 47% respectively) providing a significant resource for future studies. Likewise, both studies have also confirmed the distribution of markers through linkage mapping, providing the foundations for further trait mapping studies (Robinson et al., 2014; Khatkar et al., 2017b). Markers selected for both arrays represent only a small fraction of the total markers identified (Baranski et al., 2014; Jones et al., 2017a), with 473,600 and 234,452 SNP identified, respectively. This represents a significant pool of markers that could be utilized in ongoing studies to increase array density if required.

While it is evident that the modern high density genotyping arrays are powerful tools, and their development marks a step forward for penaeid aquaculture, the cost of genotyping (~AU\$ 40-100 per individual) on the platforms selected (Illumina I-Select and Bead chip arrays) remains generally too high for large-scale routine use by the penaeid aquaculture industry. Instead, for species where these high density arrays have been successfully implemented (most notably in agriculture and Atlantic salmon), the price of genotyping has been reduced by increased demand from centralized, cooperative, or national breeding programs which have the advanced technical expertise to utilize the data obtained. The price impediment has been further circumvented by removing the farm by farm requirement for genotyping high numbers of broodstock in each subsequent generation (Thodesen and Gjedrem, 2006; Janssen et al., 2018). In many cases these shrimp breeding programs, are operated by private enterprises that have a either specialized focus on production of seedstock (centralized nucleus) or operate a fully integrated production model (seedstock production through to harvest), however, some government or collaboratively funded programs do exist (Benzie, 2009). In private enterprise operations, the utilization of genotyping resources is reliant upon obtaining an increase in the achievable sale price of seedstock through achieving improved selection practices, as seedstock are marketed upon predicted improved grow-out performance (Janssen et al., 2017). Nevertheless, cost of genotyping has to be outweighed by increased economic returns achieved. Without centralized or highly concentrated breeding in place for all species and countries, and until a point when genotyping arrays can be delivered at a more industry accessible price, the use of these genotyping resources will remain confined to a research setting (Robledo et al., 2017). A number of sophisticated and large scale breeding programs are now underway for *L. vannamei* and *P. monodon* and hold the most promise for high density solid-state genotyping resources to be fully utilized (Castillo-Juárez et al., 2015; Li et al., 2016; Campos-Montes et al., 2017). Early DNA parentage arrays for penaeid shrimp were largely based on microsatellite DNA markers (Alcivar-Warren et al., 2003; Jerry et al., 2004, 2006), however, with the advent of next-generation sequencing it has been possible to more easily isolate SNPs. Consequently, parentage testing SNP arrays have now been developed for *P. monodon* as an intermediate step to industry application of cost-prohibitive high-density SNP genotyping (Sellars et al., 2014). While these ultra-low density arrays, containing 63 and 59 SNPs each, or 122 SNPs as a combined array (Sellars et al., 2014), lack the number of markers required for advanced genomic applications (e.g., genomic selection, GWAS, QTL mapping, and calculating genomic relationship matrixes), they do provide increased power and utility over traditional genomic markers for pedigree assignment (see section Applying Markers for Breeding Population Management) (Henshall et al., 2014). The development of a similar range of SNP arrays is currently feasible for all other commercially important species as extensive transcriptome data has been made publically available for *M. japonicus*, *F. chinensis*, *F. merguensis*, *F. indicus*, and *L. stylirostris*. Furthermore, for these species which lack any modern genotyping tools, replicating a low density array could

provide a cost effective approach for guided farm management and improvement of shrimp production (Vandeputte and Haffray, 2014). Ultimately in the selection of a genotyping method for commercial applications the time required for sample processing, genotyping, and data analysis needs to be considered, particularly given the window between pre-selection of candidate broodstock at harvest, and final breeding selection and spawning is short (often less than 3–6 months). Genotyping technologies such as the Illumina Bead Chip or Sequenom I-plex systems that can achieve both high-throughput and short turnaround times are currently most conducive to rapid decision making required in shrimp breeding.

Genotype by Sequencing as a Low Cost Approach

The regular use of solid state genotyping arrays across thousands of individuals within commercial settings is currently out of reach for the majority of the penaeid aquaculture industry. Nevertheless, the advent of a number of novel genotype-by-sequencing (GBS) approaches (e.g., double digest restriction site associated DNA, ddRAD; Diversity Arrays Technology Sequencing, DArT-Seq; specific-locus amplified fragment sequencing, SLAF) holds promise to rapidly provide a genotyping platform for less than AU\$ 40 per individual. The particular advantage of GBS approaches is the ability to discover and genotype markers (“*de-novo* marker discovery”), without requiring reference to existing genomic information (genomic sequence, transcriptomes). A number of GBS approaches (SLAF-tag, RAD, and DART-Seq) have been utilized in the study of penaeid species, with 25,140 and 23,049 markers obtained separately for *L. vannamei* (Yu et al., 2015; Wang et al., 2017), and 28,981 markers obtained for *M. japonicus* (Lu et al., 2016b). While these markers have been applied to generate linkage maps (Yu et al., 2015; Lu et al., 2016b), undertake QTL mapping (Yu et al., 2015; Lu et al., 2016b), and estimate genomic prediction accuracy (Wang et al., 2017), they have yet to be further utilized within the industry for routine genotyping applications. Further refinement of GBS protocols along with integration of targeted fragment capture technology [e.g., Bates Probes (Ali et al., 2016), Molecular Inversion Probes (Niedzicka et al., 2016)] can deliver arrays with 1,000–10,000 markers for under AU\$ 20 per individual.

The first GBS-based targeted fragment capture technology SNP assay (containing ~5,000 markers) for a penaeid species is in final stages of development (Guppy et al., 2018), and aims to provide the *P. monodon* industry the first cost-effective opportunity to undertake genomic selection. Furthermore, as with solid state sequencing platforms, GBS-based assays can be used to undertake genomic relationship calculations and examine the genome structure and architecture of traits (Zenger et al., 2017). With the development and use of GBS arrays becoming more frequent in aquaculture as a low cost alternative to solid state arrays (Holman et al., 2017; Robledo et al., 2017), it is likely that similar marker panels with tailored density and composition of markers will be developed for other penaeid species (Wang et al., 2017).

The utility of GBS for aquaculture species is reviewed in detail by Robledo et al. (2017), and for genomics more broadly in Andrews et al. (2016). While there are many benefits of GBS based genotyping in terms of flexible design, and improved accessibility through lower cost, it must be noted that, when compared to solid state arrays (Illumina I-select and bead chip arrays or Affymetrix Axiom Arrays), the genotype data obtained can be of lower quality (e.g., null alleles, read-depth dependant genotype accuracy, high missing data) if not filtered through appropriate quality control measures (Andrews et al., 2016). To account for this, custom pipelines [e.g., STACKS (Catchen et al., 2013), TASSEL (Glaubitz et al., 2014), DARTR (Bernd et al., 2018)] are being continuously refined and will provide valuable software resources to more effectively utilize data generated by this genotyping method. Likewise, as the generation of genotyping assays through GBS approaches is still an emerging area of research, the process of providing functional and validated assays is currently more complex than solid-state arrays.

Applying Markers for Breeding Population Management

Managing a sufficiently large and diverse core population of breeding individuals has always been a significant challenge for the shrimp industry and there has not been a cost-effective method to recover pedigree data on-farm. Programs relying on the separate rearing of family lines have been overshadowed by high infrastructure costs and an inability to hold sufficient numbers from each family line to make effective selection decisions. In addition, by raising individuals in separate rearing tanks, or ponds, farmers inadvertently introduce confounding environmental effects to any comparative estimates of performance between rearing systems (Sonesson and Ødegård, 2016). Furthermore, it is not possible to maintain genealogical traceability with external tags as shrimp molt as they grow, and are too small in size (~2 mm) prior to stocking into commercial ponds for feasible use of internal tags (visible implant elastomer or passive integrated transponders). Overlooking accurate tracking of pedigree has on a number of occasions led to poor management strategies being employed, and inbreeding being observed within penaeid farms (Garcia et al., 1994; Wolfus et al., 1997; Xu et al., 2001; Moss et al., 2007; Dixon et al., 2008; Knibb et al., 2014).

Nevertheless, as access to genomic markers has increased, breeding programs have been provided with the tools (i.e., genetic pedigree assignment and genetic diversity assessment) to assist in farm management. While these resources have improved substantially, industry use for penaeids has been inconsistent. The cost of genotyping (high-density solid state arrays), lack of genotyping power (microsatellites), or a combination of both factors, have been the primary impediments to wider spread use of many existing marker sets (Vandeputte and Haffray, 2014). Importantly, however, both low-density solid state arrays and GBS assays provide the penaeid industry with a cost-accessible and powerful set of tools for real world application. For instance, as demonstrated by Henshall et al. (2012, 2014), proportions of communally reared families can be determined from pooled

samples (i.e., DNA pooling of multiple individuals) that have been genotyped with low density SNP arrays (Sellars et al., 2014). Both low-cost genotyping and DNA pooling form an important progression for the industry as they allow phenotypic data from commercial ponds to be linked back to broodstock families without the need to physically track pedigree or undertake isolated rearing of individual families. By doing so, the confounding effects of variable rearing conditions upon estimates of performance can be reduced (Kinghorn et al., 2010). When phenotypic information is combined with individual genetic pedigree assignment of progeny, superior broodstock can be selected to produce improved subsequent generations (Henshall et al., 2014), as well as estimate genotype-by-environment (GxE) interactions and their effect on individual performance under commercial rearing conditions.

Furthermore, with dense marker sets now available (Jones et al., 2017a,b; Guppy et al., 2018) it is also possible to rebuild deep pedigrees and examine the hidden cryptic relatedness between individuals (Jones et al., 2017b; Khatkar, 2017). Genomic relatedness matrixes (GRMs) can be generated which quantify the extent of the genome shared between individuals. This methodology characterizes the true relatedness between individuals more accurately than traditional pedigree or parentage assignment (Hayes et al., 2009). In particular, when applied to breeding programs with complex and cryptic family lines, GRMs are able to minimize inbreeding through careful selection and allocation of mate pairings (Nielsen et al., 2011; Jones et al., 2017b; Khatkar, 2017; Toro et al., 2017; Zenger et al., 2017). The transition to higher density marker panels (i.e., >3,000 SNPs) will increase the ability to definitively assign parentage of communally reared individuals, particularly as more polymorphic markers provide a higher capacity to determine relationships between inherently related individuals produced over multiple generations of breeding. GBS assays rather than low density solid state arrays may prove to be more versatile as they maintain an accessible genotyping cost, while providing required genotyping power (Robledo et al., 2017).

An alternative approach to increase marker density is genotype imputation. Imputation relies upon genotyping a small reference panel of individuals at high-density (e.g., 1,000 individuals) to generate robust allele frequency estimates. Once paired with genetic linkage information, the genotypes of high-density markers can be inferred within a second set of individuals (e.g., 10,000 individuals) genotyped with a low-density assay (Li et al., 2009). This approach, while statistically intensive, has proved effective for genotyping in breeding programs of terrestrial species (Pryce et al., 2014), and has been proposed to improve the cost effectiveness of large scale salmon genotyping (Tsai et al., 2017). Extensive modeling of the accuracy of imputed genotypes has been undertaken in terrestrial species (Hozé et al., 2013), but the accuracy of imputation is underpinned by factors including reference population size along with the size of chromosomal fragments (haplotype or haplotype blocks) that are co-inherited (García-Ruiz et al., 2015). Currently there is a poor understanding of length and rate of rearrangement (or recombination) of haplotypes within the penaeid genome (See

section Linkage Mapping of Genetic Markers), impeding the current utilization of imputation for penaeid shrimp.

FUNCTIONAL GENOMICS

Examining the physiology, immunology, and genetics of penaeids at the transcriptional and translational level provides direct insight into the functional role of molecular mechanisms in the overall productivity of penaeid species, along with many other traits of biological and commercial interest (e.g., disease resistance, pigmentation, nutrition, reproduction). A number of methods are commonly employed to undertake this work, progressing from individual “gene by gene” characterisation (e.g., RT-qPCR, Sanger sequencing) through to profiling all transcriptionally active genomic elements simultaneously (i.e., RNA-seq). Proteomic analysis methods have traditionally focussed on characterisation and quantification of a small number of proteins (e.g., 2D gel separation), though more recent approaches can successfully characterize and quantify the complete constituency of the proteome (e.g., liquid chromatography coupled with mass spectrometry). As with other fields, functional genomics has been revolutionized by advances in next generation sequencing, mass spectrometry, and bioinformatics approaches.

Expressed Sequence Tag Sequencing and cDNA Microarray Development

Prior to the advent of high-throughput next-generation sequencing, and in the absence of a complete genomic sequence for penaeids, expressed sequence tag (EST) discovery and annotation was the focus of functional genomics studies for a significant period of time, with a substantial investment made in developing resources for research and commercial applications. Driven by the economic handicap that reproductive dysfunction and disease events have had upon the industry, these two areas have seen the majority of research effort applied.

In particular, EST studies have been conducted to identify potential genes of interest related to male (Leelatanawit et al., 2008, 2009) and female (Yamano and Unuma, 2006; Preechaphol et al., 2007, 2010) reproductive functions. Likewise, comparisons of EST libraries between challenged individuals (i.e., exposed to stress and pathogens), and non-challenged individuals, have allowed the identification of genes with putative roles in virally-induced immune response (Gross et al., 2001; Rojtinnakorn et al., 2002; Shen et al., 2004; Clavero-Salas et al., 2007; Dong and Xiang, 2007; Leu et al., 2007; Pongsomboon et al., 2008b; Maralit et al., 2014) and bacterial resistance (Bartlett et al., 2002; Supungul et al., 2002, 2004; de Lorgeril et al., 2005; Somboonwiwat et al., 2006). While these studies were effective in the initial discovery of a small number of penaeid genes (**Table 3**), alone they provide an incomplete view of the transcriptionally active genes in each biological situation investigated. Furthermore, additional complexity reduction steps [e.g., suppressive subtraction hybridisation (Leelatanawit et al., 2008; Preechaphol et al., 2010)] were required to avoid

redundant repeat sequencing of highly expressed genes such as thrombospondin from ovary libraries, or 40S ribosomal proteins from gill libraries (Clavero-Salas et al., 2007; Preechaphol et al., 2007). Without these additional steps, studies such as those by Clavero-Salas et al. (2007) contained 69% redundancy in sequencing, limiting the novel gene sequences obtained.

Through substantial investment in collaborative research programs, large scale multi-tissue and multi-life-stage studies have also been conducted with the aim of obtaining the complete transcriptional profile for penaeid species. A large number of novel ESTs have been identified in *P. monodon* (Tassanakajon et al., 2006), *L. vannamei* (O'Leary et al., 2006; Robalino et al., 2007), and *F. chinensis* (Dong and Xiang, 2007; Xiang et al., 2008), yet due to the small sequence length of each EST transcript (100–300 bp), the characterisation of transcriptional profile of these penaeids through classical sequence homology approaches has proved difficult. Many of these studies have been reviewed at length by Leu et al. (2011), with only an additional 20,318 EST tags deposited into the NCBI EST database since (was 196,248 in 2011, now 216,566 as of 8th May 2018; 709 for *P. indicus*, 5,609 for *L. vannamei*, 14,000 for *P. monodon*). While the number of ESTs is small in comparison to those generated for humans (>8 million) or Atlantic salmon (>500,000), valuable insight into genes involved in processes such as reproduction of penaeids has been obtained, with over 500 reproduction-related genes identified in *L. vannamei* and *P. monodon* (Leu et al., 2011).

Large sets of ESTs have been incorporated into a number of custom high throughput microarrays, which are able to provide relative expression profiles for 1,000's of genes simultaneously (Karoonthaisiri et al., 2009; Wongsurawat et al., 2010; Leelatanawit et al., 2011). In particular, microarrays are now available for the study of reproductive function [ReproArray (Karoonthaisiri et al., 2009; Wongsurawat et al., 2010), UniShrimpChip (Leelatanawit et al., 2011, 2017; Uawisetwathana et al., 2011)] and disease response within penaeids (Dhar et al., 2003; Wang et al., 2006; Wongpanya et al., 2007; Pongsomboon et al., 2008a, 2011; Veloso et al., 2011; Shi et al., 2016). While these studies have provided novel insight into functional genomics of penaeids (see review by Aoki et al., 2011), they provide limited novel sequence information and therefore limit the ability to differentiate between biologically significant isoforms or genetic variants. As such, high-throughput whole transcriptome sequencing approaches are instead becoming the preferred method to study the RNA profile of penaeids, due to their ability to simultaneously provide differential gene expression and sequence data.

Transcriptome Profiling With Next Generation Sequencing

With the decreasing price of sequencing, improvements in library preparation, and bioinformatic techniques, it is possible to analyse complete cDNA libraries without the need for any complexity simplification steps. In particular, without the requirement to clone and individually sequence cDNA fragments (as is needed with EST sequencing), the ability to produce

complete transcriptome profiles has increased dramatically (Li et al., 2012; Ghaffari et al., 2014; Santos et al., 2014; Huerlimann et al., 2018). It is now possible to isolate upwards of 50,000 genes, including rare variants in a single study (Table 3).

Like EST and microarray studies, the primary focus of transcriptomics studies has been immunology, disease resistance and reproductive biology. Screening experimentally challenged, or opportunistically sampled shrimp with infections, has resulted in the identification of a number of differentially expressed genes in *L. vannamei* (Chen et al., 2013; Sookruksawong et al., 2013; Xue et al., 2013; Zeng et al., 2013), *F. chinensis* (Li et al., 2013), *P. monodon* (Soonthornchai et al., 2016), and *F. merguensis* (Powell et al., 2016). Through implementation of whole transcriptome sequencing, it has been possible to gain extensive insight into the host-virus interaction underpinning mass mortality events and has allowed researchers to develop potential methods to counter infectious disease events (Sellars et al., 2011; Li et al., 2013). Comparison between the immune response invoked by different viral pathogens has highlighted common innate immune pathways that may be useful for future selection of individuals based on broad spectrum immune response (Sookruksawong et al., 2013; Tassanakajon et al., 2013; Gao et al., 2015).

With the comparative ease of generating transcriptome profiles, a number of studies have not only isolated the molecular pathways responsible for tolerance to commonly observed environmental stressors such as nitrite (Guo et al., 2013) and ammonia (Lu et al., 2016a), but have also investigated their effect on productivity. For example, many genes and pathways linked to immune response (e.g., chitinase, peritrophin, thrombospondin, and peaeidin) and growth (linoleic acid metabolism) were identified by (Lu et al., 2016a) to be suppressed when investigating the effect of ammonia exposure in *L. vannamei*. Further research, linking environmental stressors (i.e., pH, salinity, and temperature) with studies on viral immunology, will provide a more complete understanding of the role that environmentally induced immune suppression plays in mass mortality disease events and decreased productivity.

Likewise, reproductive studies comparing male and female transcriptome profiles, along with various reproductive stages within sexes, has yielded novel insight into the molecular and genetic mechanisms underpinning maturation and sex determination of *L. vannamei* (Peng et al., 2015), *F. merguensis* (Powell et al., 2015), and *P. monodon* (Rotllant et al., 2015; Goodall, 2017). As these studies are able to reconstruct complete (or near complete) RNA transcripts, it has been possible to analyse comprehensive gene pathways (Peng et al., 2015). Through the characterisation of functional gene regions it has been possible to deduce key gene to gene interactions and regulatory pathways for reproduction and maturation of male and female penaeids (Peng et al., 2015). For *P. monodon*, studies such as those undertaken by Rotllant et al. (2015) provide substantial opportunity to combat reproductive dysfunction in captive bred stocks (Arnold et al., 2013; Coman et al., 2013; Marsden et al., 2013). In particular, through differential gene expression analysis of whole-transcriptome data, genes related to fatty acid and steroid metabolism were found to have altered

TABLE 3 | Summary of currently available functional genomics studies, including the generation of ESTs, microarrays, and transcriptome data.

| Technique | Species* | Factor examined | Tissues | Number of transcripts | Number of unigenes | Annotation Rate | References | |
|--------------------------------|---------------------------|----------------------------|--|---------------------------------|--------------------|------------------------------|----------------------------|---------------------------|
| Expressed Sequence Tags (ESTs) | <i>L. vannamei</i> | WSSV | Gills | 601 ESTs | – | 87% | Clavero-Salas et al., 2007 | |
| | <i>L. vannamei</i> | Innate immune response | Hemolymph, eyestalk, hepatopancreas, lymphoid organ, nerve cord, gill | 13,656 ESTs | – | 38% | O'Leary et al., 2006 | |
| | <i>L. vannamei</i> | Antimicrobial peptides | Haemocytes | 14 clones | – | 100% | Bartlett et al., 2002 | |
| | <i>P. monodon</i> | ♀ reproduction | Ovary | 1,051 clones | 559 | – | Preschaphol et al., 2007 | |
| | <i>P. monodon</i> | ♂ reproduction | Testis | 896 clones | 601 | 45.2% | Leelatanawit et al., 2009 | |
| | <i>P. monodon</i> | <i>Vibrio harveyi</i> | Haemocytes | 1,062 ESTs | 30** | 10.8% | Supungul et al., 2004 | |
| | <i>P. monodon</i> | <i>Vibrio harveyi</i> | Haemocytes | 267 clones | 24 | 31% | Somboonwiwat et al., 2006 | |
| | <i>P. monodon</i> | <i>Vibrio harveyi</i> | Lymphoid organ | 1,033 clones | 22 | 25.2% | Pongsomboon et al., 2008b | |
| | <i>P. monodon</i> | <i>Vibrio harveyi</i> | Haemocytes | 615 ESTs | 21** | 44% | Supungul et al., 2002 | |
| | <i>P. monodon</i> | Gene discovery | Eyestalk, ovary, hepatopancreas, haematopoietic tissues, haemocyte, lymphoid organ | 10,100 clones | 4,845 | 48.8% | Tassanakajon et al., 2006 | |
| | Subtractive Hybridisation | <i>F. chinensis</i> | Innate immune response | Haemocytes | 2,371 clones | 482 | 49.6% | Dong and Xiang, 2007 |
| | | <i>F. chinensis</i> | Innate immune response | Cephalothorax | 10,446 EST | 3,120 | 44.0% | Xiang et al., 2008 |
| | | <i>L. setiferus</i> | Antimicrobial peptides/ | Haemocytes | 7 clones | – | 100% | Bartlett et al., 2002 |
| | | <i>P. monodon</i> | WSSV | Gills | 9,597 ESTs | 65 | 8.6% | Maralit et al., 2014 |
| | | <i>P. monodon</i> | ♀ reproduction | Ovary | 452 | 109 | 48.0% | Preschaphol et al., 2010 |
| | | <i>P. monodon</i> | ♂ reproduction | Testis | 365 | 71 | 45.5% | Leelatanawit et al., 2008 |
| <i>M. japonicus</i> | | Sex and reproductive genes | Male and female gonads | 296 transcripts | 147, 36 | 7.3–43.8% | Callaghan et al., 2010 | |
| Microarray | | <i>L. vannamei</i> | TSV, YHV | Hepatopancreas | 2,469 | – | – | Veloso et al., 2011 |
| | | <i>L. vannamei</i> | Immune response | Hemocytes, gill, hepatopancreas | 7,021 | 2,469 | 36% | Robalino et al., 2007 |
| | | <i>P. monodon</i> | <i>Vibrio harveyi</i> , YHV, WSSV | Haemocytes | 9,990 ESTs | 420, 1,136, 1,954 | 15–25% | Pongsomboon et al., 2011 |
| | <i>P. monodon</i> | YHV | Haemolymph | 2,028 | 1,269 | 47% of the 105 DEG | Pongsomboon et al., 2008a | |
| | <i>P. monodon</i> | WSSV | Hepatopancreas | 47 ESTs, 37 cDNA | – | 100% | Dhar et al., 2003 | |
| | <i>P. monodon</i> | WSSV | Haemocytes | 1,026 | 30–135 | 52.4% | Wongpanya et al., 2007 | |
| | <i>P. monodon</i> | Stress | Hemocytes | 3,853 | – | 17% of the 145 DEG | Vega et al., 2007 | |
| | <i>P. monodon</i> | ♂ reproduction | Testis | 4,803 | 2,702 | 41.9% | Wongsurawat et al., 2010 | |
| | <i>P. monodon</i> | ♂ reproduction | Libraries from Tassanakajon et al. (2006) | 10,536 | 5,568 | 70.6% | Leelatanawit et al., 2011 | |
| | <i>P. monodon</i> | ♂ reproduction | Testis and vas deferens | 6,072 | – | 42.6% of the 162 DGE | Leelatanawit et al., 2017 | |
| <i>P. monodon</i> | ♀ reproduction | Ovaries and testes | 4,992 | – | – | Karoonuthaisiri et al., 2009 | | |
| | ♀ reproduction | Eyestalk | 1,988 | 1,386 | 16.7% | Yamano and Unuma, 2006 | | |

(Continued)

TABLE 3 | Continued

| Technique | Species* | Factor examined | Tissues | Number of transcripts | Number of unigenes | Annotation Rate | References |
|----------------------|------------------------------|--|--|-----------------------|--------------------|-----------------------|----------------------------|
| Transcriptomics | <i>M. japonicus</i> | YHV | Haemocytes | 2,028 | 759 | 47% of 105 DEG | Pongsomboon et al., 2008a |
| | <i>F. chinensis</i> | WSSV | Cephalothorax | 3,136 ESTs | – | – | Wang et al., 2006 |
| | <i>F. chinensis</i> | WSSV | Hepatopancreas | – | 59,137 | 75 genes** | Shi et al., 2016 |
| | <i>L. vannamei</i> | WSSV | Hemocytes | 101,479 contigs | 52,073 | 45.3% | Xue et al., 2013 |
| | <i>L. vannamei</i> | WSSV | Hepatopancreas | – | 14,583 | 73.24% | Chen et al., 2013 |
| | <i>L. vannamei</i> | WSSV and growth | Hepatopancreas and muscle | 63,105 | 14,124 | 33% | Santos et al., 2018 |
| | <i>L. vannamei</i> | TSV | Hemolymph and hemocytes | 61,937 | – | 20.0% | Sookruksawang et al., 2013 |
| | <i>L. vannamei</i> | TSV | Hepatopancreas | – | 15,004 | 69.5% | Zeng et al., 2013 |
| | <i>L. vannamei</i> | Gonadal development | Testis and Ovaries | 65,218 | 30,304 | 46.5% | Peng et al., 2015 |
| | <i>L. vannamei</i> | Acute ammonia stress | Hepatopancreas | 94,627 | 78,636 | 28.84% | Lu et al., 2016b |
| | <i>L. vannamei</i> | Osmoregulatory Stress | Hepatopancreas | 26,034 | 38,237 | 25.2% | Chen et al., 2015 |
| | <i>L. vannamei</i> | Osmoregulatory Stress | Gills | 466,293 | 349,012 | 7.03% | Zhang et al., 2016 |
| | <i>L. vannamei</i> | Nitrite | Hepatopancreas and hemocytes | 92,821 contigs | 42,336 | 55.6% | Guo et al., 2013 |
| | <i>L. vannamei</i> | Embryo development | Whole larvae (ZODAH) | 162,342 scaffolds | 73,505 | 44.1% | Li et al., 2012 |
| | <i>L. vannamei</i> | Larval Development | Embryo, Nauplius, zoea, mysis, post larvae | – | 66,815 | 48.4% | Wei et al., 2014 |
| | <i>L. vannamei</i> | Molting | Whole shrimp | – | 93,756 | 16.6% | Gao et al., 2015 |
| | <i>L. vannamei</i> | Feed efficiency | Muscle | – | 72,120 | 28.63% | Dai et al., 2017 |
| | <i>L. vannamei</i> | None specifically | muscle, hepatopancreas, gills, pleopods | 110,474 | 87,307 | 24.4% | Ghaffari et al., 2014 |
| | <i>P. monodon</i> | Growth | Heart, muscle, hepatopancreas and eyestalk | 239,135 (raw) | 69,089 | 17.8% | Nguyen et al., 2016 |
| | <i>P. monodon</i> | Reproduction and development | Hepatopancreas and ovary | 13,288 contigs | 15,867 | – | Rotliant et al., 2015 |
| <i>P. monodon</i> | Gene discovery | Eyestalk, stomach, female gonad, male gonad, gill, haemolymph, hepatopancreas, lymphoid organ, tail muscle, embryos, nauplii, zoea, and mysis, whole larvae (PL1, 4, 10, 15) | 236,388 | – | – | Huefmann et al., 2018 | |
| <i>M. japonicus</i> | Embryo development | Fertilized eggs, embryos and vegetal halves | 46,781, 41,567 | – | 7.4–10% | Sellars et al., 2015 | |
| <i>F. chinensis</i> | WSSV | Cephalothorax | – | 46,676 | 46.28% | Li et al., 2013 | |
| <i>F. merguensis</i> | Color | Cuticle, muscle, androgenic gland, hepatopancreas, stomach, nervous system, eyestalk, male gonads, female gonads | 5,990 | – | 59.8% | Ertl et al., 2013 | |
| <i>F. merguensis</i> | Reproduction and development | Hepatopancreas, stomach, eye stalk, nerve cord, male gonad, female gonad, androgenic gland region, muscle and cuticle | 124,631 | 59,179 proteins | 34.52% | Powell et al., 2015 | |

**L. vannamei* (*Litopenaeus vannamei*), *P. monodon* (*Penaeus monodon*), *M. japonicus* (*Marsupenaeus japonicus*), *F. chinensis* (*Fenneropenaeus chinensis*), *F. merguensis* (*Fenneropenaeus merguensis*), *L. setiferus* (*Litopenaeus setiferus*); **Information on remaining transcripts/ESTs not reported; ♂ Male, ♀ Female; WSSV, White spot syndrome virus; YHV, Yellow Head Virus; TSV, Taura Syndrome Virus.

expression patterns when compared between wild sourced and domesticated stock (Rotllant et al., 2015). Similarly, exploration of the effects variable ovarian arachidonic acid content has on a number of key ovarian development pathways, revealed distinct differences between domesticated individuals (Goodall, 2017). Both studies provide a strong case for continued research into the informed improvement of reproductive maturation diets and may help the industry overcome commonly poor breeding performance (Rotllant et al., 2015; Goodall, 2017).

Interestingly, while functional genomics studies have also investigated the mechanisms involved in embryonic development (Li et al., 2012; Sellars et al., 2015) and molting (Gao et al., 2015), little whole transcriptome research has focused explicitly on muscle growth (Nguyen et al., 2016), to date being limited to confirmation of conserved growth genes, potentially overlooking important shrimp specific genes related to growth. Likewise, differential gene expression analysis undertaken by Dai et al. (2017) found expression of 383 genes relating to cell proliferation, growth, and energy and nutrient metabolism, showed distinct differences between selected lines with high and low feed efficiency, however only 57% (220) of these differentially expressed genes were annotated. Further research investment is required to gain an enhanced understanding of the functional genomic elements that dictate superior or inferior performance for both of these key economic traits.

As with earlier EST studies, functional annotation of data from transcriptome studies through classical sequence homology approaches has been difficult. Predominately, shrimp transcriptome studies are able to annotate between 20 and 50% of transcripts with Kyoto Encyclopedia Genes and Genomes (KEGG), Gene Ontology (GO), or NCBI databases (Table 3). As these database have inadequate penaeid sequence information, along with poor characterized sequence information from close relatives (Decapods, or Arthropoda), they hold limited utility in characterizing shrimp or crustacean specific transcripts which may be of high importance for traits such as disease resistance, reproduction, or growth (Huerlimann et al., 2018). Short-read based RNA sequencing (e.g., Illumina Hi-Seq) is high-throughput and relatively inexpensive, and as such has been the most widely applied method for recent penaeid transcriptomics studies. However, reconstructing full-length transcripts during transcriptome assembly is often problematic, with misassemblies commonly observed, as well as poor representation of many alternatively spliced isoforms. In contrast, emerging long-read RNA sequencing methods (e.g., Pacific Bioscience Iso-Seq- or Nanopore Direct RNA sequencing) can sequence individual full-length RNA fragments and may provide increased insight into isoform diversity particular in non-model species (Kim et al., 2017; Xiaoxian et al., 2017). Further studies incorporating epigenomics methodologies such as whole genome bisulphite sequencing and DNA methylation analysis will also provide insight into regulation of RNA expression, however to date this research has been limited to development of a reduced cost methodology (He et al., 2015).

It is not possible to apply whole- transcriptome sequencing approaches on a commercially relevant scale (1000's of individuals). As such, a concerted effort needs to be applied

to utilize existing data to develop commercially applicable transcriptomics tools for Penaeids. Unlike the extensive number available for livestock species there has been no commercially available tests developed from the transcriptomics research completed, despite the significant of number studies on penaeids to date (Table 3).

Proteomics

Proteomics techniques have been applied across a range of aquaculture species (reviewed in Rodrigues et al., 2012), often with a focus on environmental toxicology. Within penaeid shrimp, proteomic analyses have been focused on understanding the effects of nutrition (Silvestre et al., 2010; Qiao et al., 2011), immunity (Robalino et al., 2009), viral infection (Chongsatja et al., 2007; Rattanarojpong et al., 2007; Wang et al., 2007; Bourchookarn et al., 2008; Robalino et al., 2009; Chai et al., 2010; Kulkarni et al., 2014; Li et al., 2014; Chen et al., 2016), bacterial infection (Wu et al., 2007; Somboonwivat et al., 2010; Zhang J. et al., 2010; Chaikerasitak et al., 2012), and environmental stress (Jiang et al., 2009; Fan et al., 2013, 2016; Xu C. et al., 2017). In much the same way as transcriptomics, these proteomic studies examine differentially expressed proteins from shrimp tissues under various pathogenic or environmental stressors compared with those under normal states. The most widely applied proteomics workflows to date have utilized a range of gel-based separations often over 2-dimensions (2D) that separates proteins by isoelectric point then by size. This is followed by image analysis, proteolytic digestion and identification of differentially expressed protein spots using various mass spectrometry (MS) or tandem mass spectrometry (MS/MS) techniques (Table 4). Although technically demanding, 2D gels are capable of identifying large numbers of unique protein spots, as well as determining post-translational modifications such as glycosylation and phosphorylation. However, quantifying differential changes in protein abundance in 2D requires specialized software, highly technical skills and establishment of appropriate intensity thresholds to prevent over or under estimations due to electrophoretic artifacts. This may allow the detection of several hundred protein spots, but vastly reduces the successful quantification or identification of proteins to between 10 and 50 proteins (Table 4). When successful, 2D gels have identified a small numbers of proteins that are likely to have functional significance under relevant culture or stress conditions, with some studies confirming as functional importance through complementary techniques (e.g., RNAi; Robalino et al., 2009).

Further methodological advances now allow analysis of complex mixtures of soluble proteins without the need for difficulties of running 2D gels and subsequently characterizing each protein spot. By coupling liquid chromatography (LC) with MS/MS in an approach termed shotgun proteomics, databases of known proteins can then be utilized to identify peptide mass fingerprints. Shotgun proteomics has yet to be applied widely to the study of penaeid shrimp, but holds potential to rapidly improve protein identification and functional characterisation. Of the only two examples to date in shrimp (Wu et al., 2007; Xu C. et al., 2017), the first applied two proteomic approaches,

TABLE 4 | Summary of currently available penaeid proteomic studies.

| Technique | Species | Factor examined | Tissue | Proteins Identified | Differentially Expressed | Detailed technique | References |
|--------------------|------------------------------|-----------------------------|---|---------------------------|--------------------------|---|-----------------------------|
| 2D LC-MS/MS | <i>P. monodon</i> | Antibiotics | Hemolymph | 2394 spots 2 proteins | 9 | 2D DIGE nanoLC-ESI-MS/MS | Silvestre et al., 2010 |
| | <i>P. monodon</i> | <i>Vibrio harveyi</i> | Lymphoid organ | ~1000 spots 20 proteins | 27 | 2D gels nanoLC-ESI-MS/MS | Chaikheeratsak et al., 2012 |
| | <i>P. monodon</i> | <i>Vibrio harveyi</i> | Hemocytes | nd spots 29 proteins | 27 | 2D gels nanoLC-ESI-MS/MS | Somboonwivat et al., 2010 |
| | <i>P. monodon</i> | WSSV | Stomach | >200 spots 17 proteins | 20 | 2D gels nanoLC-ESI-MS/MS | Kulkarni et al., 2014 |
| | <i>P. monodon</i> | YHV | Lymphoid organ | 370–420 spots 26 proteins | 33 | MALDI-TOF MS, nanoLC-ESI-MS/MS | Bourchookarn et al., 2008 |
| | <i>L. vannamei</i> | WSSV | Gills | ~500 spots 40 proteins | 20 | nanoLC-ESI-MS/MS | Chen et al., 2016 |
| | <i>L. vannamei</i> | TSV | Hemocytes | 320–400 spots 20 proteins | 32 | 2D gels nanoLC-ESI-MS/MS | Chongsatja et al., 2007 |
| | <i>L. vannamei</i> | WSSV | Gills | nd spots 37 proteins | 37 | Not described | Robalino et al., 2009 |
| | <i>L. vannamei</i> | Vitamin C and Chinese herbs | Hemocytes Hepatopancreas | 220–400 spots 10 proteins | 29/28 | 2D gels MALDI-TOF/MS | Qiao et al., 2011 |
| | <i>L. vannamei</i> | Cold stress | Hepatopancreas | 913 spots 37 proteins | 37 | 2D gels MALDI-TOF/TOF MS | Fan et al., 2016 |
| | <i>L. vannamei</i> | Cold stress | Hemocytes | 755 spots 30 proteins | 30 | 2D gels MALDI-TOF/TOF MS | Fan et al., 2013 |
| | <i>L. vannamei</i> | YHV | Gills | >200 spots 20 proteins | 18 | 2D gels nanoLC-ESI-MS/MS | Rattanaojpong et al., 2007 |
| | <i>L. vannamei</i> | WSSV | stomach | 500 spots 53 proteins | 27 | 2D gels nanoLC-ESI-MS/MS | Wang et al., 2007 |
| | <i>F. chinensis</i> | Hypoxic stress | Hepatopancreas | 620–640 spots 33 proteins | 67 | 2D gels LC-ESI-MS/MS | Jiang et al., 2009 |
| | <i>F. chinensis</i> | WSSV | Hemocytes | 580 spots 33 proteins | 45 | 2D gels MALDI-TOF MS/MS | Li et al., 2014 |
| | <i>F. chinensis</i> | WSSV | Hepatopancreas | 580–600 spots 81 proteins | 81 | 2D gels MALDI-TOF/TOF MS | Chai et al., 2010 |
| | <i>F. chinensis</i> | <i>Vibrio anguillarum</i> | Lymphoid organ | 700 spots 95 proteins | 17 | 2D gels nanoLC-ESI-MS/MS | Zhang J. et al., 2010 |
| Shotgun Proteomics | <i>Penaeus monodon</i> | WSSV | Cuticular epithelium | 429 proteins | 12 | iCAT or SCX MALDI-TOF/TOF-MS/MS | Wu et al., 2007 |
| | <i>Litopenaeus vannamei</i> | Salinity | Hepatopancreas | 533 proteins | 84 | iTRAQ isobaric labeling, SCX, nanoLC-MS/MS | Xu C. et al., 2017 |
| | <i>Geocarcinus lateralis</i> | Eyestalk ablation | Y-organ | 543 proteins | 259 | 2D gels LC MALDI-TOF MS/MS | Lee and Mykles, 2006 |
| | <i>Cancer borealis</i> | None | Pericardial organ Sinus glands (eyestalk) brain | 142 proteins | – | MALDI FTMS; MALDI TOF MS/MS; nanoLC-ESI MS/MS | Ma et al., 2009 |

ESI, electrospray ionization; FTMS, Fourier transform mass spectrometry; iCAT, isotope-coded affinity tags; LC, liquid chromatography; MALDI, matrix-assisted laser desorption/ionization; MS, mass spectrometry; MS/MS, tandem mass spectrometry; SCX, strong cation exchange; TOF, time of flight.

[sub-cellular fractionation and a cleavable isotope-coded affinity tag (iCAT)], each followed by peptide identification with matrix-assisted laser desorption/ionization tandem time of flight mass spectrometry (MALDI-TOF/TOF-MS/MS) to examine WSSV-infected shrimp. Once critical filtering and error rate calculations were performed, a total of 429 proteins were identified with a high degree of confidence. In addition, iCAT labeling was applied to assessing changes in protein abundance over a time series post-infection. The second used 8-plex isobaric tags for relative and absolute quantification (iTRAQ), followed by nano LC-MS/MS analysis to identify and quantify differently expressed proteins in the hepatopancreas of *L. vannamei* exposed to low salinity (Xu C. et al., 2017). A total of 529 proteins were identified, of which 84 were statistically different between treatments. Both these studies represent an approximate 10-fold improvement in protein identifications over previous techniques (Table 4). Tissues that contain highly abundant proteins, such as hemocyanin contained within crustacean hemolymph, remain an obstacle to global scale identification of proteins using shotgun proteomics, but applications in other tissues or using protein labels await investigation.

The potential of shotgun proteomics has been demonstrated in other crustaceans, with studies of neuropeptide signaling pathways controlling molting in crabs successfully identified 543 proteins from the Y organ in the land crab *Geocarcinus lateralis* (Lee and Mykles, 2006) and 142 from *Cancer borealis* (Ma et al., 2009). These studies culminated in the proposition of a novel mechanism by which molt inhibiting hormone and nitric oxide synthase activates a signaling cascade of phosphoproteins that regulate molting and reproductive development (Lee and Mykles, 2006). A gel-free proteomic workflow using LC-ESI-MS/MS study revealed the identity of 62 proteins differentially expressed in hepatopancreas of prawn *Macrobrachium rosenbergii* after exposure to the insecticide chlordecone (Lafontaine et al., 2017). Similar studies in penaeid shrimp could be employed to better understand the occurrence of reproductive dysfunction in domesticated shrimp stocks, in particular in *P. monodon*.

Recently, the combination of molecular weight cut-off during sample preparation (Wiśniewski et al., 2009) and unbiased MS analysis, called sequential window acquisition of all theoretical fragment ion spectra (SWATH), has provided a method to simultaneously detect and quantify thousands of proteins, without labeling or previous knowledge of the precursor peptide ions (Gillet et al., 2012; Chapman et al., 2014). As its name denotes, SWATH independently collects information based on time windows enabling detection of low-level ions rather than only selecting the most intense compounds in the sample as normally occurs with information dependent acquisition MS modes. Additionally, protein abundance quantified using SWATH has been shown to correlate with RNAseq transcriptomic data, as well as identify clinically relevant cancer biomarkers (Gao et al., 2017). SWATH information can also be combined with the more traditional multiple reaction monitoring mass spectrometry (MRM-MS) methods for protein quantification. MRM-MS is a double-level specificity tool to quantify specific targets in complex samples using existing knowledge of peptide m/z and fragment ion m/z

transitions (Liebler and Zimmerman, 2013). The abundance of targets of interest is often assessed by three peptides and three transitions. Quantification of proteins employing MRM-MS has been successfully applied for the quantification of allergens in penaeids including *P. monodon* and other crustaceans (Korte et al., 2016) as well as for the detection of antibiotics in shrimp muscle (Tyagi et al., 2008).

As for most “omics” research, data accessibility through publicly available repositories is key to maintaining research quality and building upon past work. Examples of this include ProteomeXchange (<http://www.proteomexchange.org>) or the PRoteomics IDentifications (PRIDE) database. These repositories contain curated MS spectral profiles that are standards compliant and facilitate peptide identification from existing proteomics data. High quality species-specific transcriptomes greatly improve peptide fragment and protein identification within proteomics workflows, even if that protein has no known function. While the penaeid research community may be dissuaded from publically releasing data due to the investment in data generation and perceived value, without concerted data sharing efforts, large proportions of data will become duplicated, while inevitably remaining poorly utilized due to incomplete characterisation of proteins and peptides.

For the purposes of this review, we have not included classical protein or gene function studies of small numbers of individual proteins through expression of recombinant proteins, pulldown assays, transient knockdown or western blotting. Complementary techniques such as targeted gene knockdown (RNAi and CRISPR), gene regulation (Methyl-seq, ChIPseq), protein-protein interactions (protein expression, pulldown assays), *in situ* hybridization, western blotting, cell culture and fluorescence microscopy can each aid in elucidating novel gene functions in shrimp. Overlaying and comparing results with those obtained in complementary metabolomics studies will also provide improved insight into functional response pathways to commercially relevant environmental stressors (Schock et al., 2013; Li et al., 2017) and diseases (Huynh et al., 2018).

DISSECTING AND EXPLOITING THE GENETIC VARIATION UNDERLYING PHENOTYPES

One of the primary drivers behind integrating genomics research into food production industries is to gain an understanding of the relationship between genetic variation and phenotypes of commercially important traits (Abdelrahman et al., 2017). In particular, through quantitative trait locus (QTL) mapping, and genome-wide association studies (GWAS), it may be possible to identify the number, location and effect size of genetic elements (i.e., genes, loci, and regions) that are linked with the observed phenotypic variation of a trait (Mackay et al., 2009). At a fundamental level, these studies provide novel insight into the genetic basis and architecture of traits, and when findings are compared across species, they allow the study of evolutionary processes underpinning trait characteristics. Furthermore, on an

applied level, identifying markers that are highly predictive for a superior or inferior phenotype is fundamental to improving the selection of elite individuals for breeding programs (Thorgaard et al., 2006). Recently, breeding programs for many agricultural species have utilized genome-wide estimated of performance (genomic breeding values), within the framework of genomic selection programs, to rapidly improve both simple and complex traits (Meuwissen et al., 2001, 2016). Similar efforts are to implement genomic selection underway for shrimp [*P. monodon* (Khatkar et al., 2017a), *L. vannamei* (Wang et al., 2016, 2017)], and promise to bring a substantial breeding improvements to the industry (Zenger et al., 2017).

Quantitative Trait Loci Mapping

As the availability of genomic resources for penaeid species has until recently been somewhat limited, the ability to undertake robust QTL studies has also been restricted. Early QTL studies examining penaeid production traits were limited in their power and ability to accurately isolate QTL effects (Alcivar-Warren et al., 2007; Du et al., 2010; Wang et al., 2012; Zhang et al., 2013). In particular, studies have been limited by incomplete genome coverage, along with the inclusion of insufficient individuals to ascertain precise QTL locations, or detect QTL of moderate or small effects. For instance, maps constructed and utilized by Wang et al. (2012) covered only 63% of the genome, and contained few markers (average of 5.1–5.4) per linkage group. At this resolution, it is possible to identify linkage groups that may contain QTL of interest, but isolating specific genes, or genomic regions, is generally not possible due to the large physical distance between markers (Alcivar-Warren et al., 2007; Andriantahina et al., 2013; Zhang et al., 2013). For instance Lyons et al. (2007), following on from research by Li Y. et al. (2006) and Li Z. et al. (2006), located a growth-related gene (EVOVL-like gene) through characterizing the region surrounding a growth-related QTL in *M. japonicus*. However, it remains unclear if this finding is indicative of the true genetic factor driving phenotypic variation, or if the association occurs through the co-localisation of a number of growth related genes in penaeids as suggested by Andriantahina et al. (2013).

Recently, a small number of QTL studies have been undertaken with vastly improved marker densities (3,959–4,626 SNPs), providing an improved ability to dissect the underlying genetic variation of traits. Robinson et al. (2014) utilized a SNP-based linkage map developed by Baranski et al. (2014) to investigate differences in survival time post-WSSV infection in *P. monodon*. Likewise, Yu et al. (2015) and Lu et al. (2016b) identified QTLs associated with growth in *L. vannamei*, and growth and high-temperature tolerance in *M. japonicus*. Unlike Robinson et al. (2014) both Yu et al. (2015) and Lu et al. (2016b) did not look to characterize the regions surrounding the identified QTLs, furthermore Lu et al. (2016b) did not compare findings with earlier growth QTLs observed by Li Y. et al. (2006). As such, while these studies provide the fundamental insight into the genetic architecture of these traits, they fall short of providing the opportunity to apply the findings to improve the selection of broodstock. Integration of dense linkage maps, transcriptome datasets, and genomic assemblies will facilitate the

characterisation of QTLs and further extend the utility of the research findings and should be required for publication of future penaeid QTL studies.

Genome-Wide Association Studies

In the quantitative study of traits, the commonly expected paradigm has largely shifted away from searching for the presence of a single gene or locus driving trait performance, to that where traits are more likely underpinned by tens to hundreds of genes, or regions of small effect, often dispersed across multiple chromosomes of the genome. Since it is possible to generate a high density of genome-wide markers, livestock, and crop species now predominately undertake a number of high resolution genome-wide association studies [GWAS; including regression models, linear mixed models, and a number of Bayesian mixed models (e.g., Bayes R, Bayesian Lasso, Bayes A)] in an effort to obtain a more accurate understanding of the number and distribution of trait associated loci along with their effect on commercially relevant traits.

To date, only two GWAS studies (Robinson et al., 2014; Khatkar et al., 2017b) have been published for penaeids. Both studies successfully identifying a number of loci significantly associated with sex confirming of observations of earlier QTL studies (Staelens et al., 2008) by concluding that sex in penaeids was likely a simple, genetically determined trait. Furthermore, Robinson et al. (2014) utilized three GWAS analysis approaches in tandem, to identify two markers associated with WSSV resistance (one marker with GRAMMAS and FASTA, one marker with QFAM). The limited identification of markers associated with this complex trait may stem from the limited samples size utilized (1,024 individuals) in the case the trait is highly polygenic, or as suggested previously by Hayes et al. (2010) there may be limited genetic variation underlying WSSV resistance of *P. monodon* (heritability < 0.001). While both markers were derived from RNA-Seq transcripts they were not further characterized in the study. Given the desire to selectively breed for disease resistance, further study of this trait, along with characterization of the markers associated should be completed to inform selective breeding programs.

In the second penaeid GWAS study, Khatkar et al. (2017b) found no significant associations of markers with the growth of *L. vannamei* after completing false-discovery-rate corrections for multiple tests. It is suggested by this study that no gene of large effect regulates growth, however, this finding contrasts those obtained earlier through QTL analysis by Yu et al. (2015) with the largest QTL found to explain 17.9% of phenotypic variation in *L. vannamei*. This may have resulted from the small number of phenotypic records used by Khatkar et al. (2017b), along with the use of genotypic information derived from pooled DNA samples. Furthermore additional individuals with genotype information and phenotypic records are required (~10,000) to undertake GWAS studies when including a large number of markers (5000–10,000), as each additional marker adds increasingly stringent multiple testing corrections (i.e., false-discovery-rate or FDR, Bonferroni corrections) that without sufficient detection power can eliminate otherwise potentially associated markers (Korte and Farlow, 2013). When investigating complex traits (those

underpinned by multiple genomic regions of small effect) the expected effect size of each individual marker is small (each driving <1% of phenotypic variation) and can be overlooked by QTL and GWAS analysis methods, particularly when a small number of individuals (e.g., <5000–10,000) are utilized (Visscher et al., 2014). To date, no Bayesian association approaches have been used in GWAS, with linear mixed models (ASReml, FASTA, GRAMMAS) or linear regression (QFAM) employed instead. Use of Bayesian approaches may decrease bias in estimates of SNP effect size, by allowing SNP effects to be modeled from a mixture of normal distributions [i.e., small, medium, and large effect sizes (Goddard et al., 2016)]. These Bayesian approaches have been used in initial exploration of genomic selection models of penaeids, but did not explore or comment specifically on any associations found with traits (Wang et al., 2017).

Use of Genome-Wide Trait Association Information in Penaeid Breeding

While GWAS can identify loci underpinning the performance of traits, it is unlikely single or only a few markers can be utilized in isolation to select superior animals (e.g., Marker Assisted Selection, MAS), as a limited number of loci may not capture sufficient contribution of the true trait performance. Successful attempts to integrate MAS in aquaculture are limited, and are increasingly being overlooked in favor of alternative methods, and as such are not discussed in detail in this review. Like those of livestock species, panels containing 50–100s of markers [e.g., for feed efficiency in beef cattle (Abo-Ismael et al., 2018)] can be developed to allow low cost prediction of animal performance, but can often be poor at capturing sufficient trait variation to make accurate breeding decisions. Ultimately, it is expected as additional genomic resources become available and are integrated (i.e., linkage maps used for QTL studies are integrated into genomic assemblies), it will be possible to characterize trait associated genomic regions to the resolution of each underlying gene or possible causative mutations (Quantitative Trait Nucleotides, or QTN). However, due to the high commercial value of robust predictive markers for advanced penaeid breeding, it is likely that some research findings will be restricted from public scrutiny (Benzie, 2009). As an example of this, a patent (Vuylsteke et al., 2012) has been obtained for a sex-specific genetic marker identified by Staelens et al. (2008), due to the potential economic value of use within the production process of monosex populations of *P. monodon*. The publication of full datasets is often restricted due to the considerable investment required to generate large phenotypic and genotype datasets, along with an increasing recognition of the competitive value of genomic-based information generated under industry based partnerships.

In the long term, it is likely that selection for polygenic traits will instead be enhanced through genomic selection (GS), where genome wide information is used to ascertain genomic relationships among individuals, and combined with the effect size of each trait associated SNP markers to generate a genomic estimate breeding value (gEBV). Unlike the low density trait specific panels available for livestock, by utilizing genome-wide

information GS allows recalculation and reconfiguration of the models underlying gEBVs, based on progressively updated genetic maps (e.g., linkage and LDU map). In addition, SNP effects are also continuously updated due to the observed LD decay between SNPs and trait locus across subsequent generations (Amaral et al., 2008). LD decay is expected to occur more rapidly given the relatively outbred status of penaeids in current and future breeding programs (Wang et al., 2016, 2017), when compared to most livestock species (e.g., cattle, sheep, horses, pigs, goats; Hall, 2016). Early modeling of genomic selection and prediction accuracy has been completed in *L. vannamei* (Wang et al., 2016, 2017), and indicates the ability to undertake genomic selection with a density of genome-wide markers (~3,200) currently available in commercial products (e.g., the Infinium ShrimpLD-24 v1.0 Bead Chip; Jones et al., 2017a). The effects of population stratification and marker number have been explored, however, prediction accuracy has only been estimated in internal subsets, rather than in subsequent novel generations. Mean reliability of the gEBVs (generated with BayesA, rrBLUP, and Bayesian LASSO) for body weight and body length, were 0.411 and 0.029 respectively (Wang et al., 2016), however, these values may decrease under empirical forward-prediction validation (Daetwyler et al., 2013). Additional samples [currently 200 individuals from 13 full-sib families (Wang et al., 2016)], as well as inclusion of half-sib families would provide more robust prediction capacity through improved estimates of SNP effects and division of maternal and paternal influences on trait performance. Further exploration of genomic selection breeding program designs (within- vs. between-family GS) and genotyping marker density should also be completed to ensure optimisation of genetic gain, in relation to cost of genotyping and inbreeding rates achieved (Lillehammer et al., 2013).

CONCLUSIONS

The field of penaeid genomics has seen considerable research applied over the past 30 years. Recent advances in “omics” research has continued to push the industry toward broad-scale uptake of genomic resources, particularly toward the implementation of genomic selection, and has begun to provide the understanding to overcome a number of the impediments faced by industry. However, the resources available currently require further development to ensure wide-scale use is feasible, and that studies being undertaken provide the most informative results possible to aid industry development. With this in mind, suggestions for areas of research focus are;

- Incorporation of error-corrected long-read sequencing information and optical mapping to complete long standing genome assembly projects.
- Integration of resources (linkage maps, physical maps, annotated transcriptome, characterized proteome data, and genome sequence) to form a complete and highly informative dataset that can be provided in a genome browser
- Development and public provision of cost effective genotyping resources for use in commercial scale genomic selection programs, as well as quantitative traits studies (e.g., GWAS).

Where possible consistent marker sets should be used to allow progressive improvement of datasets and comparison between studies.

- Development and release of high-density linkage maps (with minimal “binned” markers) and LDU maps to understand genome architecture, genome evolution, and inform genomic selection practices.
- Annotation of shrimp specific proteins and transcripts, as well as those conserved across taxa, to provide a complete understanding of the functional elements dictating shrimp performance. Cataloging of these should be maintained in public repositories to allow follow up studies (CRISPER or RNAi knockdown) to confirm role of identified targets of potential commercial importance.
- Undertaking of empirical forward-prediction genomic selection studies which incorporate traits of commercial importance (e.g., growth, disease resistance), to allow determination the most effective approach for the industry

breeding programs to improve productivity and overcome current constraints (i.e., disease prevalence).

AUTHOR CONTRIBUTIONS

All authors contributed significantly to the manuscript. The manuscript was prepared by JG. Manuscript structure was guided and first draft reviewed extensively by DBJ, DRJ, and KZ. Amendment and guidance was provided by HR, and RH. The section discussing proteomics was authored by NW. All authors reviewed the manuscript and approved its content.

FUNDING

The first author was supported by an Australian Postgraduate Award Scholarship. Funding for this work was provided from the Australian Research Council Industrial Transformation Research Program.

REFERENCES

- Abdelrahman, H., ElHady, M., Alcivar-Warren, A., Allen, S., Al-Tobasei, R., Bao, L., et al. (2017). Aquaculture genomics, genetics and breeding in the United States: current status, challenges, and priorities for future research. *BMC Genomics* 18:191. doi: 10.1186/s12864-017-3557-1
- Abo-Ismael, M. K., Lansink, N., Akanno, E., Karisa, B. K., Crowley, J. J., Moore, S. S., et al. (2018). Development and validation of a small SNP panel for feed efficiency in beef cattle. *J. Anim. Sci.* 96, 375–397. doi: 10.1093/jas/sky020
- Agrawl, R., and Narayan, J. (2015). Unraveling the impact of bioinformatics and omics in agriculture. *Int. J. Plant Biol. Res.* 3:1039.
- Alcivar-Warren, A., Meehan-Meola, D., Park, S. W., Xu, Z., Delaney, M., and Zuniga, G. (2007). ShrimpMap: A low-density, microsatellite-based linkage map of the Pacific whiteleg shrimp, *Litopenaeus vannamei*: Identification of sex-linked markers in linkage group 4. *J. Shellfish Res.* 26, 1259–1277. doi: 10.2983/0730-8000(2007)26:1259:SALMLM2.0.CO;2
- Alcivar-Warren, A., Meehan-Meola, D., Wang, Y., Guo, X., Zhou, L., Xiang, J., et al. (2006). Isolation and mapping of telomeric pentanucleotide (TAACC)n repeats of the Pacific whiteleg shrimp, *Penaeus vannamei*, using fluorescence in situ hybridization. *Mar. Biotechnol.* 8, 467–480. doi: 10.1007/s10126-005-6031-z
- Alcivar-Warren, A., Xu, Z., Meehan, D., Zuniga, G., and Bell, K. (2003). “Over 150 polymorphic microsatellite markers available for parentage testing, pedigree tracing, genetic diversity and genome mapping of the Pacific White Shrimp, *Penaeus (Litopenaeus) vannamei*,” in *Aquaculture America* (Louisville, KY).
- Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., et al. (2016). RAD Capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics* 202, 389–400. doi: 10.1534/genetics.115.183665
- Amaral, A. J., Megens, H.-J., Crooijmans, R. P. M. A., Heuven, H. C. M., and Groenen, M. A. M. (2008). Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* 179, 569–579. doi: 10.1534/genetics.107.084277
- Anderson, R. (2014). Development of a high density (600K) “Illumina Ovine SNP chip and its use to fine map the yellow fat locus,” in *Plant and Animal Genome XXII Conference* (San Diego, CA).
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* 17, 81–92. doi: 10.1038/nrg.2015.28
- Andriantahina, F., Liu, X., and Huang, H. (2013). Genetic map construction and quantitative trait locus (QTL) detection of growth-related traits in *Litopenaeus vannamei* for selective breeding applications. *PLoS ONE* 8:e75206. doi: 10.1371/journal.pone.0075206
- Aoki, T., Wang, H.-C., Unajak, S., Santos, M. D., Kondo, H., and Hirono, I. (2011). Microarray analyses of shrimp immune responses. *Mar. Biotechnol.* 13, 629–638. doi: 10.1007/s10126-010-9291-1
- Arnold, S. J., Coman, G. J., and Emerenciano, M. (2013). Constraints on seedstock production in eighth generation domesticated *Penaeus monodon* broodstock. *Aquaculture* 410, 95–100. doi: 10.1016/j.aquaculture.2013.06.023
- Baranski, M., Gopikrishna, G., Robinson, N. A., Katneni, V. K., Shekhar, M. S., Shanmugakarthik, J., et al. (2014). The development of a high density linkage map for black tiger shrimp (*Penaeus monodon*) based on cSNPs. *PLoS ONE* 9:e85413. doi: 10.1371/journal.pone.0085413
- Bartlett, T. C., Cuthbertson, B. J., Shepard, E. F., Chapman, R. W., Gross, P. S., and Warr, G. W. (2002). Crustins, homologues of an 11.5-kDa antibacterial peptide, from two species of penaeid shrimp, *Litopenaeus vannamei* and *Litopenaeus setiferus*. *Mar. Biotechnol.* 4, 278–293. doi: 10.1007/s10126-002-0020-2
- Benzie, J. A. (1998). Penaeid genetics and biotechnology. *Aquaculture* 164, 23–47. doi: 10.1016/S0044-8486(98)00175-6
- Benzie, J. A. (2009). Use and exchange of genetic resources of penaeid shrimps for food and aquaculture. *Rev. Aquacult.* 1, 232–250. doi: 10.1111/j.1753-5131.2009.01018.x
- Bernd, G., Unmack, P. J., Berry, O. F., and Arthur, G. (2018). dartr: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol. Ecol. Resour.* 18, 691–699. doi: 10.1111/1755-0998.12745
- Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., et al. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.* 5:3657. doi: 10.1038/ncomms4657
- Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., et al. (2017). Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. *Nat. Genet.* 49, 643–650. doi: 10.1038/ng.3802
- Bourchookarn, A., Havanapan, P. O., Thongboonkerd, V., and Krittanai, C. (2008). Proteomic analysis of altered proteins in lymphoid organ of yellow head virus infected *Penaeus monodon*. *Bba Proteins Proteomics* 1784, 504–511. doi: 10.1016/j.bbapap.2007.12.006
- Brozynska, M., Copetti, D., Furtado, A., Wing, R. A., Crayn, D., Fox, G., et al. (2016). Sequencing of Australian wild rice genomes reveals ancestral relationships with domesticated rice. *Plant Biotechnol. J.* 15, 765–777. doi: 10.1111/pbi.12674
- Callaghan, T. R., Degnan, B. M., and Sellars, M. J. (2010). Expression of sex and reproduction-related genes in *Marsupenaeus japonicus*. *Mar. Biotechnol.* 12, 664–677. doi: 10.1007/s10126-009-9254-6
- Campos-Montes, G. R., Montaldo, H. H., Armenta-Córdova, M., Martínez-Ortega, A., Caballero-Zamora, A., and Castillo-Juárez, H. (2017). Incorporation of tail weight and tail percentage at harvest size in selection programs for the Pacific white shrimp *Penaeus (Litopenaeus) vannamei*. *Aquaculture* 468, 293–296. doi: 10.1016/j.aquaculture.2016.10.034

- improved annotation of the whiteleg shrimp (*Litopenaeus vannamei*), a dominant crustacean in global seafood mariculture. *Sci. Rep.* 4:7081. doi: 10.1038/srep07081
- Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., et al. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* 11:O111.01671. doi: 10.1074/mcp.O111.016717
- Gjedrem, T., and Rye, M. (2016). Selection response in fish and shellfish: a review. *Rev. Aquacult.* 10, 168–179. doi: 10.1111/raq.12154
- Gjedrem, T., Robinson, N., and Rye, M. (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. *Aquaculture* 350, 117–129. doi: 10.1016/j.aquaculture.2012.04.008
- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., et al. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* 9:e90346. doi: 10.1371/journal.pone.0090346
- Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136, 245–257. doi: 10.1007/s10709-008-9308-0
- Goddard, M. E., Kemper, K. E., MacLeod, I. M., Chamberlain, A. J., and Hayes, B. J. (2016). Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proc. R. Soc. B Biol. Sci.* 283:20160569. doi: 10.1098/rspb.2016.0569
- Gonçalves, M. M., de Almeida Regitano, L. C., Dami, C., Salgado, C. C., de Freitas, P. D., Teixeira, A. K. G., et al. (2014). Inheritance of AFLP markers and genetic linkage analysis in two full-sib families of the marine shrimp *Litopenaeus vannamei* (Crustacea, Decapoda). *Adv. Biosci. Biotechnol.* 5, 273–281. doi: 10.4236/abb.2014.53034
- Goodall, J. (2017). *Improving the Reproductive Performance of Domesticated Giant Tiger Shrimp, Penaeus monodon*. PhD Thesis. The University of Queensland.
- Gross, P., Bartlett, T., Browdy, C., Chapman, R., and Warr, G. (2001). Immune gene discovery by expressed sequence tag analysis of hemocytes and hepatopancreas in the Pacific White Shrimp, *Litopenaeus vannamei*, and the Atlantic White Shrimp, *L. setiferus*. *Dev. Comp. Immunol.* 25, 565–577. doi: 10.1016/S0145-305X(01)00018-0
- Guo, H., Ye, C.-X., Wang, A.-L., Xian, J.-A., Liao, S.-A., Miao, Y.-T., et al. (2013). Transcriptome analysis of the Pacific white shrimp *Litopenaeus vannamei* exposed to nitrite by RNA-seq. *Fish Shellfish Immunol.* 35, 2008–2016. doi: 10.1016/j.fsi.2013.09.019
- Guppy, J. L., Jones, D. B., Steinig, E. J., Khatkar, M. S., Jerry, D. R., and Zenger, K. R. (2018). “Development of a highly informative and cost effective genotyping panel for advanced black tiger prawn breeding programs,” in *International Symposium on Genetics in Aquaculture* (Cairns, QLD).
- Habier, D., Fernando, R. L., and Dekkers, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177, 2389–2397. doi: 10.1534/genetics.107.081190
- Hall, S. J. G. (2016). Effective population sizes in cattle, sheep, horses, pigs and goats estimated from census and herdbook data. *Animal* 10, 1778–1785. doi: 10.1017/S1751731116000914
- Hayes, B. J., Gitterle, T., Gopikrishna, G., Gopal, C., Krishna, G., Jahageerda, S., et al. (2010). Limited evidence for genetic variation for resistance to the white spot syndrome virus in Indian populations of *Penaeus monodon*. *Aquac. Res.* 41, e872–e877. doi: 10.1111/j.1365-2109.2010.02611.x
- Hayes, B. J., Visscher, P. M., and Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* 91, 47–60. doi: 10.1017/S0016672308009981
- He, Y., Du, Y., Li, J., Liu, P., Wang, Q., and Li, Z. (2015). Analysis of DNA methylation in different tissues of *Fenneropenaeus chinensis* from the wild population and Huanghai No. 1. *Acta Oceanol. Sin.* 34, 175–180. doi: 10.1007/s13131-015-0765-x
- Henshall, J. M., Dierens, L., and Sellars, M. J. (2014). Quantitative analysis of low-density SNP data for parentage assignment and estimation of family contributions to pooled samples. *Genet. Select. Evol.* 46:51. doi: 10.1186/s12711-014-0051-y
- Henshall, J. M., Hawken, R. J., Dominik, S., and Barendse, W. (2012). Estimating the effect of SNP genotype on quantitative traits from pooled DNA samples. *Genet. Select. Evol.* 44:12. doi: 10.1186/1297-9686-44-12
- Henson, J., Tischler, G., and Ning, Z. (2012). Next-generation sequencing and large genome assemblies. *Pharmacogenomics* 13, 901–915. doi: 10.2217/pgs.12.72
- Holman, L. E., Garcia de la serrana, D., Onoufriou, A., Hillestad, B., and Johnston, I. A. (2017). A workflow used to design low density SNP panels for parentage assignment and traceability in aquaculture species and its validation in Atlantic salmon. *Aquaculture* 476, 59–64. doi: 10.1016/j.aquaculture.2017.04.001
- Houston, R. D., Taggart, J. B., Cézard, T., Bekaert, M., Lowe, N. R., Downing, A., et al. (2014). Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics* 15:90. doi: 10.1186/1471-2164-15-90
- Howe, K., and Wood, J. M. D. (2015). Using optical mapping data for the improvement of vertebrate genome assemblies. *Gigascience* 4:10. doi: 10.1186/s13742-015-0052-y
- Hozé, C., Fouilloux, M.-N., Venot, E., Guillaume, F., Dassonneville, R., Fritz, S., et al. (2013). High-density marker imputation accuracy in sixteen French cattle breeds. *Genet. Select. Evol.* 45:33. doi: 10.1186/1297-9686-45-33
- Huan, P., Zhang, X., Li, F., Zhao, C., Zhang, C., and Xiang, J. (2010). Chromosomal localization of 5S rDNA in Chinese shrimp (*Fenneropenaeus chinensis*): a chromosome-specific marker for chromosome identification. *Chinese J. Oceanol. Limnol.* 28, 233–238. doi: 10.1007/s00343-010-9278-9
- Huang, S.-W., Lin, Y.-Y., You, E.-M., Liu, T.-T., Shu, H.-Y., Wu, K.-M., et al. (2011). Fosmid library end sequencing reveals a rarely known genome structure of marine shrimp *Penaeus monodon*. *BMC Genomics* 12:242. doi: 10.1186/1471-2164-12-242
- Huerlimann, R., Wade, N. M., Gordon, L., Montenegro, J. D., Goodall, J., McWilliam, S., et al. (2018). *De novo* assembly, characterization, functional annotation and expression patterns of the black tiger shrimp (*Penaeus monodon*) transcriptome. *bioRxiv*. doi: 10.1101/280420
- Hulse-Kemp, A. M., Maheshwari, S., Stoffel, K., Hill, T. A., Jaffe, D., Williams, S. R., et al. (2018). Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Hortic. Res.* 5, 4. doi: 10.1038/s41438-017-0011-0
- Huynh, T.-G., Cheng, A.-C., Chi, C.-C., Chiu, K.-H., and Liu, C.-H. (2018). A synbiotic improves the immunity of white shrimp, *Litopenaeus vannamei*: metabolomic analysis reveal compelling evidence. *Fish Shellfish Immunol.* 79, 284–293. doi: 10.1016/j.fsi.2018.05.031
- International Chicken Genome Sequencing, C. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695. doi: 10.1038/nature03154
- Janssen, K., Chavanne, H., Berentsen, P., and Komen, H. (2017). Impact of selective breeding on European aquaculture. *Aquaculture* 472, 8–16. doi: 10.1016/j.aquaculture.2016.03.012
- Janssen, K., Saatkamp, H., and Komen, H. (2018). Cost-benefit analysis of aquaculture breeding programs. *Genet. Select. Evol.* 50:2. doi: 10.1017/9781108660624
- Jerry, D. R., Evans, B. S., Kenway, M., and Wilson, K. (2006). Development of a microsatellite DNA parentage marker suite for black tiger shrimp *Penaeus monodon*. *Aquaculture* 255, 542–547. doi: 10.1016/j.aquaculture.2005.11.026
- Jerry, D. R., Preston, N. P., Crocos, P. J., Keys, S., Meadows, J. R., and Li, Y. (2004). Parentage determination of Kuruma shrimp *Penaeus (Marsupenaeus) japonicus* using microsatellite markers (Bate). *Aquaculture* 235, 237–247. doi: 10.1016/j.aquaculture.2004.01.019
- Jiang, H., Li, F., Xie, Y., Huang, B., Zhang, J., Zhang, J., et al. (2009). Comparative proteomic profiles of the hepatopancreas in *Fenneropenaeus chinensis* response to hypoxic stress. *Proteomics* 9, 3353–3367. doi: 10.1002/pmic.200800518
- Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., et al. (2017). Improved maize reference genome with single-molecule technologies. *Nature* 546, 524. doi: 10.1038/nature22971
- Jones, D. B., Jerry, D. R., Khatkar, M. S., Raadsma, H. W., Steen, H. V. D., Prochaska, J., et al. (2017a). A comparative integrated gene-based linkage and locus ordering by linkage disequilibrium map for the Pacific white shrimp, *Litopenaeus vannamei*. *Sci. Rep.* 7:10360. doi: 10.1038/s41598-017-10515-7
- Jones, D. B., Zenger, K. R., Khatkar, M. S., Raadsma, H. W., Steen, H. A. M. V. D., Prochaska, J., et al. (2017b). “Development of a low-density commercial genotyping array for the white legged shrimp, *Litopenaeus vannamei*,” in *AAABG*, Edited by Genetics AftAoABa (Townsville, QLD).
- Karoonuthaisiri, N., Sittikankeaw, K., Preechaphol, R., Kalachikov, S., Wongsurawat, T., Uawisetwathana, U., et al. (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. *Comp. Biochem. Physiol. Part D Genomics Proteomics* 4, 90–99. doi: 10.1016/j.cbd.2008.11.003
- Khatkar, M. S. (2017). Genomic selection in aquaculture breeding programs. *Bioinform. Aquacult.* 2017, 380–391. doi: 10.1002/9781118782392.ch21
- Khatkar, M., Coman, G., Thomson, P., and Raadsma, H. (2017a). “Comparison of different breeding design options for long term genetic gain and diversity in aquaculture species,” in *Proc Assoc Advmt Anim Breed Genet* (Townsville, QLD), 449–452.
- Khatkar, M., Zenger, K., Jones, D., Prochasca, J., van der Steen, H., Jerry, D., et al. (2017b). “Quantitative genomic analyses in the pacific whiteleg shrimp *Litopenaeus vannamei*,” in *Association for the Advancement of Animal Breeding and Genetics* (Townsville, QLD).
- Kim, M., Rhee, J.-S., Kim, T., Lee, J., Choi, A.-Y., Choi, B.-S., et al. (2017). Alternative splicing profile and sex-preferential gene expression in the female and male Pacific Abalone *Haliotis discus hannai*. *Genes* 8:99. doi: 10.3390/genes8030099
- Kinghorn, B., Bastiaansen, J., Ciobanu, D., and van der Steen, H. (2010). Quantitative genotyping to estimate genetic contributions to pooled samples and genetic merit of the contributing entities. *Acta Agric. Scand. Sect. A* 60, 3–12. doi: 10.1080/09064701003801922
- Knibb, W., Whatmore, P., Lamont, R., Quinn, J., Powell, D., Elizur, A., et al. (2014). Can genetic diversity be maintained in long term mass selected populations without pedigree information? — A case study using banana shrimp *Fenneropenaeus merguensis*. *Aquaculture* 428, 71–78. doi: 10.1016/j.aquaculture.2014.02.026
- Kong, J., and Gao, H. (2005). Analysis of tandem repeats in the genome of Chinese shrimp *Fenneropenaeus chinensis*. *Chinese Sci. Bull.* 50, 1462–1469. doi: 10.1360/982005-270
- Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9:29. doi: 10.1186/1746-4811-9-29
- Korte, R., Monneuse, J.-M., Gemrot, E., Metton, I., Humpf, H.-U., and Brockmeyer, J. (2016). New high-performance liquid chromatography coupled mass spectrometry method for the detection of lobster and shrimp allergens in food samples via multiple reaction monitoring and multiple reaction monitoring cubed. *J. Agric. Food Chem.* 64, 6219–6227. doi: 10.1021/acs.jafc.6b02620
- Koyama, T., Asakawa, S., Katagiri, T., Shimizu, A., Fagutao, F. F., Mavichak, R., et al. (2010). Hyper-expansion of large DNA segments in the genome of kuruma shrimp, *Marsupenaeus japonicus*. *BMC Genomics* 11:141. doi: 10.1186/1471-2164-11-141
- Kulkarni, A. D., Kiron, V., Rombout, J. H. W. M., Brinchmann, M. F., Fernandes, J. M. O., Sudheer, N. S., et al. (2014). Protein profiling in the gut of *Penaeus monodon* gavaged with oral WSSV-vaccines and live white spot syndrome virus. *Proteomics* 14, 1660–1673. doi: 10.1002/pmic.201300405
- Lafontaine, A., Baiwir, D., Joaquim-Justo, C., De Pauw, E., Lemoine, S., Boulangé-Lecomte, C., et al. (2017). Proteomic response of *Macrobrachium rosenbergii* hepatopancreas exposed to chlordecone: Identification of endocrine disruption biomarkers? *Ecotoxicol. Environ. Saf.* 141, 306–314. doi: 10.1016/j.ecoenv.2017.03.043
- Lee, S. G., and Mykles, D. L. (2006). Proteomics and signal transduction in the crustacean molting gland. *Integr. Comp. Biol.* 46, 965–977. doi: 10.1093/icb/icl047
- Leelatanawit, R., Klinbunga, S., Aoki, T., Hirono, I., Valyasevi, R., and Menasveta, P. (2008). Suppression subtractive hybridization (SSH) for isolation and characterization of genes related to testicular development in the giant tiger shrimp *Penaeus monodon*. *BMB Rep.* 41, 796–802. doi: 10.5483/BMBRep.2008.41.11.796
- Leelatanawit, R., Sittikankeaw, K., Yocawibun, P., Klinbunga, S., Roytrakul, S., Aoki, T., et al. (2009). Identification, characterization and expression of sex-related genes in testes of the giant tiger shrimp *Penaeus monodon*. *Comp. Biochem. Physiol. Part A* 152, 66–76. doi: 10.1016/j.cbpa.2008.09.004
- Leelatanawit, R., Uawisetwathana, U., Klanhui, A., Khudet, J., Phomklad, S., Wongtriphop, S., et al. (2017). Transcriptomic ANALYSIS of Male Black Tiger Shrimp (*Penaeus monodon*) after polychaete feeding to enhance testicular maturation. *Mar. Biotechnol.* 19, 125–135. doi: 10.1007/s10126-017-9738-8
- Leelatanawit, R., Uawisetwathana, U., Klinbunga, S., and Karoonuthaisiri, N. (2011). A cDNA microarray, UniShrimpChip, for identification of genes relevant to testicular development in the black tiger shrimp (*Penaeus monodon*). *BMC Mol. Biol.* 12:15. doi: 10.1186/1471-2199-12-15
- Leu, J.-H., Chang, C.-C., Wu, J.-L., Hsu, C.-W., Hirono, I., Aoki, T., et al. (2007). Comparative analysis of differentially expressed genes in normal and white spot syndrome virus infected *Penaeus monodon*. *BMC Genomics* 8:120. doi: 10.1186/1471-2164-8-120
- Leu, J.-H., Chen, S.-H., Wang, Y.-B., Chen, Y.-C., Su, S.-Y., Lin, C.-Y., et al. (2011). A review of the major penaeid shrimp EST studies and the construction of a shrimp transcriptome database based on the ESTs from four penaeid shrimp. *Mar. Biotechnol.* 13, 608–621. doi: 10.1007/s10126-010-9286-y
- Li, C., Liu, X., Liu, B., Ma, B., Liu, F., Liu, G., et al. (2018). Draft genome of the Peruvian scallop *Argopecten purpuratus*. *Gigascience* 2018:giy031. doi: 10.1093/gigascience/giy031
- Li, C., Weng, S., Chen, Y., Yu, X., Lü, L., Zhang, H., et al. (2012). Analysis of *Litopenaeus vannamei* transcriptome using the next-generation DNA sequencing technique. *PLoS ONE* 7:e47442. doi: 10.1371/journal.pone.0047442
- Li, S., Zhang, X., Sun, Z., Li, F., and Xiang, J. (2013). Transcriptome analysis on Chinese shrimp *Fenneropenaeus chinensis* during WSSV acute infection. *PLoS ONE* 8:e58627. doi: 10.1371/journal.pone.0058627
- Li, T., Li, E., Suo, Y., Xu, Z., Jia, Y., Qin, J. G., et al. (2017). Energy metabolism and metabolomics response of Pacific white shrimp *Litopenaeus vannamei* to sulfide toxicity. *Aquat. Toxicol.* 183, 28–37. doi: 10.1016/j.aquatox.2016.12.010
- Li, W., Tang, X. Q., Xing, J., Sheng, X. Z., and Zhan, W. B. (2014). Proteomic analysis of differentially expressed proteins in *Fenneropenaeus chinensis* hemocytes upon white spot syndrome virus infection. *PLoS ONE* 9:e8996. doi: 10.1371/journal.pone.0089962
- Li, Y., Byrne, K., Miggiano, E., Whan, V., Moore, S., Keys, S., et al. (2003). Genetic mapping of the kuruma prawn *Penaeus japonicus* using AFLP markers. *Aquaculture* 219:143–156. doi: 10.1016/S0044-8486(02)00355-1
- Li, Y., Dierens, L., Byrne, K., Miggiano, E., Lehnert, S., Preston, N., et al. (2006). QTL detection of production traits for the Kuruma prawn *Penaeus japonicus* (Bate) using AFLP markers. *Aquaculture* 258, 198–210. doi: 10.1016/j.aquaculture.2006.04.027
- Li, Y., Jiang, S., Huang, J., Ma, Z., and Zhou, F. (2016). Growth and survival variations of *Penaeus monodon* from six breeding families. *Int. J. Aquacult.* 5, 1–9. doi: 10.5376/ija.2015.05.0039
- Li, Y., Willer, C., Sanna, S., and Abecasis, G. (2009). Genotype Imputation. *Annu. Rev. Genomics Hum. Genet.* 10, 387–406. doi: 10.1146/annurev.genom.9.081307.164242
- Li, Z., Li, J., Wang, Q., He, Y., and Liu, P. (2006). AFLP-based genetic linkage map of marine shrimp *Penaeus (Fenneropenaeus) chinensis*. *Aquaculture* 261, 463–472. doi: 10.1016/j.aquaculture.2006.07.002
- Liebler, D. C., and Zimmerman, L. J. (2013). Targeted quantitation of proteins by mass spectrometry. *Biochemistry* 52, 3797–3806. doi: 10.1021/bi400110b
- Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., et al. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* 533, 500–505. doi: 10.1038/nature17164
- Lillehammer, M., Meuwissen, T. H. E., and Sonesson, A. K. (2013). A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. *Genet. Select. Evol.* 45:39. doi: 10.1186/1297-9686-45-39
- Lio-Po, G. D., and Leño, E. M. (2016). “Important diseases of Penaeid shrimps,” in *Progress of Shrimp and Prawn Aquaculture in the World*. Keelung: National Taiwan Ocean University, 269–315.
- Liu, B., Wang, Q., Li, J., Liu, P., and He, Y. (2010). A genetic linkage map of marine shrimp *Penaeus (Fenneropenaeus) chinensis* based on AFLP, SSR, and RAPD markers. *Chinese J. Oceanol. Limnol.* 28, 815–825. doi: 10.1007/s00343-010-9915-3
- Liu, S., Sun, L., Li, Y., Sun, F., Jiang, Y., Zhang, Y., et al. (2014). Development of the catfish 250K SNP array for genome-wide association studies. *BMC Res. Notes* 7:135. doi: 10.1186/1756-0500-7-135
- Liu, Z. J. (2007). Fish genomics and analytical genetic technologies, with examples of their potential applications in management of fish genetic resources. *America* 5, 145–179.
- Liu, Z., Liu, S., Yao, J., Bao, L., Zhang, J., Li, Y., et al. (2016). The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nat. Commun.* 7:11757. doi: 10.1038/ncomms11757

- Lu, X., Kong, J., Luan, S., Dai, P., Meng, X., Cao, B., et al. (2016a). Transcriptome analysis of the hepatopancreas in the Pacific White Shrimp (*Litopenaeus vannamei*) under acute ammonia stress. *PLoS ONE* 11:e0164396. doi: 10.1371/journal.pone.0164396
- Lu, X., Luan, S., Hu, L. Y., Mao, Y., Tao, Y., Zhong, S. P., et al. (2016b). High-resolution genetic linkage mapping, high-temperature tolerance and growth-related quantitative trait locus (QTL) identification in *Marsupenaeus japonicus*. *Mol. Genet. Genomics* 291, 1391–1405. doi: 10.1007/s00438-016-1192-1
- Lu, X., Luan, S., Kong, J., Hu, L., Mao, Y., and Zhong, S. (2017). Genome-wide mining, characterization, and development of microsatellite markers in *Marsupenaeus japonicus* by genome survey sequencing. *Chinese J. Oceanol. Limnol.* 35, 203–214. doi: 10.1007/s00343-016-5250-7
- Lyons, R., Dierens, L., Tan, S., Preston, N., and Li, Y. (2007). Characterization of AFLP markers associated with growth in the Kuruma prawn, *Marsupenaeus japonicus*, and identification of a candidate gene. *Mar. Biotechnol.* 9, 712–721. doi: 10.1007/s10126-007-9022-4
- Ma, M., Wang, J., Chen, R., and Li, L. (2009). Expanding the Crustacean neuropeptidome using a multifaceted mass spectrometric approach. *J. Proteome Res.* 8, 2426–2437. doi: 10.1021/pr801047v
- Mackay, T. F., Stone, E. A., and Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10, 565–577. doi: 10.1038/nrg2612
- Maneeruttanarungroj, C., Pongsomboon, S., Wuthisuthimethavee, S., Klinbunga, S., Wilson, K. J., Swan, J., et al. (2006). Development of polymorphic expressed sequence tag-derived microsatellites for the extension of the genetic linkage map of the black tiger shrimp (*Penaeus monodon*). *Anim. Genet.* 37, 363–368. doi: 10.1111/j.1365-2052.2006.01493.x
- Maralit, B. A., Ventolero, M. F. H., Maningas, M. B. B., Amar, E. C., and Santos, M. D. (2014). "Subtracted transcriptome profile of tiger shrimp (*Penaeus monodon*) that survived WSSV challenge," in *Dataset Papers in Science*. doi: 10.1155/2014/807806
- Marsden, G., Richardson, N., Mather, P., and Knibb, W. (2013). Reproductive behavioural differences between wild-caught and pond-reared *Penaeus monodon* prawn broodstock. *Aquaculture* 402, 141–145. doi: 10.1016/j.aquaculture.2013.03.019
- Mathew, B., Léon, J., and Sillanpää, M. J. (2018). A novel linkage-disequilibrium corrected genomic relationship matrix for SNP-heritability estimation and genomic prediction. *Heredity* 120:356. doi: 10.1038/s41437-017-0023-4
- Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., et al. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4:e5350. doi: 10.1371/journal.pone.0005350
- McEwan, J. C., and Consortium (2009). "TISG: Skim sequencing the sheep genome, and SNP discovery, using Roche 454 FLX sequencing," in *Plant and Animal Genomes XVII Conference* (San Diego, CA).
- Meuwissen, T., Hayes, B., and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819.
- Meuwissen, T., Hayes, B., and Goddard, M. (2016). Genomic selection: a paradigm shift in animal breeding. *Anim. Front.* 6, 6–14. doi: 10.2527/af.2016-0002
- Ming, R., VanBuren, R., Wai, C. M., Tang, H., Schatz, M. C., Bowers, J. E., et al. (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* 47, 1435–1442. doi: 10.1038/ng.3435
- Montenegro, J. D., Huerlimann, R., Simmering, K., Gordon, L., Tinning, M., Jabbari, J. S., et al. (2018). "Hybrid assembly of the black tiger shrimp genome (*Penaeus monodon*)," in *International Symposium on Genetics in Aquaculture* (Cairns, QLD).
- Moss, D. R., Arce, S. M., Otsu, C. A., Doyle, R. W., and Moss, S. M. (2007). Effects of inbreeding on survival and growth of Pacific white shrimp *Penaeus (Litopenaeus) vannamei*. *Aquaculture* 272, S30–S37. doi: 10.1016/j.aquaculture.2007.08.014
- Mouse Genome Sequencing, C. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520. doi: 10.1038/nature01262
- Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., et al. (2000). A whole-genome assembly of *Drosophila*. *Science* 287, 2196–2204. doi: 10.1126/science.287.5461.2196
- Nguyen, C., Nguyen, T. G., Van Nguyen, L., Pham, H. Q., Nguyen, T. H., Pham, H. T., et al. (2016). De novo assembly and transcriptome characterization of major growth-related genes in various tissues of *Penaeus monodon*. *Aquaculture* 464, 545–553. doi: 10.1016/j.aquaculture.2016.08.003
- Niedzicka, M., Fijarczyk, A., Dudek, K., Stuglik, M., and Babik, W. (2016). Molecular inversion probes for targeted resequencing in non-model organisms. *Sci. Rep.* 6:24051. doi: 10.1038/srep24051
- Nielsen, H. M., Sonesson, A. K., and Meuwissen, T. H. E. (2011). Optimum contribution selection using traditional best linear unbiased prediction and genomic breeding values in aquaculture breeding schemes. *J. Anim. Sci.* 89, 630–638. doi: 10.2527/jas.2009-2731
- Odde, S., Zelig, A., and Kaplan, N. (2018). Three invariant Hi-C interaction patterns: applications to genome assembly. *Methods* 142, 89–99. doi: 10.1016/j.ymeth.2018.04.013
- O'Leary, N. A., Trent, H. F., Robalino, J., Peck, M. E., McKillen, D. J., and Gross, P. S. (2006). Analysis of multiple tissue-specific cDNA libraries from the Pacific whiteleg shrimp, *Litopenaeus vannamei*. *Integr. Comp. Biol.* 46, 931–939. doi: 10.1093/icb/icl006
- Peng, J., Wei, P., Zhang, B., Zhao, Y., Zeng, D., Chen, X., et al. (2015). Gonadal transcriptomic analysis and differentially expressed genes in the testis and ovary of the Pacific white shrimp (*Litopenaeus vannamei*). *BMC Genomics* 16:1006. doi: 10.1186/s12864-015-2219-4
- Pérez, F., Erazo, C., Zhinaula, M., Volckaert, F., and Calderón, J. (2004). A sex-specific linkage map of the white shrimp *Penaeus (Litopenaeus) vannamei* based on AFLP markers. *Aquaculture* 242, 105–118. doi: 10.1016/j.aquaculture.2004.09.002
- Pérez-de-Castro, A. M., Vilanova, S., Cañizares, J., Pascual, L., Blanca, J. M., Diez, M. J., et al. (2012). Application of genomic tools in plant breeding. *Curr. Genomics* 13, 179–195. doi: 10.2174/138920212800543084
- Phillippy, A. M. (2017). New advances in sequence assembly. *Genome Res.* 27, xi–xiii. doi: 10.1101/gr.223057.117
- Pongsomboon, S., Tang, S., Boonda, S., Aoki, T., Hirono, I., and Tassanakajon, A. (2011). A cDNA microarray approach for analyzing transcriptional changes in *Penaeus monodon* after infection by pathogens. *Fish Shellfish Immunol.* 30, 439–446. doi: 10.1016/j.fsi.2010.10.015
- Pongsomboon, S., Tang, S., Boonda, S., Aoki, T., Hirono, I., Yasuie, M., et al. (2008a). Differentially expressed genes in *Penaeus monodon* hemocytes following infection with yellow head virus. *BMB Rep.* 41, 670–677. doi: 10.5483/BMBRep.2008.41.9.670
- Pongsomboon, S., Wongpanya, R., Tang, S., Chalorsrikul, A., and Tassanakajon, A. (2008b). Abundantly expressed transcripts in the lymphoid organ of the black tiger shrimp, *Penaeus monodon*, and their implication in immune function. *Fish Shellfish Immunol.* 25, 485–493. doi: 10.1016/j.fsi.2008.07.010
- Powell, D., Knibb, W., Nguyen, N. H., and Elizur, A. (2016). Transcriptomic Profiling of Banana Shrimp *Fenneropenaeus merguensis* with Differing Levels of Viral Load. *Integr. Comp. Biol.* 56, 1131–1143. doi: 10.1093/icb/icw029
- Powell, D., Knibb, W., Remilton, C., and Elizur, A. (2015). De-novo transcriptome analysis of genes associated with reproduction and development. *Mar. Genomics* 22, 71–78. doi: 10.1016/j.margen.2015.04.006
- Preechaphol, R., Klinbunga, S., Khamnamtong, B., and Menasveta, P. (2010). Isolation and characterization of genes functionally involved in ovarian development of the giant tiger shrimp *Penaeus monodon* by suppression subtractive hybridization (SSH). *Genet. Mol. Biol.* 33, 676–685. doi: 10.1590/S1415-4752010000400014
- Preechaphol, R., Leelatanawit, R., Sittikankeaw, K., Klinbunga, S., Khamnamtong, B., Puanglarp, N., et al. (2007). Expressed sequence tag analysis for identification and characterization of sex-related genes in the giant tiger shrimp *Penaeus monodon*. *J. Biochem. Mol. Biol.* 40, 501–510. doi: 10.5483/BMBRep.2007.40.4.501
- Pryce, J. E., Johnston, J., Hayes, B. J., Sahana, G., Weigel, K. A., McParland, S., et al. (2014). Imputation of genotypes from low density (50,000 markers) to high density (700,000 markers) of cows from research herds in Europe, North America, and Australasia using 2 reference populations. *J. Dairy Sci.* 97, 1799–1811. doi: 10.3168/jds.2013-7368
- Przytycki, L. P., and Gabaldón, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* 44:e113. doi: 10.1093/nar/gkw294
- Qiao, J., Du, Z. H., Zhang, Y. L., Du, H., Guo, L. L., Zhong, M. Q., et al. (2011). Proteomic identification of the related immune-enhancing proteins in shrimp

- Litopenaeus vannamei* stimulated with vitamin C and Chinese herbs. *Fish Shellfish Immunol.* 31, 736–745. doi: 10.1016/j.fsi.2011.07.005
- Rattanaoijpong, T., Wang, H. C., Lo, C. F., and Flegel, T. W. (2007). Analysis of differently expressed proteins and transcripts in gills of *Penaeus vannamei* after yellow head virus infection. *Proteomics* 7, 3809–3814. doi: 10.1002/pmic.200700202
- Robalino, J., Almeida, J. S., McKillen, D., Colglazier, J., Trent, H. F., Chen, Y. A., et al. (2007). Insights into the immune transcriptome of the shrimp *Litopenaeus vannamei*: tissue-specific expression profiles and transcriptomic responses to immune challenge. *Physiol. Genomics* 29, 44–56. doi: 10.1152/physiolgenomics.00165.2006
- Robalino, J., Carnegie, R. B., O'Leary, N., Ouvry-Patat, S. A., de la Vega, E., Prior, S., et al. (2009). Contributions of functional genomics and proteomics to the study of immune responses in the Pacific white leg shrimp *Litopenaeus vannamei*. *Vet. Immunol. Immunopathol.* 128, 110–118. doi: 10.1016/j.vetimm.2008.10.329
- Robinson, N. A., Gopikrishna, G., Baranski, M., Katneni, V. K., Shekhar, M. S., Shanmugakarthik, J., et al. (2014). QTL for white spot syndrome virus resistance and the sex-determining locus in the Indian black tiger shrimp (*Penaeus monodon*). *BMC Genomics* 15:731. doi: 10.1186/1471-2164-15-731
- Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., and Houston, R. (2017). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev. Aquacult.* 1–13. doi: 10.1111/raq.12193
- Rodrigues, P. M., Silva, T. S., Dias, J., and Jessen, F. (2012). PROTEOMICS in aquaculture: applications and trends. *J. Proteomics* 75, 4325–4345. doi: 10.1016/j.jprot.2012.03.042
- Rojtinnakorn, J., Hirono, I., Itami, T., Takahashi, Y., and Aoki, T. (2002). Gene expression in haemocytes of kuruma prawn, *Penaeus japonicus*, in response to infection with WSSV by EST approach. *Fish Shellfish Immunol.* 13, 69–83. doi: 10.1006/fsim.2001.0382
- Rothschild, M. F., and Plastow, G. S. (2008). Impact of genomics on animal agriculture and opportunities for animal health. *Trends Biotechnol.* 26, 21–25. doi: 10.1016/j.tibtech.2007.10.001
- Rotllant, G., Wade, N. M., Arnold, S. J., Coman, G. J., Preston, N. P., and Glencross, B. D. (2015). Identification of genes involved in reproduction and lipid pathway metabolism in wild and domesticated shrimps. *Mar. Genomics* 22, 55–61. doi: 10.1016/j.margen.2015.04.001
- Rubin, C.-J., Zody, M. C., Eriksson, J., Meadows, J. R. S., Sherwood, E., Webster, M. T., et al. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464:587. doi: 10.1038/nature08832
- Santos, C. A., Andrade, S. C. S., Teixeira, A. K., Farias, F., Kurkjian, K., Guerrelhas, A. C., et al. (2018). *Litopenaeus vannamei* transcriptome profile of populations evaluated for growth performance and exposed to White Spot Syndrome Virus (WSSV). *Front. Genet.* 9:120. doi: 10.3389/fgene.2018.00120
- Santos, C. A., Blanck, D. V., and de Freitas, P. D. (2014). RNA-seq as a powerful tool for penaeid shrimp genetic progress. *Front. Genet.* 5:298. doi: 10.3389/fgene.2014.00298
- Schock, T. B., Duke, J., Goodson, A., Weldon, D., Brunson, J., Leffler, J. W., et al. (2013). Evaluation of Pacific White Shrimp (*Litopenaeus vannamei*) health during a superintensive aquaculture growout using NMR-based metabolomics. *PLoS ONE* 8:e59521. doi: 10.1371/journal.pone.0059521
- Sedlazeck, F. J., Lee, H., Darby, C. A., and Schatz, M. C. (2018). Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat. Rev. Genet.* 19, 329–346. doi: 10.1038/s41576-018-0003-4
- Sellers, M. J., Dierens, L., McWilliam, S., Little, B., Murphy, B., Coman, G. J., et al. (2014). Comparison of microsatellite and SNP DNA markers for pedigree assignment in Black Tiger shrimp, *Penaeus monodon*. *Aquac. Res.* 45, 417–426. doi: 10.1111/j.1365-2109.2012.03243.x
- Sellers, M. J., Rao, M., Arnold, S. J., Wade, N. M., and Cowley, J. A. (2011). *Penaeus monodon* is protected against gill-associated virus by muscle injection but not oral delivery of bacterially expressed dsRNAs. *Dis. Aquat. Org.* 95, 19–30. doi: 10.3354/dao02343
- Sellers, M. J., Trewin, C., McWilliam, S. M., Graves, R., and Hertzler, P. L. (2015). Transcriptome profiles of *Penaeus (Marsupenaeus) japonicus* animal and vegetal half-embryos: identification of sex determination, germ line, mesoderm, and other developmental genes. *Mar. Biotechnol.* 17, 252–265. doi: 10.1007/s10126-015-9613-4
- Sharp, A. R., and Udall, J. A. (2016). OMWare: a tool for efficient assembly of genome-wide physical maps. *BMC Bioinformatics* 17:241. doi: 10.1186/s12859-016-1099-1
- Shen, Y., Xiang, J., Wang, B., Li, F., and Tong, W. (2004). Discovery of immune related factors in *Fenneropenaeus chinensis* by annotation of ESTs. *Prog. Nat. Sci.* 14, 47–54. doi: 10.1080/10020070412331343131
- Shi, X., Kong, J., Meng, X., Luan, S., Luo, K., Cao, B., et al. (2016). Comparative microarray profile of the hepatopancreas in the response of "Huanghai No. 2" *Fenneropenaeus chinensis* to white spot syndrome virus. *Fish Shellfish Immunol.* 58, 210–219. doi: 10.1016/j.fsi.2016.07.031
- Silvestre, F., Huynh, T. T., Bernard, A., Dorts, J., Dieu, M., Raes, M., et al. (2010). A differential proteomic approach to assess the effects of chemotherapeutics and production management strategy on giant tiger shrimp *Penaeus monodon*. *Comp. Biochem. Phys. D* 5, 227–233. doi: 10.1016/j.cbd.2010.06.003
- Small, K. S., Brudno, M., Hill, M. M., and Sidow, A. (2007). A haplome alignment and reference sequence of the highly polymorphic *Ciona savignyi* genome. *Genome Biol.* 8:R41. doi: 10.1186/gb-2007-8-3-r41
- Somboonwiwat, K., Chaikeratsak, V., Wang, H. C., Lo, C. F., and Tassanakajon, A. (2010). Proteomic analysis of differentially expressed proteins in *Penaeus monodon* hemocytes after *Vibrio harveyi* infection. *Proteome Sci.* 8:39. doi: 10.1186/1477-5956-8-39
- Somboonwiwat, K., Supungul, P., Rimphanitchayakit, V., Aoki, T., Hirono, I., and Tassanakajon, A. (2006). Differentially expressed genes in hemocytes of *Vibrio harveyi*-challenged shrimp *Penaeus monodon*. *BMB Rep.* 39, 26–36. doi: 10.5483/BMBRep.2006.39.1.026
- Sonesson, A. K., and Ødegård, J. (2016). Mating structures for genomic selection breeding programs in aquaculture. *Genet. Select. Evol.* 48, 46. doi: 10.1186/s12711-016-0224-y
- Sookruksawong, S., Sun, F., Liu, Z., and Tassanakajon, A. (2013). RNA-Seq analysis reveals genes associated with resistance to Taura syndrome virus (TSV) in the Pacific white shrimp *Litopenaeus vannamei*. *Dev. Comp. Immunol.* 41, 523–533. doi: 10.1016/j.dci.2013.07.020
- Soonthornchai, W., Chaiyapechara, S., Klinbunga, S., Thongda, W., Tangphatsornruang, S., Yoocha, T., et al. (2016). Differentially expressed transcripts in stomach of *Penaeus monodon* in response to AHPND infection. *Dev. Comp. Immunol.* 65, 53–63. doi: 10.1016/j.dci.2016.06.013
- Staelens, J., Rombaut, D., Vercauteren, I., Argue, B., Benzie, J., and Vuylsteke, M. (2008). High-density linkage maps and sex-linked markers for the black tiger shrimp (*Penaeus monodon*). *Genetics* 179, 917–925. doi: 10.1534/genetics.107.080150
- Sun, Z., Liu, P., Li, J., Meng, X., and Zhang, X. (2008). Construction of a genetic linkage map in *Fenneropenaeus chinensis* (Osbeck) using RAPD and SSR markers. *Hydrobiologia* 596, 133–141. doi: 10.1007/s10750-007-9063-8
- Supungul, P., Klinbunga, S., Pichyangkura, R., Hirono, I., Aoki, T., and Tassanakajon, A. (2004). Antimicrobial peptides discovered in the black tiger shrimp *Penaeus monodon* using the EST approach. *Dis. Aquat. Org.* 61, 123–135. doi: 10.3354/dao061123
- Supungul, P., Klinbunga, S., Pichyangkura, R., Jitrapakdee, S., Hirono, I., Aoki, T., et al. (2002). Identification of immune-related genes in hemocytes of black tiger shrimp (*Penaeus monodon*). *Mar. Biotechnol.* 4, 487–494. doi: 10.1007/s10126-002-0043-8
- Tassanakajon, A., Klinbunga, S., Paunglar, N., Rimphanitchayakit, V., Udomkit, A., Jitrapakdee, S., et al. (2006). *Penaeus monodon* gene discovery project: the generation of an EST collection and establishment of a database. *Gene* 384, 104–112. doi: 10.1016/j.gene.2006.07.012
- Tassanakajon, A., Somboonwiwat, K., Supungul, P., and Tang, S. (2013). Discovery of immune molecules and their crucial functions in shrimp immunity. *Fish Shellfish Immunol.* 34, 954–967. doi: 10.1016/j.fsi.2012.09.021
- Taylor, J. F., Taylor, K. H., and Decker, J. E. (2016). Holsteins are the genomic selection poster cows. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7690–7692. doi: 10.1073/pnas.1608144113
- Thodesen, J., and Gjedrem, T. (2006). "Development of aquatic animal genetic improvement and dissemination programs: current status and action plans," in *WorldFish Center Conference Proceedings*, eds R. W. Ponzoni, B. O. Acosta, and A. G. Ponniah (Penang), 22–26.
- Thorgaard, G. H., Nichols, K. M., and Phillips, R. B. (2006). Comparative gene and QTL mapping in aquaculture species. *Israeli J. Aquacult. Bamidgheh* 58, 4.

- Tian, Y., Kong, J., and Wang, W. (2008). Construction of AFLP-based genetic linkage maps for the Chinese shrimp *Fenneropenaeus chinensis*. *Chinese Sci. Bull.* 53, 1205–1216. doi: 10.1007/s11434-008-0170-3
- Toro, M., Saura, M., Fernandez, J., and Villanueva, B. (2017). Accuracy of genomic within-family selection in aquaculture breeding programmes. *J. Anim. Breed. Genet.* 134, 256–263. doi: 10.1111/jbg.12272
- Tsai, H.-Y., Matika, O., Edwards, S. M., Antolin-Sánchez, R., Hamilton, A., Guy, D. R., et al. (2017). Genotype imputation to improve the cost-efficiency of genomic selection in farmed Atlantic Salmon. *G3*, 1377–1383. doi: 10.1534/g3.117.040717
- Tyagi, A., Vernekar, P., Karunasagar, I., and Karunasagar, I. (2008). Determination of chloramphenicol in shrimp by liquid chromatography-electrospray ionization tandem mass spectrometry (LC-ESI-MS-MS). *Food Addit. Contam. Part A* 25, 432–437. doi: 10.1080/02652030701586665
- Uawisetwathana, U., Leelatanawit, R., Klanchui, A., Prommoon, J., Klinbunga, S., and Karoonuthaisiri, N. (2011). Insights into eyestalk ablation mechanism to induce ovarian maturation in the black tiger shrimp. *PLoS ONE* 6:e24427. doi: 10.1371/journal.pone.0024427
- Van Emon, J. M. (2015). The omics revolution in agricultural research. *J. Agric. Food Chem.* 64, 36–44. doi: 10.1021/acs.jafc.5b04515
- Vandeputte, M., and Haffray, P. (2014). Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Front. Genet.* 5:432. doi: 10.3389/fgene.2014.00432
- Vega, E. D. L., Hall, M. R., Wilson, K. J., Reverter, A., Woods, R. G., and Degnan, B. M. (2007). Stress-induced gene expression profiling in the black tiger shrimp *Penaeus monodon*. *Physiol. Genomics* 31, 126–138. doi: 10.1152/physiolgenomics.00068.2007
- Veloso, A., Warr, G. W., Browdy, C. L., and Chapman, R. W. (2011). The transcriptomic response to viral infection of two strains of shrimp (*Litopenaeus vannamei*). *Dev. Comp. Immunol.* 35, 241–246. doi: 10.1016/j.dci.2010.10.001
- Vij, S., Kuhl, H., Kuznetsova, I. S., Komissarov, A., Yurchenko, A. A., Van Heusden, P., et al. (2016). Chromosomal-level assembly of the Asian seabass genome using long sequence reads and multi-layered scaffolding. *PLoS Genet.* 12:e1005954. doi: 10.1371/journal.pgen.1005954
- Visscher, P. M., Hemani, G., Vinkhuyzen, A. A. E., Chen, G.-B., Lee, S. H., Wray, N. R., et al. (2014). Statistical power to detect genetic (Co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet.* 10:e1004269. doi: 10.1371/journal.pgen.1004269
- Vuylsteke, M., Van Breusegem, F., and Staelens, J. (2012). *Sex-Specific Marker for Shrimps and Prawns*. Google Patent. Brussels: European Patents Office.
- Wang, B., Li, F., Dong, B., Zhang, X., Zhang, C., and Xiang, J. (2006). Discovery of the Genes in Response to White Spot Syndrome Virus (WSSV) Infection in *Fenneropenaeus chinensis* Through cDNA Microarray. *Mar. Biotechnol.* 8, 491–500. doi: 10.1007/s10126-005-6136-4
- Wang, H. C., Wang, H. C., Leu, J. H., Kou, G. H., Wang, A. H. J., and Lo, C. F. (2007). Protein expression profiling of the shrimp cellular response to white spot syndrome virus infection. *Dev. Comp. Immunol.* 31, 672–686. doi: 10.1016/j.dci.2006.11.001
- Wang, Q., Yu, Y., Li, F., Zhang, X., and Xiang, J. (2016). Predictive ability of genomic selection models for breeding value estimation on growth traits of Pacific white shrimp *Litopenaeus vannamei*. *Chinese J. Oceanol. Limnol.* 35:1221. doi: 10.1007/s00343-017-6038-0
- Wang, Q., Yu, Y., Yuan, J., Zhang, X., Huang, H., Li, F., et al. (2017). Effects of marker density and population structure on the genomic prediction accuracy for growth trait in Pacific white shrimp *Litopenaeus vannamei*. *BMC Genet.* 18:45. doi: 10.1186/s12863-017-0507-5
- Wang, W., Tian, Y., Kong, J., Li, X., Liu, X., and Yang, C. (2012). Integration genetic linkage map construction and several potential QTLs mapping of Chinese shrimp (*Fenneropenaeus chinensis*) based on three types of molecular markers. *Russ. J. Genet.* 48, 422–434. doi: 10.1134/S102279541204014X
- Wang, Z. Y., Tsai, K. H., and Chu, K. H. (2004). Applications of AFLP technology in genetic and phylogenetic analysis of penaeid shrimp. *Biochem. Syst. Ecol.* 32, 399–407. doi: 10.1016/j.bse.2003.10.006
- Wei, J., Zhang, X., Yu, Y., Li, F., and Xiang, J. (2014). RNA-Seq reveals the dynamic and diverse features of digestive enzymes during early development of Pacific white shrimp *Litopenaeus vannamei*. *Comp. Biochem. Physiol. Part D Genomics Proteomics* 11, 37–44. doi: 10.1016/j.cbd.2014.07.001
- Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., and Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome Res.* 27, 757–767. doi: 10.1101/gr.214874.116
- Wilson, K., Li, Y. T., Whan, V., Lehnert, S., Byrne, K., Moore, S., et al. (2002). Genetic mapping of the black tiger shrimp *Penaeus monodon* with amplified fragment length polymorphism. *Aquaculture* 204, 297–309. doi: 10.1016/S0044-8486(01)00842-0
- Winfield, M. O., Allen, A. M., BurrIDGE, A. J., Barker, G. L., Benbow, H. R., Wilkinson, P. A., et al. (2015). High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol. J.* 14, 1195–1210. doi: 10.1111/pbi.12485
- Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6:359. doi: 10.1038/nmeth.1322
- Wolc, A., Zhao, H. H., Arango, J., Settar, P., Fulton, J. E., O’Sullivan, N. P., et al. (2015). Response and inbreeding from a genomic selection experiment in layer chickens. *Genet. Select. Evol.* 47:59. doi: 10.1186/s12711-015-0133-5
- Wolfus, G. M., Garcia, D. K., and Alcivar-Warren, A. (1997). Application of the microsatellite technique for analyzing genetic diversity in shrimp breeding programs. *Aquaculture* 152, 35–47. doi: 10.1016/S0044-8486(96)01527-X
- Wongpanya, R., Aoki, T., Hirono, I., Yasuie, M., and Tassanakajon, A. (2007). Analysis of gene expression in haemocytes of shrimp *Penaeus monodon* challenged with white spot syndrome virus by cDNA microarray. *Sci. Asia* 33, 165–174. doi: 10.2306/scienceasia1513-1874.2007.33.165
- Wongsurawat, T., Leelatanawit, R., Thamniemdee, N., Uawisetwathana, U., Karoonuthaisiri, N., Menasveta, P., et al. (2010). Identification of testis-relevant genes using *in silico* analysis from testis ESTs and cDNA microarray in the black tiger shrimp (*Penaeus monodon*). *BMC Mol. Biol.* 11:55. doi: 10.1186/1471-2199-11-55
- Wu, J., Lin, Q., Lim, T. K., Liu, T., and Hew, C. L. (2007). White spot syndrome virus proteins and differentially expressed host proteins identified in shrimp epithelium by shotgun proteomics and cleavable isotope-coded affinity tag. *J. Virol.* 81, 11681–11689. doi: 10.1128/JVI.01006-07
- Wuthisuthimethavee, S., Lumubol, P., Toojinda, T., Tragoonrun, S., and Vanavichita, A. (2005). SSLP-based linkage map construction in black tiger shrimp (*Penaeus monodon* Fabricius). *Sci. Asia* 31, 91–97. doi: 10.2306/scienceasia1513-1874.2005.31.091
- Xiang, J. (2016). “Progress of the shrimp genome project,” in *Plant & Animal Genome Conference XXIV* (San Diego, CA).
- Xiang, J., Wang, B., Li, F., Liu, B., Zhou, Y., and Tong, W. (2008). “Generation and analysis of 10,443 ESTs from cephalothorax of *Fenneropenaeus chinensis*,” in *Bioinformatics and Biomedical Engineering, 2008 ICBBE 2008 The 2nd International Conference on 2008* (Shanghai: IEEE), 74–80. doi: 10.1109/ICBBE.2008.25
- Xiaoxian, L., Wenbin, M., Soltis, S. P., Soltis, S. D., and Brad, B. W. (2017). Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome. *Mol. Ecol. Resour.* 17, 1243–1256. doi: 10.1111/1755-0998.12670
- Xu, C., Li, E., Liu, Y., Wang, X., Qin, J. G., and Chen, L. (2017). Comparative proteome analysis of the hepatopancreas from the Pacific white shrimp *Litopenaeus vannamei* under long-term low salinity stress. *J. Proteomics* 162, 1–10. doi: 10.1016/j.jprot.2017.04.013
- Xu, Y., Jiang, Y., Shi, T., Cai, H., Lan, X., Zhao, X., et al. (2017). Whole-genome sequencing reveals mutational landscape underlying phenotypic differences between two widespread Chinese cattle breeds. *PLoS ONE* 12:e0183921. doi: 10.1371/journal.pone.0183921
- Xu, Z., Primavera, J. H., de la Pena, L. D., Pettit, P., Belak, J., and Alcivar-Warren, A. (2001). Genetic diversity of wild and cultured black tiger shrimp (*Penaeus monodon*) in the Philippines using microsatellites. *Aquaculture* 199, 13–40. doi: 10.1016/S0044-8486(00)00535-4
- Xue, S., Liu, Y., Zhang, Y., Sun, Y., Geng, X., and Sun, J. (2013). Sequencing and *de novo* analysis of the hemocytes transcriptome in *Litopenaeus vannamei* response to white spot syndrome virus infection. *PLoS ONE* 8:e76718. doi: 10.1371/journal.pone.0076718
- Yamano, K., and Unuma, T. (2006). Expressed sequence tags from eyestalk of kuruma prawn, *Marsupenaeus japonicus*. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* 143, 155–161. doi: 10.1016/j.cbpa.2005.11.005

- Yáñez, J. M., Naswa, S., López, M., Bassini, L., Correa, K., Gilbey, J., et al. (2016). Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Mol. Ecol. Resour.* 16, 1002–1011. doi: 10.1111/1755-0998.12503
- You, E. M., Liu, K. F., Huang, S. W., Chen, M., Groumellec, M., Fann, S. J., et al. (2010). Construction of integrated genetic linkage maps of the tiger shrimp (*Penaeus monodon*) using microsatellite and AFLP markers. *Anim. Genet.* 41, 365–376. doi: 10.1111/j.1365-2052.2009.02014.x
- Yu, Y., Wei, J., Zhang, X., Liu, J., Liu, C., Li, F., et al. (2014). SNP Discovery in the transcriptome of white pacific shrimp *Litopenaeus vannamei* by next generation sequencing. *PLoS ONE* 9:e87218. doi: 10.1371/journal.pone.0087218
- Yu, Y., Zhang, X., Yuan, J., Li, F., Chen, X., Zhao, Y., et al. (2015). Genome survey and high-density genetic map construction provide genomic and genetic resources for the Pacific White Shrimp *Litopenaeus vannamei*. *Sci. Rep.* 5:15612. doi: 10.1038/srep15612
- Zeng, D., Chen, X., Xie, D., Zhao, Y., Yang, C., Li, Y., et al. (2013). Transcriptome analysis of Pacific white shrimp (*Litopenaeus vannamei*) hepatopancreas in response to Taura syndrome Virus (TSV) experimental infection. *PLoS ONE* 8:e57515. doi: 10.1371/journal.pone.0057515
- Zeng, Q., Fu, Q., Li, Y., Waldbieser, G., Bosworth, B., Liu, S., et al. (2017). Development of a 690 K SNP array in catfish and its application for genetic mapping and validation of the reference genome sequence. *Sci. Rep.* 7:40347. doi: 10.1038/srep40347
- Zenger, K. R., Khatkar, M. S., Jerry, D. R., and Raadsma, H. W. (2017). “The next wave in selective breeding: implementing genomic selection in aquaculture,” in *AAABG*, Edited by Genetics AftAoABa (Townsville, QLD).
- Zhang, D., Wang, F., Dong, S., and Lu, Y. (2016). De novo assembly and transcriptome analysis of osmoregulation in *Litopenaeus vannamei* under three cultivated conditions with different salinities. *Gene* 578, 185–193. doi: 10.1016/j.gene.2015.12.026
- Zhang, J., Li, F., Jiang, H., Yu, Y., Liu, C., Li, S., et al. (2010). Proteomic analysis of differentially expressed proteins in lymphoid organ of *Fenneropenaeus chinensis* response to *Vibrio anguillarum* stimulation. *Fish Shellfish Immunol.* 29, 186–194. doi: 10.1016/j.fsi.2010.03.011
- Zhang, J.-Y., Wang, W.-J., Kong, J., and Wang, Q.-Y. (2013). Construction of a genetic linkage map in *Fenneropenaeus chinensis* using SNP markers. *Russ. J. Mar. Biol.* 39, 136–142. doi: 10.1134/S1063074013020107
- Zhang, L., Yang, C., Zhang, Y., Li, L., Zhang, X., Zhang, Q., et al. (2007). A genetic linkage map of Pacific white shrimp (*Litopenaeus vannamei*): sex-linked microsatellite markers and high recombination rates. *Genetica* 131, 37–49. doi: 10.1007/s10709-006-9111-8
- Zhang, X., Zhang, Y., Scheuring, C., Zhang, H.-B., Huan, P., Wang, B., et al. (2010). Construction and characterization of a bacterial artificial chromosome (BAC) library of Pacific white shrimp, *Litopenaeus vannamei*. *Mar. Biotechnol.* 12, 141–149. doi: 10.1007/s10126-009-9209-y
- Zhao, C., Zhang, X., Liu, C., Huan, P., Li, F., Xiang, J., et al. (2012). BAC end sequencing of Pacific white shrimp *Litopenaeus vannamei*: a glimpse into the genome of Penaeid shrimp. *Chinese J. Oceanol. Limnol.* 30, 456–470. doi: 10.1007/s00343-012-1159-y
- Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., et al. (2009). A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10:R42. doi: 10.1186/gb-2009-10-4-r42

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Guppy, Jones, Jerry, Wade, Raadsma, Huerlimann and Zenger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.