The NGS Magic Pudding: A Nanopore-Led Long-Read Genome Assembly for the Commercial Australian Freshwater Crayfish, *Cherax destructor*

Christopher M. Austin1,2*, Laurence J. Croft1,2, Frederic Grandjean3 and Han Ming Gan4

1Deakin Genomics Centre, Deakin University, Geelong, VIC, Australia, 2Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Geelong, VIC, Australia, 3Laboratoire Ecologie et Biologie des Interactions, Équipe Écologie Évolution Symbiose, Unité Mixte de Recherche 7267 Centre National de la Recherche Scientifique, Université de Poitiers, Poitiers, France, 4GeneSEQ Sdn Bhd, Rawang, Malaysia

Cherax destructor, the yabby, is an iconic Australian freshwater crayfish species, which, similar to other major invertebrate groups, is grossly under-represented in genomic databases. The yabby is also the principal commercial freshwater crustacean species in Australia subject to exploitation via inland fisheries and aquaculture. To address the genomics knowledge gap for this species and explore cost effective and efficient methods for genome assembly, we generated 106.8 gb of Nanopore reads and performed a long-read only assembly of the *Cherax destructor* genome. On a mini-server configured with an ultra-fast swap space, the *de novo* assembly took 131 h (~5.5 days). Genome polishing with 126.3 gb of PCR-Free Illumina reads generated an assembled genome size of 3.3 gb (74.6% BUSCO completeness) with a contig N50 of 80,900 bp, making it the most contiguous for freshwater crayfish genome assemblies. We found an unusually large number of cellulase genes within the yabby genome which is relevant to understanding the nutritional biology, commercial feed development, and ecological role of this species and crayfish more generally. These resources will be useful for genomic research on freshwater crayfish and our methods for rapid and super-efficient genome assembly will have wide application.

Keywords: genome, annotation, nanopore, cellulase, aquaculture, decapoda, Parastacidae

1 INTRODUCTION

Australia’s freshwater crayfish are highly diverse and as charismatic as the country’s better known avian and mammalian fauna, but far less appreciated and studied. Crayfish are found in a range of freshwater environments, include some exceptionally large species in Australia, and can reach very high densities in both natural and cultured environments (Nyström and Strand, 1996; Whitlegde and Rabeni, 1997; Jones and Ruscoe, 2000; Reynolds and Richardson, 2013). As a result, they often represent keystone species and ecosystem engineers in permanent and semi-permanent freshwater systems in many parts of the world. This also means they are an important part of food webs as significant prey items for fish, birds and mammals (Hicks and McCaughan, 1997; Jones and Grey, 2016), and for humans, including indigenous communities (Eyre et al., 1845; Austin, 1998; Kusabs
Crayfish also have significant ecological roles within inland aquatic systems as they can consume and process sizeable volumes of a range of organic matter and detritus (Nyström and Strand, 1996; Whittledge and Rabeni, 1997; Reynolds and Richardson, 2013; Jones and Grey, 2016). While crayfish are widely considered as omnivorous and opportunistic feeders their exact ecological role and nutritional biology has been controversial (Momot, 1995) and are assuming greater importance with the frequent translocation of crayfish species and their potential to cause a range of negative ecological impacts both locally and globally (Austin and Ryan, 2002; Lodge et al., 2001; Reynolds and Richardson, 2013; Jones and Grey, 2016). Some authors have postulated that freshwater crayfish are primarily carnivorous (Momot, 1995; Weinländer and Fürderer, 2012), however molecular and limited NGS-based studies have revealed the presence of cellulase and a diversity of carbohydrate-active related genes supporting an adaption to the processing of plant-based food (Crawford et al., 2005; Tan et al., 2016). The first cellulase reported for freshwater crayfish was from the GH9 family which was found to be especially diverse in Cherax quadricarinatus based on a transcriptomic study by Tan et al. (2015) Tan et al. (2016). The second cellulase was fragmented to 8 kb using Covaris g-tube and 2–4 ug of the fragmented gDNA was subsequently used to construct a Nanopore library and was sequenced on the LS830 sequencing machine. The sequencing was done on a NovaSeq6000 using a run configuration of 2 × 150 bp. A total of 418,053,185 paired-end reads were generated totaling to 126.3 gigabases.
two 1 TB drives. Nanopore reads and intermediate assembly files were all stored on a separate RAID 0 partition comprising four 4 TB hard drives. De novo assembly of the Nanopore reads used wtdbg 2.5 (Ruan and Li, 2019) with the options “-l 60 -p 19 -AS 2 -s 0.05 -L 3000 -g 6G -edge-min 2 --reduce-low-cov-edges”. Using this configuration, the de novo assembly took 131 h (∼5.5 days) to complete with a maximum memory usage of 607 GB. After the wtdbg assembly, one round of polishing with long reads was performed using the wtdbg 2.5 internal polishing tool, wtpoa-cns. For genome polishing with Illumina reads, two rounds of polishing with Pilon v1.22 (Walker et al., 2014) were carried out. The raw paired-end reads were first adapter, quality and poly-G trimmed with fastp v0.20.0 (Chen et al., 2018). For each round of pilon-polishing, the trimmed reads were aligned to the genome using bwa-mem v0.7.17-r1188 (Li, 2013) followed by correction of individual base errors (SNPs) and small indels using the options “--diploid --fix bases”. To overcome memory limitation in Pilon due to large genome size, the genome was split into 10 smaller fasta files, processed with Pilon separately and merged back into a single fasta file. Transcriptome-guided scaffolding of the polished contigs was performed with P_RNA_scaffold v1 (Zhu et al., 2018) using publicly available transcriptome data (Ali et al., 2015). The genome completeness was assessed using BUSCO v5 (Waterhouse et al., 2017) with the Arthropoda ortholog dataset (Arthropod odb10). Statistics of the resulting assembly were generated using QUAST v5.0.2 (Gurevich et al., 2013) and are presented in Table 1. Illumina and Nanopore reads were mapped to the final assembly using bwa-mem (Li, 2013) and minimap2 v2.17 (Li, 2018), respectively. The BAM files were separately processed in Qualimap2 v2.2.1 (Okonechnikov et al., 2016) to generate additional statistics for the genome assembly based on read alignment.

### Table 1: Genome assembly and annotation statistics.

| Parameter | Details |
|-----------|---------|
| Organism  | Cherax destructor (Australian yabby) |
| Isolate   | CDF2 (female, wild population) |
| Bioproject| PRJNA588861 |
| Biosample | SAMN13295587 |
| Whole-genome GenBank accession | WNWK00000000 |
| Assembled scaffold/contig length | 3,336,744,225 bp / 3,336,542,869 bp |
| Scaffold N50 (number of sequences) | 87,184 bp (98,662) |
| Contig N50 (number of sequences) | 80,900 bp (100,635) |
| GC content | 41.43% |
| BUSCO completeness | 74.6% Single-copy, 1.1% Duplicated |
| Number of predicted protein-coding genes | 45,673 |
| Number of predicted proteins | 47,377 |
| With InterPro signature | 21,102 (44.5%) |
| With gene ontology (GO) term | 14,068 (29.7%) |

2.3 Repeat Annotation and Protein-Coding Gene Prediction

Repetitive regions were identified using RepeatModeler v1.0.11 (Smit and Hubley, 2010). The de novo generated repeat library (Gan et al., 2020) was subsequently used to soft-mask the genome assembly with RepeatMaskerv4.0.7 (Tarailo-Graovac and Chen, 2009) with the options "-no_is -div 40 -xsmall". Using this repeat annotation approach, 61.34% of the genome has been repeat-masked with long interspersed nuclear elements (LINEs) being the most common repeat annotated (31%). For protein-coding gene prediction, BRAKER v2.1.4 (Hoff et al., 2019) was chosen since it can incorporate both RNA-sequencing data and closely related proteins for gene prediction training. Publicly available Cherax destructor transcriptome datasets (Ali et al., 2015) were downloaded and aligned to the genome using STAR v2.7.1a (Dobin et al., 2013). To obtain closely related protein sequences, all publicly available Cherax quadricarinatus transcriptome data were downloaded from NCBI-SRA as of 2nd December 2019, individually assembled using rnaspades v3.13.0 (Bushmanova et al., 2019) followed by redundancy removal of the concatenated transcripts using EvidentialGene v2013.03.11 (Gilbert, 2019). Cherax quadricarinatus translated open reading frames that are larger than 200 amino acid residues and labelled as “complete” e.g., with intact 5’ and 3’ ends, were selected as the protein input (Gan et al., 2020) for training in BRAKER2 using default settings. Using Orthofinder v2.3.8 (Emms and Kelly, 2018), the initial predicted proteins from BRAKER2 were used as the input for orthologous clustering with the available proteomes of the red claw crayfish (C. quadricarinatus) (Tan et al., 2020a), pacific white shrimp (Litopenaeus vannamei) (Zhang et al., 2019), black tiger prawn (Penaeus monodon) (Quyen et al., 2020), marbled crayfish (Procambarus virginalis) (Gutekunst et al., 2018), and amphipod (Parhyale hawaiensis) (Kao et al., 2016). Then, the predicted C. destructor proteins that formed orthologous clusters with at least one of the decapod species were used for subsequent annotation and analysis. Specific comparisons of peptide homology were made with several decapod crustaceans including the recently published clawed lobster genome (clawed lobsters are from the clade most closely related to the freshwater crayfish) (Polinski et al., 2021), the southern hemisphere crayfish (Cherax quadricarinatus) using NCBI’s blastx (evalue 1e-10). Putative protein functions were inferred using Interproscan v5.35-74.0 (Jones et al., 2014) with the options “--iprlookup –gotermms --dp”. Identification of Carbohydrate-Active enzymes (CAZy) in the selected crustacean proteomes used dbCAN2 v2.0.0 (Zhang et al., 2018) and the identified GH9 cellulases were further extracted and their diversity explored by phylogenetic analysis. The GH9 cellulases were first aligned with MUSCLE v3.8.31 (Edgar, 2004) followed by trimming in trimal v1.9 (Capella-Gutierrez et al., 2009) (“automated” option) and phylogenetic construction in IQTree v1.6.10 (Nguyen et al., 2014) (“-m TESTNEW –bb 1,000” options). The unrooted IQTree maximum likelihood tree was annotated and visualized in TreeFig v1.4.3 (Rambaut, 2009).

2.4 Data Availability

Raw sequencing libraries have been deposited in NCBI-SRA under the BioProject PRJNA588861. The genome assembly has been deposited in GenBank under the accession number WNWK0000000 (the version described in this paper is
WNWK01000000). The wtdbg2.5 assembly log file, intermediate
C. destructor genome assemblies, repeat annotation, CAZy
annotation, protein-coding gene prediction (GTF format),
predicted genes, and proteins have been deposited in the
Zenodo repository (Gan et al., 2020). The C. quadricarinatus
RNAspades transcriptome assemblies, QUAST-generated
genome statistics for all Decapod genomes re-analyzed in this
paper and their BUSCO calculations are also deposited at Zenodo
(Gan et al., 2020).

3 RESULTS AND DISCUSSION

An alignment rate of more than 99.5% was observed for both
Illumina and Nanopore reads with the most frequently observed
sequencing depth of 29× and 23×, respectively. Assuming the
sequencing depth with the highest observed frequency represents
the coverage of the single-copy genomic region, the genome size of
Cherax destructor is estimated to be 4.36–4.64 gb (Total sequencing
yield in gigabases divided by single-copy coverage). This is consistent
with genome size estimates for the northern hemisphere crayfish
Procambarus virginalis (∼3.5 gb) and Cherax quadricarinatus (∼5 gb) (Tan et al., 2020a) making Australian crayfish larger than
all other crustaceans so far sequenced with the exception of the
prawn Exoplaemon carinicauda (9.5 gb).

Using 106.8 gb and 126.3 gb of Nanopore and Illumina data,
respectively, a 3.3 gb genome assembly was generated with an
estimated BUSCO score of 89.7% in less than a week. The
assembled genome size was ∼27.0% smaller than the genome
size estimate. This is quite a common outcome for decapod
genome assemblies due to sequencing bias and their repetitive
genomes (Tan et al., 2020a; Polinski et al., 2021) and was reflected
in the uneven distribution of read depths across scaffolds in our
study. Over 3,000 scaffolds have over 300× coverage, compared
with an average read depth of 111×, consistent with the occurrence
of a significant proportion of repeat regions and potentially
contributing to the discrepancy between the assembled
genome size and the genome size estimate.

FIGURE 2 | Statistics of publicly available decapod crustacean genome assemblies. (A) Number of sequences, Genome N50, and total assembled length (B)
BUSCO completeness based on the Arthropoda ortholog dataset.
The contig N50 of 80,900 bp is the longest to date among currently available for freshwater crayfish genome assemblies. Comparisons with the recently sequenced *Cherax quadricarinatus* genome (Tan et al., 2020a) initially assembled using short reads followed by scaffolding with low coverage Nanopore long reads, show that a long-read led assembly is more efficient though more costly. However, the higher cost of long reads is more than compensated for by increased computational efficiencies due to the availability of speed and memory-efficient long-read assemblers (wtdbg2) (Ruan and Li, 2019) and lack of reliance on the need to generate large volumes of Illumina reads during the initial assembly stage.

The cumulative scaffold length of *Cherax destructor* is similar to the *C. quadricarinatus* genome (~3 gb) that was assembled using Illumina reads (191x) followed by scaffolding with low coverage Nanopore reads (x7). In comparison with the other two crayfish assemblies the advantages of a Nanopore-based assembly with an increased volume of long reads can be seen from Figure 2, where the difference between the contig and scaffold level assemblies is greatly reduced leading to a less gappy assembly. Also the need for high volumes of short reads is also greatly reduced with only 123.6 gb used in their study compared with used for the assemblies of *C. quadricarinatus* (964 gb) and *Procambarus virginalis* (350 gb).

It is also worth noting that this *C. destructor* genome assembly exhibits a contig N50 length of nearly 100 kb which is longest among freshwater crayfish genome assemblies. Recent decapod assemblies increasingly using both short and long reads and Hi-C data which is assisting in more robust decapod crustacean genome assemblies (Zhang et al., 2019; Tang et al., 2020) especially for those species with large repetitive genomes such as *Macrobrachium* shrimps (Jin et al., 2021). The reported *C. destructor* BUSCO genome completeness in this study is also one of the highest to date for freshwater crayfish (Figure 2B). A logical next step, given the large and repetitive genomes exhibited by freshwater crayfish, is to attempt to improve this genome assembly via the inclusion of HiC data (Jin et al., 2021).

An initial 187,638 of putative unigenes were predicted by BRAKER2. The final protein set consisted of 47,377 transcripts (45,673 genes) of which 21,102 and 14,068 were identified with InterPro signature and Gene Ontology term, respectively. The number of predicted proteins with InterPro signatures is very similar to other species of decapod crustaceans. A total of 68.97% of *C. destructor* peptides mapped to the related *C. quadricarinatus* annotation (evalue $1e^{-10}$) (Tan et al., 2016). More specifically, we get 32,677 peptides in common with *C. quadricarinatus*, 25,129 with *Procambarus virginalis*, 23,008 with *Penaeus monodon*, 17,159 with *Litopenaeus vannamei*, and 10,318 with *Homarus americanus*. The number of predicted proteins with InterPro signatures is very similar to other species of decapod crustaceans (Tan et al., 2016). While the total number of predicted protein-coding genes is large (45,673) relative to those that have an Interpro signature, this number does not differ greatly from the recently published genome for the clawed lobster, *Homarus americanus*, which identified 40,732 peptides (Polinski et al., 2021). This high proportion of unique genes is most likely a function of the evolution of a large repetitive genome and the limited genomic data for crayfish and lobsters as pointed out by Polinski et al. (2021) in their recent study of the American lobster (Polinski et al., 2021). Significantly, *Cherax destructor* harbours the highest number of cellulase genes among the currently sequenced decapod crustaceans (Figure 3A) with a substantially higher number of GH9 cellulase genes comparable to its close relative, *C. quadricarinatus*, which was previously highlighted in an earlier transcriptomic study (Tan et al., 2016). Phylogenetic analysis of the GH9 cellulases showed a clustering pattern first by the GH9 cellulase variants and then by species relatedness (Figure 3B).
Despite the high number of GH9 cellulases identified among the *Cherax* spp., they were generally closely related and localized in a few major clades (Figure 3B). Although there were a few that claded with those from the northern hemisphere crayfish *P. virginalis*, indicating a more ancient origin, *Cherax destructor*, is considered to be versatile in its nutrient utilisation based on both dietary and field-based studies (Jones and De Silva, 1997; Beatty, 2006; Giling et al., 2009; Johnston et al., 2011) and is considered an opportunistic omnivorous generalist, that can derive nutrition directly from both animal and plant material and detritus.

A common view is that crayfish, in general, have a trophic role primarily as predators (Momot, 1995) may need to be reassessed, given the antiquity, and diversity of cellulase and related genes in this group. However there also may be wide variation within and among crayfish species and the diet of particular species can vary in time and space (Beatty, 2006; Giling et al., 2009; Johnston et al., 2011) which has contributed to conflicting views. For example, Johnston et al. (2011) found variation between species from the same crayfish community ranging from primarily herbivorous species to primarily carnivorous species. Other species from this crayfish community, including *C. destructor*, had either mixed diets or switched between plant, and animal diets at different sites. It will, therefore, be of great interest to further examine cellulase diversity and expression in a range of crayfishes species from different environments including under aquaculture conditions and the ability of different crayfish species to utilise plant material in the field and through laboratory trials and how this relates to cellulase gene profiles and their expression.

In general, a significant limitation in further advancing the study of the genomics of non-model organisms is the computational resources and time needed to assemble genomes from predominately short reads, even when aided with long reads for scaffolding (Lewin et al., 2018). This problem is further exacerbated for groups with larger repetitive genomes, which means analyses can take months if not years and still lead to poor quality assemblies. In this study, we demonstrate that a high-quality genome assembly for a decapod crustacean with a large (>3 gb) and repetitive genome can be achieved with modest sequencing volumes, that take advantage of rapid and ongoing developments in third generation sequencing technologies, and can be completed in under 1 week of computation time on a high performance desktop machine.

### 4 CONCLUSION

This reference genome, along with its annotation, will be useful for future functional, ecological, aquaculture-related and evolutionary genomic studies, and genome-based selection and targeted genetic manipulation of this emerging aquaculture species. Given our finding of an evolutionary proliferation of cellulase genes, we are hoping these data will stimulate new research into the nutritional biology and trophic roles of freshwater crayfish in freshwater ecosystems. We see the continuing advances in Nanopore and other third generation sequencing technologies like the fabled “magic pudding” from a well known Australian children’s story (Norman, 1918), it keeps on “giving”, similar to the continuing improvements in efficiency, output volume, and accuracy making the intractable, tractable when it comes to genome sequencing and assembly of non-model species. As a consequence we are able to provide a new model with respect to sequencing platforms, hardware configuration and assembly strategy to enable an ultrafast and efficient genome assembly that can be potentially applied to any species, including those with large and repetitive genomes. We anticipate our strategy and methodology will help elevate the study of interesting and important invertebrate genomes.

### DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, PRJNA588861 https://www.ncbi.nlm.nih.gov/genbank/, SAMN13258587 https://www.ncbi.nlm.nih.gov/genbank/, WNWK00000000.1.

### AUTHOR CONTRIBUTIONS

HG—Conceived and designed the analysis, collected the data, performed the analysis, and wrote the paper. FG—Conceived and designed the study, contributed to the paper. LC—Contributed to bioinformatics and discussion. CA—Conceived and designed the analysis, collected the data, contributed data, and wrote the paper.

### FUNDING

Funding was provided by Deakin University and the University of Poitiers.

### ACKNOWLEDGMENTS

We would like to thank Julian Vreugdenburg for the technical support and configuration of the mini-server which enabled the rapid completion of the memory-intensive *de novo* assembly.

Austin, C. M. (1998). Potential for the Commercial Exploitation of Freshwater Crayfish via Aquaculture in the Mt Bosavi Region of Papua New Guinea - a Preliminary Report. Geelong, Australia: Unpublished Report for the World Wide Fund for Nature (WWF).

Austin, C. M., Tan, M. H., Harrisson, K. A., Lee, Y. P., Croft, L. J., Sunnucks, P., et al. (2017). De Novo genome Assembly and Annotation of Australia’s Largest

### REFERENCES

Ali, M. Y., Pavasovic, A., Amin, S., Mather, P. B., and Prentis, P. J. (2015). Comparative Analysis of Gill Transcriptomes of Two Freshwater Crayfish, Cherax Cainii and C. Destructor. *Mar. Genomics* 22, 11–13. doi:10.1016/j.margen.2015.03.004
Okonechnikov, K., Conesa, A., and Garcia-Alcalde, F. (2016). Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data. Bioinformatics 32 (2), 292–294. doi:10.1093/bioinformatics/btv566

Piper, L. (2000). Potential for Expansion of the Freshwater Crayfish Industry in Australia: a Report for the Rural Industries Research and Development Corporation. Available at: https://www.agrifutures.com.au/wp-content/uploads/publications/00-142.pdf

Polinski, J. M., Zimin, A. V., Clark, K. F., Kohn, A. B., Sadowski, N., Timp, W., et al. (2021). The American Lobster Genome Reveals Insights on Longevity, Neural, and Immune Adaptations. Sci. Adv. 7 (26), eabe8290. doi:10.1126/sciadv.ab8290

Quyen, D. V., Gan, H. M., Lee, Y. P., Nguyen, D. D., Tran, X. T., Nguyen, V. S., et al. (2020). Improved Genomic Resources for the Black Tiger Prawn (Penaeus monodon). Marine Genomics 52, 100751. doi:10.1016/j.margen.2020.100751

Rambaut, A. (2013). FigTree. Available at: http://tree.bio.ed.ac.uk/software/figtree/ (Accessed on 9th January 2020).

Reynolds, J., Souty-Grosset, C., and Richardson, A. (2013). Ecological Roles of Crayfish in Freshwater and Terrestrial Habitats. Freshwater. Crayfish 19 (2), 197–218. doi:10.5869/fc.2013.v19.2.197

Ruan, J., and Li, H. (2019). Fast and Accurate Long-Read Assembly with Wtdbg2. Nat. Methods. 17. doi:10.1038/s41592-019-0669-3

Smit, A. F., and Hubley, R. (2010). RepeatModeler Open-1.0. Souty-Grosset, C., Anastácio, P. M., Aquiloni, L., Banha, F., Choquer, J., Chucholl, C., et al. (2016). The Red Swamp Crayfish Procambarus clarkii in Europe: Impacts on Aquatic Ecosystems and Human Well-Being. Limnologica 58, 78–93. doi:10.1016/j.limno.2016.03.003

Tan, M. H., Austin, C. M., Hammer, M. P., Lee, Y. P., Croft, L. J., and Gan, H. M. (2018). Finding Nemo: Hybrid Assembly with Oxford Nanopore and Illumina Reads Greatly Improves the Clownfish (Amphiprion ocellaris) Genome Assembly. GigaScience 7 (3), 1–6. doi:10.1093/gigascience/gix137

Tan, M. H., Gan, H. M., Gan, H. Y., Lee, Y. P., Croft, L. J., Schultz, M. B., et al. (2016). First Comprehensive Multi-Tissue Transcriptome of Cherax quadricarinatus (Decapoda: Parastacidae) Reveals Unexpected Diversity of Endogenous Cellulase. Org. Divers. Evol. 16 (1), 185–200. doi:10.1007/s13127-015-0237-3

Tan, M. H., Gan, H. M., Lee, Y. P., Grandjean, F., Croft, L. J., and Austin, C. M. (2020a). A Giant Genome for a Giant Crayfish (Cherax quadricarinatus) with Insights into Cox1 Pseudogenes in Decapod Genomes. Front. Genet. 11, 201. doi:10.3389/fgene.2020.00201

Tang, B., Zhang, D., Li, H., Jiang, S., Zhang, H., Xuan, F., et al. (2020). Chromosome-level Genome Assembly Reveals the Unique Genome Evolution of the Swimming Crab (Portunus Trituberculatus). Gigascience 9 (1), 1–10. doi:10.1093/gigascience/giz161

Tataro-Graovac, M., and Chen, N. (2009). Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. Curr. Protoc. Bioinformatics 25 (1), 4–14. doi:10.1002/0471250953.bi0410s25

Ventura, T., Stewart, M. J., Chandler, J. C., Rotgans, B., Elizur, A., and Hewitt, A. W. (2019). Molecular Aspects of Eye Development and Regeneration in the Australian Redclaw Crayfish, Cherax quadricarinatus. Aquacult. Fish. 4 (1), 27–36. doi:10.1016/j.aaf.2018.04.001

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakhikumar, S., et al. (2014). Pilon: an Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS one 9 (11), e112963. doi:10.1371/journal.pone.0112963

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Kloutchnikov, G., et al. (2017). BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol. Biol. Evol. 35 (3), 543–548. doi:10.1093/molbev/msx319

Weinländer, M., and Füreder, L. (2012). Associations between Stream Habitat Characteristics and Native and Alien Crayfish Occurrence. Hydrobiologia 693 (1), 237–249. doi:10.1007/s10750-012-1125-x

Whitlege, G. W., and Rabeni, C. F. (1997). Energy Sources and Ecological Role of Crayfishes in an Ozark Stream: Insights from Stable Isotopes and Gut Analysis. Can. J. Fish. Aquat. Sci. 54 (11), 2555–2563. doi:10.1139/f97-173

Wingfield, M. (2008). An Updated Overview of the Australian Freshwater Crayfish Farming Industry. Freshw. Crayfish 16, 15–18. doi:10.5869/fc.2008.v16.15

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: a Meta Server for Automated Carbohydrate-Active Enzyme Annotation. Nucleic Acids Res. 46 (W1), W95–W101. doi:10.1093/nar/gky418

Zhang, X., Yuan, J., Sun, Y., Li, S., Gao, Y., Yu, Y., et al. (2019). Penaeid Shrimp Genome Provides Insights into Benthic Adaptation and Frequent Molting. Nat. Commun. 10 (1), 356. doi:10.1038/s41467-018-08197-4

Zhu, B.-H., Xiao, J., Xue, W., Xu, G.-C., Sun, M.-Y., and Li, J.-T. (2018). Annotation. dbCAN2: a Meta Server for Automated Carbohydrate-Active Enzyme Annotation. Nucleic Acids Res. 46 (W1), W95–W101. doi:10.1093/nar/gky418

Conflict of Interest: Author HMG was employed by company GeneSEQ.

The remaining 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.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.