Org Divers Evol (2016) 16:185–200
DOI 10.1007/s13127-015-0237-3

ORIGINAL ARTICLE

ORGANISMS DIVERSITY & EVOLUTION

First comprehensive multi-tissue transcriptome of *Cherax quadricarinatus* (Decapoda: Parastacidae) reveals unexpected diversity of endogenous cellulase

Mun Hua Tan[1,2] · Han Ming Gan[1,2] · Huan You Gan[1,2] · Yin Peng Lee[1,2] · Larry J. Croft[1,2,3] · Mark B. Schultz[4] · Adam D. Miller[5,6] · Christopher M. Austin[1,2]

Received: 10 June 2015 / Accepted: 2 September 2015 / Published online: 11 September 2015

Abstract The Australian freshwater crayfish species, *Cherax quadricarinatus* Von Martens, 1868, is an important commercial and invasive species that is also being increasingly used as a model organism to address important and interesting questions in crustacean biology. Through deep sequencing of the transcriptome of *C. quadricarinatus* from the hepatopancreas and four other tissues, we examine the evolution of endogenously transcribed cellulase genes and provide new insights into controversial issues regarding the nutritional biology of crayfishes. A cluster assembly approach yielded one of the highest quality transcriptome assemblies for a decapod crustacean to date. A total of 206,341,872 reads with an average read length of 80 bp were generated from sequencing the transcriptomes from the heart, kidney, hepatopancreas, nerve, and testis tissues. The assembled transcriptome contains a total of 44,525 transcripts. A total of 65 transcripts coding for carbohydrate-active enzymes (CAZy) were identified based on hidden Markov model (HMM), and a majority of them display high relative transcript abundance in the hepatopancreas tissue, supporting their role in nutrient digestion. Comprehensive phylogenetic analyses of proteins belonging to two main glycosyl hydrolase families (GH9 and GH5) suggest shared ancestry of *C. quadricarinatus* cellulases with other characterized crustacean cellulases. Our study significantly expands the number of known crustacean-derived CAZy-coding transcripts. More importantly, the surprising level of evolutionary diversification of these proteins in *C. quadricarinatus* suggests that these enzymes may have been of critical importance in the adaptation of freshwater crayfishes to new plant-based food sources as part of their successful invasion of freshwater systems from marine ancestors.

Keywords Freshwater crayfish · Genomics · Pancrustacea · Cellulase

Electronic supplementary material The online version of this article (doi:10.1007/s13127-015-0237-3) contains supplementary material, which is available to authorized users.

✉ Han Ming Gan
gan.han.ming@monash.edu

[1] School of Science, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway, 47500 Petaling Jaya, Selangor, Malaysia

[2] Monash University Malaysia Genomics Facility, Jalan Lagoon Selatan, Bandar Sunway, 47500 Petaling Jaya, Selangor, Malaysia

[3] Malaysian Genomics Resource Centre Berhad, Boulevard Signature Office, Kuala Lumpur, Malaysia

[4] Faculty of Medical and Dental Health Sciences, The University of Melbourne, Bio21 Research Institute, 30 Flemington Rd, Parkville, Victoria 3010, Australia

[5] School of BioSciences, The University of Melbourne, Parkville, Victoria 3010, Australia

[6] Nature Glenelg Trust, Warrnambool, Victoria 3280, Australia

Introduction

The ecological role and nutritional biology of crayfishes has been controversial (Momot 1995) and is of resurgent interest given the increasing number and spread of invasive freshwater crayfish species and their potential to cause negative ecological impacts (James et al. 2014; Lodge et al. 2012). While many crustaceans are detritivorous or herbivorous, traditional views maintain that they, like all animals, are unable to directly digest plant material as they lack the genes for endogenous cellulase production, which are otherwise widespread in the microbial world. However, it is now emerging that cellulase genes have an ancient origin in animals and are widespread in

the Metazoa (Davison and Blaxter 2005). As a result, there has been an increase in research efforts to catalogue the presence of cellulase genes in animals and understand the complex enzymology of cellulose hydrolysis for biotechnology applications, which is benefiting from increased comparative genetic information (Chen 2015). Recent research on cellulase has generated rapid growth in the discovery of cellulase genes exemplified by the increase of GH5 sequences from approximately 2300 sequences (Aspeborg et al. 2012) to close to 5000 sequences in the current carbohydrate-active enzyme (CAZy) database (http://www.cazy.org/) (Lombard et al. 2014). In this context, an especially noteworthy finding is by Byrne et al. (1999) who discovered the occurrence of endogenous GH9 cellulase in *Cherax quadricarinatus*, the first report of this digestive enzyme in crustaceans. This work was then followed by the characterization of this gene in *C. quadricarinatus* by Crawford et al. (2004). While this led to the demonstration of a functional cellulase in crayfishes, analyses based on traditional methods involving cDNA constructs are not exhaustive, resulting in a lack of information pertaining to the potential diversity of cellulase genes in crayfishes. The increasing availability of new approaches to sequencing offers an alternative approach to investigating cellulase diversity in crayfishes and crustaceans more generally.

Next-generation sequencing technology provides opportunities to characterize the genomic composition of individuals, populations, and species in unprecedented detail. It is now possible to study patterns of genic diversity across the genome and to address a variety of questions relating to systematic and comparative genomic research (Amemiya et al. 2013; Eaton and Ree 2013; McCormack et al. 2013; Oakley et al. 2013). Comparative genomic applications are diverse and range from investigations of specific genes or gene families all the way through to whole genomes (Alföldi and Lindblad-Toh 2013), providing insights into gene function, evolutionary mechanisms and adaptive genetic variation, and the co-evolution of disease organisms, parasites, and their hosts (Hayward et al. 2015; Jarvis et al. 2014; You et al. 2014; Ward and Kellis 2012; Rubin et al. 2012; Zhernakova et al. 2010).

Yet, despite the significant advances in sequencing technology and bioinformatics, genomic research for many animal groups remains limited by a lack of available genomic data. Of the 540 species listed on GenBank with a significant level of genomic resources, 54 % are vertebrates, and for the non-vertebrates, resources are dominated by taxa of medical or agricultural importance, predominantly hexapods and nematode worms, many of which are used as model organisms. Thus, major groups of invertebrates such as the annelids, molluscs, and crustaceans are poorly represented, despite their rich and extensive fossil records and major contributions to metazoan diversity and the ecology of many of the earth's major ecosystems (Scientists 2014). Until these groups are better represented in genomic databases, it will not be possible to properly understand the evolutionary history of the animal kingdom and identify major elements of genomic evolution and adaptation within lineages, including the diversification and characteristics of protein families of biological or biomedical significance.

In recent years, the collection and assembly of transcriptomic data has become faster and less costly than whole genome sequencing (WGS) and can provide substantial datasets for studies of gene expression, quantitative trait variation as well as for comparative genomic and phylogenetic purposes. However, the gene coverage and quality of transcriptomic datasets can vary widely especially when only a single tissue type is analyzed. Optimally, transcriptome sampling from different tissues and developmental stages of an organism sequenced at deep coverage helps to ensure the recovery of a comprehensive transcriptome profile (Clarke et al. 2014; Pallavicini et al. 2013).

The crustaceans, including the highly diverse, ecologically and economically important decapods (shrimps, crabs, crayfishes, and lobsters), are notably lacking in their representation on genomic databases. Complete genomes are available for only one branchiopod (*Daphnia pulex*), two copepods (*Lepeophtheirus salmonis*, *Eurytemora affinis*), and one amphipod (*Hyalella azteca*) (Colbourne et al. 2011; Consortium 2013). Although several crustacean transcriptomic datasets are available, these are highly variable in terms of tissue source, data output, assembly methods, sequencing platforms, and, therefore, gene coverage and data quality (Lv et al. 2013; Ma et al. 2014; Sahoo et al. 2013). For a group containing more than 50,000 species, less than 50 crustacean transcriptomic datasets are deposited in public databases at the time of this study.

A major crustacean group, as they have traditionally been considered, is the Malacostraca containing the order Decapoda, which is both highly diverse and contains all the significant crustacean seafood species. Within the Decapoda, the freshwater crayfishes are a conspicuous group that has a geographic range encompassing all major land masses with the exception of Africa and Antarctica (Bracken-Grissom et al. 2014; Porter et al. 2005; Toon et al. 2010). They reach their greatest diversity in the south eastern USA and south-eastern Australia and are considered to have a monophyletic origin having evolved from marine ancestors related to lobsters (Nephropoidea) during the Permian or Triassic (Porter et al. 2005; Martin et al. 2008).

In Australia, *Cherax* is one of the most familiar genera of freshwater crayfishes, commonly known as smooth yabbies, containing several large and economically significant species (Austin 1996; Austin and Knott 1996). The best known species within *Cherax* is the red claw crayfish, *C. quadricarinatus*, endemic to northern Australia and southern New Guinea (Austin 1996). Unfortunately, due to extensive translocation resulting from its popularity as an

aquaculture and ornamental species and its adaptability, it now also has an extensive global distribution and is acknowledged as a major invasive species of inland aquatic ecosystems in the tropics (Ahyong and Yeo 2007; Saoud et al. 2013; Larson and Olden 2012). This species is also being increasingly used as a model organism to address questions of fundamental interest to crustacean evolution and biology, including the presence and role of cellulase genes (Davison and Blaxter 2005), chitin metabolism and molting (Fernández et al. 2012; Pamuru et al. 2012), sex determination and reproduction (Sagi and Khalaila 2001; Shechter et al. 2005), and for infection-related studies (Hayakijkosol et al. 2011; Nguyen et al. 2014).

The genetics of Australian freshwater crayfishes, including *Cherax* and *C. quadricarinatus*, have been well studied using PCR-based methods and Sanger sequencing. To our knowledge, other than mitogenomic data (Gan et al. 2014), there are no comprehensive genomic studies on *C. quadricarinatus*, although there are several small-scale and focused studies that have been carried out investigating specific nuclear gene sequences in the species (Abdu et al. 2002; Fang et al. 2011; Liu et al. 2011; Yudkovski et al. 2010). In addition to these studies, the *C. quadricarinatus* transcriptome from gastrolith-forming and subticular epithelia has been sequenced using 454 pyrosequencing and assembled (Glazer et al. 2013); however, this represents only a modest dataset. In this study, we present a comprehensive transcriptomic analysis of *C. quadricarinatus* sequenced at high coverage and assembled based on sequences from five tissues. We use our *C. quadricarinatus* data to significantly extend the studies of Byrne et al. (1999) and Davison and Blaxter (2005) on genes from cellulase families to further explore the evolution of these interesting and important gene families and report an unexpectedly high level of diversification of these genes in *C. quadricarinatus*. Throughout this paper, the term "cellulase" can also be used to refer to enzymes in the cellulase family, which also includes other enzymes such as mannanase that are capable of digesting other plant-derived material.

Material and methods

Sequencing of the *C. quadricarinatus* transcriptome

A sample of *C. quadricarinatus* was obtained from Rapid Creek in Darwin, Australia. Approximately 50 mg of heart, kidney, hepatopancreas, central nerve cord, and testis tissues were dissected from the euthanized crayfish and immediately homogenized in 600 μL of RNA shield (Zymo Research, Irvine, CA). Total RNA extraction was performed on the homogenate using Quick-RNA MicroPrep (Zymo Research, Irvine, CA) and stored at −80 °C. Poly-A enrichment and library preparation were done using NEXTflex Poly(A) beads and NEXTflex RNA-Seq kit respectively (Bioo Scientific, Austin, TX) according to the manufacturer's instructions. The concentration and fragment size of the libraries were determined using a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). Then, all five libraries were normalized, pooled, and sequenced three times on the MiSeq (2× 80 bp run) (Illumina, San Diego, CA).

De novo assembly of the transcriptome

The workflow for assembly, annotation, and analysis of the transcriptome is summarized in Fig. 1. Paired-end transcriptome reads from the heart, kidney, hepatopancreas, nerve, and testis tissues were pre-processed with Trimmomatic v.0.32 (Bolger et al. 2014) by trimming low-quality bases (reads were trimmed if an average quality fell below 20 within a window size of four bases) and retaining reads of at least 50 bp in length. Reads were then aligned to complete *C. quadricarinatus* 18S [GenBank: AF235966.1] and 28S (Online Resource 1) ribosomal RNA (rRNA) gene sequences with Bowtie2 v.2.1.0 (Langmead and Salzberg 2012), and those which mapped perfectly to these rRNA sequences were excluded. The remaining reads from all tissues were pooled and assembled de novo using two independent transcriptome assembler programs: Trinity v.20140717 (Grabherr et al. 2011) and IDBA-tran v.1.1.1 (Peng et al. 2013). Bowtie2 v.2.1.0 (Langmead and Salzberg 2012) was used to multi-map reads to each assembly separately before clustering the transcripts from both assemblers with Corset v.1.03 (Davidson and Oshlack 2014). Subsequently, the longest transcript was selected to represent each cluster, and only sequences longer than 200 bp were retained for further analyses. These representative transcripts were translated into amino acid sequences using TransDecoder v.20140704 (http://transdecoder.sourceforge.net/) while including PFAM domain searches. Finally, CEGMA v.2.4 (Parra et al. 2007) was used to detect the presence of a subset of core eukaryotic proteins in the assembled transcriptome.

Transcriptome annotation and transcript abundance analysis

For functional annotation, transcripts were aligned to protein sequences from the Swiss-Prot database as well as NCBI's non-redundant (NR) protein database using BLASTX v.2.2.29+ (Altschul et al. 1990) at an e-value threshold of $1e^{-10}$. In addition, translated protein sequences were searched against various databases using InterProScan v.5.8–49.0 (Jones et al. 2014) to further identify protein families, domains, and functional sites. Gene ontology (GO) terms were also assigned from the InterProScan results.

Pre-processed paired-end reads from individual tissues were mapped to the assembled transcriptome using Bowtie v.0.12.7 (Langmead et al. 2009), and the RSEM v.1.2.19 software package (Li and Dewey 2011) was used to assign reads to genes in order to calculate the relative transcript abundance

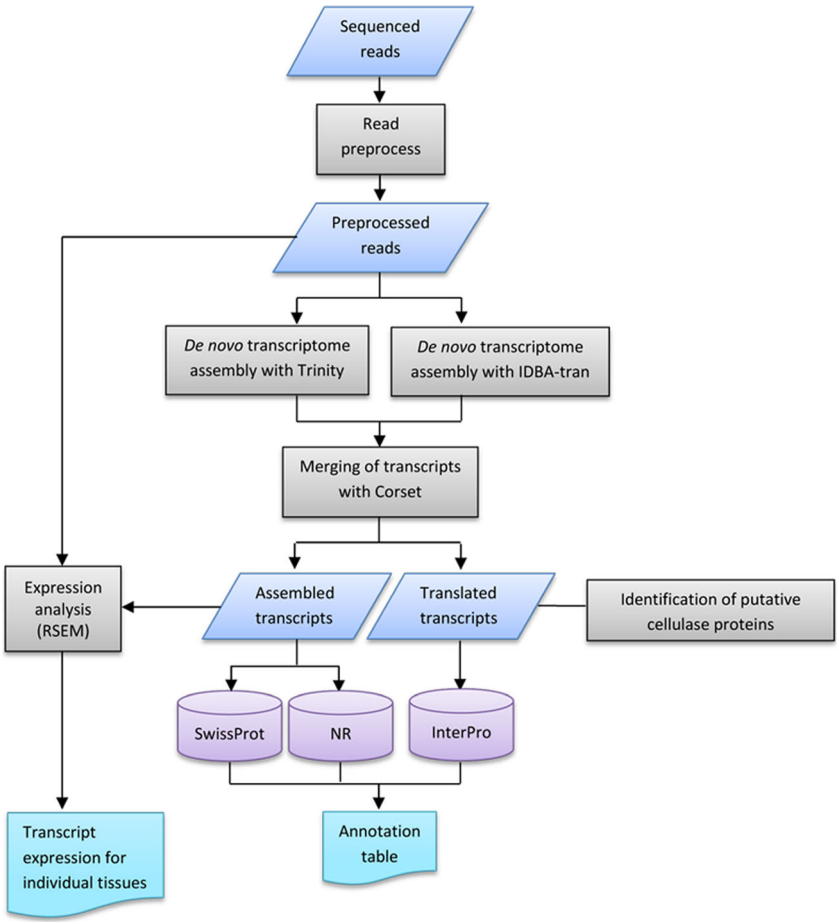


Fig. 1 Workflow diagram for the analysis of the *Cherax quadricarinatus* transcriptome. Analyses included the de novo assembly of the transcriptome, annotation of transcripts, expression analysis in multiple tissues, and the identification of putative cellulases

in each tissue type. Transcripts per million (TPM) values were used as expression values. The overlap between all tissue types was analyzed by identifying common and unique transcripts expressed at a TPM threshold of 100 in each tissue. A Venn diagram representing tissue-specific and overlapping transcripts was generated using the "VennDiagram" R package v.1.6.0 (Chen and Boutros 2011). The expression of second-level GO terms relating to biological process, cellular component, and molecular function categories were inspected by summing up the TPM expression value of transcripts pertaining to the specific second-level GO term. A detailed examination of the expression of third-level GO terms in the biological process category was also analyzed as above. Heat maps to represent the relative expression for all tissues were drawn using the "pheatmap" R package v.0.7.7 (Kolde 2012).

Identification of genes encoding CAZy enzymes and an exploration of cellulase genes in GH9 and GH5 and their evolution

To identify putative cellulase genes within the *C. quadricarinatus* dataset, protein sequences were queried against data from the CAZy database (http://www.cazy.org/) (Lombard et al. 2014) through the dbCAN web server (http://csbl.bmb.uga.edu/dbCAN/) (Yin et al. 2012). Sequences with hits to a GH family with its highest expression in any tissue exceeding TPM of 10 were retained. The pheatmap R package v.0.7.7 (Kolde 2012) was used to plot a heat map of relative expression for all tissues.

For each GH9 and GH5 family, cellulase genes from three domains, Bacteria, Archaea, and Eukarya (limited to kingdoms Viridiplantae, Fungi, and Metazoa), were extracted from CAZy (Lombard et al. 2014). To reduce redundancy, highly similar proteins for a same species were clustered with CD-HIT v. 4.6 (W. Li et al. 2001) (99 % identity, word length of 5). Remaining sequences were aligned with MAFFT v.7.127b (Katoh and Standley 2013) using the *L-INS-i* accurate option, and a tree was inferred from this large alignment using FastTree v.2.1.7 (Price et al. 2010). Using the *trim_tips.py* script from Yang and Smith (2014), unusually long tips were removed by discarding a tip if it was longer than an absolute length (0.6 for GH9, 1.1 for GH5) or was more than ten times longer than tips in its sister clade. Unrooted circular phylograms for both GH families were annotated and visualized with GraPhlAn v.0.9.7 (http://huttenhower.sph.harvard.edu/graphlan).

Subtrees containing sequences of interest, particularly metazoan GH9 proteins and members of subfamily 10 of

GH5 (GH5_10), were inferred separately. Sequences longer than 100 amino acids were aligned with MAFFT v.7.127b (Katoh and Standley 2013) using the *L-INS-i* accurate option, and ambiguously aligned regions were trimmed with BMGE v.1.1 (Criscuolo and Gribaldo 2010). The best model of protein evolution was determined with ProtTest3 v.3.4 (Darriba et al. 2011; Guindon and Gascuel 2003). A maximum-likelihood tree for each of these subtrees was estimated under the WAG + G model using RAxML v.8.1.5 (Stamatakis 2006), and nodal support was inferred with standard bootstrap replicates (with *autoMRE* criteria, 500 replicates for GH5 and 450 replicates for GH9).

All putative GH9 sequences in addition to the two public *C. quadricarinatus* protein sequences: AAD38027.1 (Byrne et al. 1999) and AAO61672.2 (Crawford et al. 2004), were aligned in a pairwise manner using SDT v.1.2 (Muhire et al. 2014) (MAFFT was selected as aligner). Multiple sequence alignments of these sequences and several GH9 cellulases from other phyla were also carried out with MAFFT v.7.127b (Katoh and Standley 2013) to identify conserved functional regions and active sites which were identified by Davison and Blaxter (2005).

Results

Assembly and annotation of *C. quadricarinatus* transcriptome reveals the paucity of publicly available genetic resources for crustaceans

A total of 206,341,872 reads with an average read length of 80 bp was generated from sequencing the transcriptomes from the heart, kidney, hepatopancreas, nerve, and testis tissues at fragment sizes of 380, 376, 368, 388, and 384 bp, respectively. Subsequently, all raw reads were submitted to the Sequence Read Archive [SRA: ERR391748–ERR391752]. Following the pre-processing of reads, 188,234,126 reads (91.22 %) were retained for assembly. A summary of read statistics for individual tissues is reported in Online Resource 1.

The Trinity-assembled transcriptome contained 105,211 transcripts [TSA: HACB2000001–HACB2105211] while 75,424 transcripts were assembled by IDBA-tran [TSA: HACK01000001–HACK01075424]. Corset grouped these transcripts into 44,525 clusters with the longest transcript selected as the representative for each cluster. The final set of representative transcripts (see Online Resource 2 for sequence file) resulted in a total transcriptome size of 71,428,145 bp with an N50 length of 2383 bp (see Table 1 for assembly statistics and Fig. 2a for distribution of transcript lengths). Out of the 248 highly conserved core eukaryotic proteins in the CEGMA gene set, 235 complete proteins (94.76 %) were represented in the transcriptome.

A total of 16,623 transcripts (37.33 %) had annotations from at least one database (NR: 15,664 transcripts, Swiss-Prot: 11,177 transcripts, InterPro: 11,570 transcripts). Annotations for all transcripts are listed in Online Resource 1. The top ten species hits from the NR database are also shown in Online Resource 1, with the termite *Zootermopsis nevadensis* as the top hit, and *D. pulex* and *C. quadricarinatus* were the only crustacean species on this list. Less than 1 % (408) of the transcripts were annotated to *C. quadricarinatus* sequences, reflecting the paucity of protein sequences available for freshwater crayfishes and crustaceans in general in the NR database. A total of 8570 transcripts were associated with at least one GO term from the three main GO categories with 1701 being unique GO terms. Of these, 7765 transcripts were assigned to a molecular function category, 4427 transcripts in the biological process category, and 2479 transcripts in the cellular component category.

The longest assembled transcript from this analysis was 32,959 bases long which translated into a protein of 8431 amino acids in length and appeared to be complete with a start and stop codon. This transcript was mainly expressed in the heart and testis with TPM values of 48.51 and 65.32, respectively. Annotation results against protein databases showed the highest sequence homology of this transcript to the *Procambarus clarkii* I-connectin protein [NR: BAB64297.1,

Table 1 Summary of assembly statistics for the *Cherax quadricarinatus* transcriptome

Summary	Trinity assembly	IDBA-Tran assembly	Clustered (longest transcript selected as representative)
Number of transcripts	105,211	75,424	44,525
Total transcriptome size (bp)	91,952,851	108,990,202	71,428,145
Longest transcript length (bp)	19,914	32,959	32,959
Shortest transcript length (bp)	201	300	201
Mean transcript length (bp)	874	1445	1604
Median transcript length (bp)	411	739	1018
Transcript N50 length (bp)	1697	2612	2383
% GC	41.27	41.50	40.72

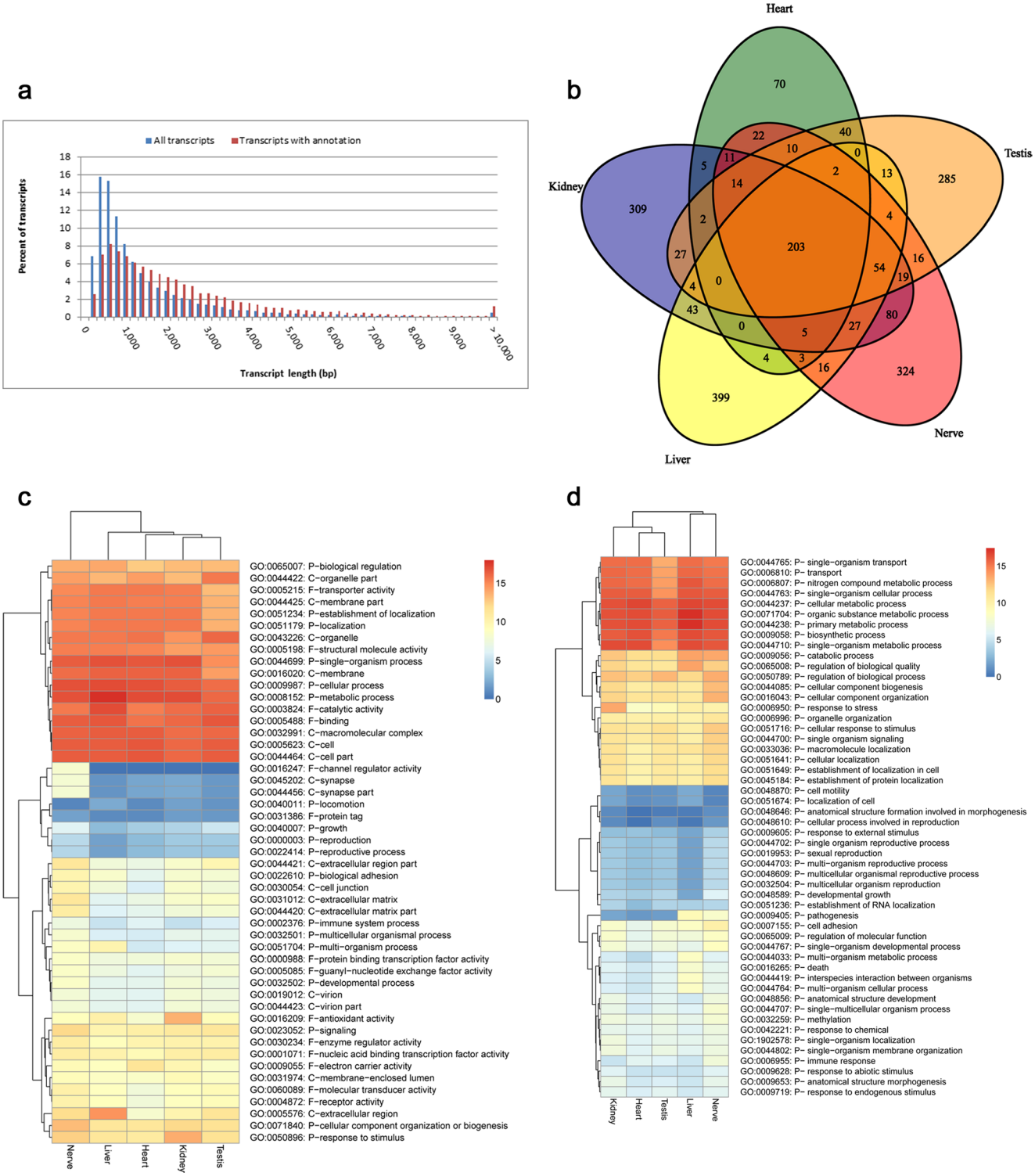


Fig. 2 Annotation and information on the *Cherax quadricarinatus* transcriptome. **a** Distribution of transcript lengths with *blue bars* representing the lengths of all transcripts and *red bars* showing the length distribution only for transcripts with at least one annotation. **b** Venn diagram showing the number of shared and unique transcripts in different tissue types expressed at a TPM threshold of 100. **c** Transcript expression according to gene ontology (GO) terms in different tissues measured in $\log_2$ of (TPM + 1) values. *Horizontal categories* represent second-level GO terms in three categories namely biological process (P), cellular component (C), and molecular function (F). **d** Same expression analysis as in **c**, but with *horizontal categories* representing third-level GO terms contained in the biological process category specifically

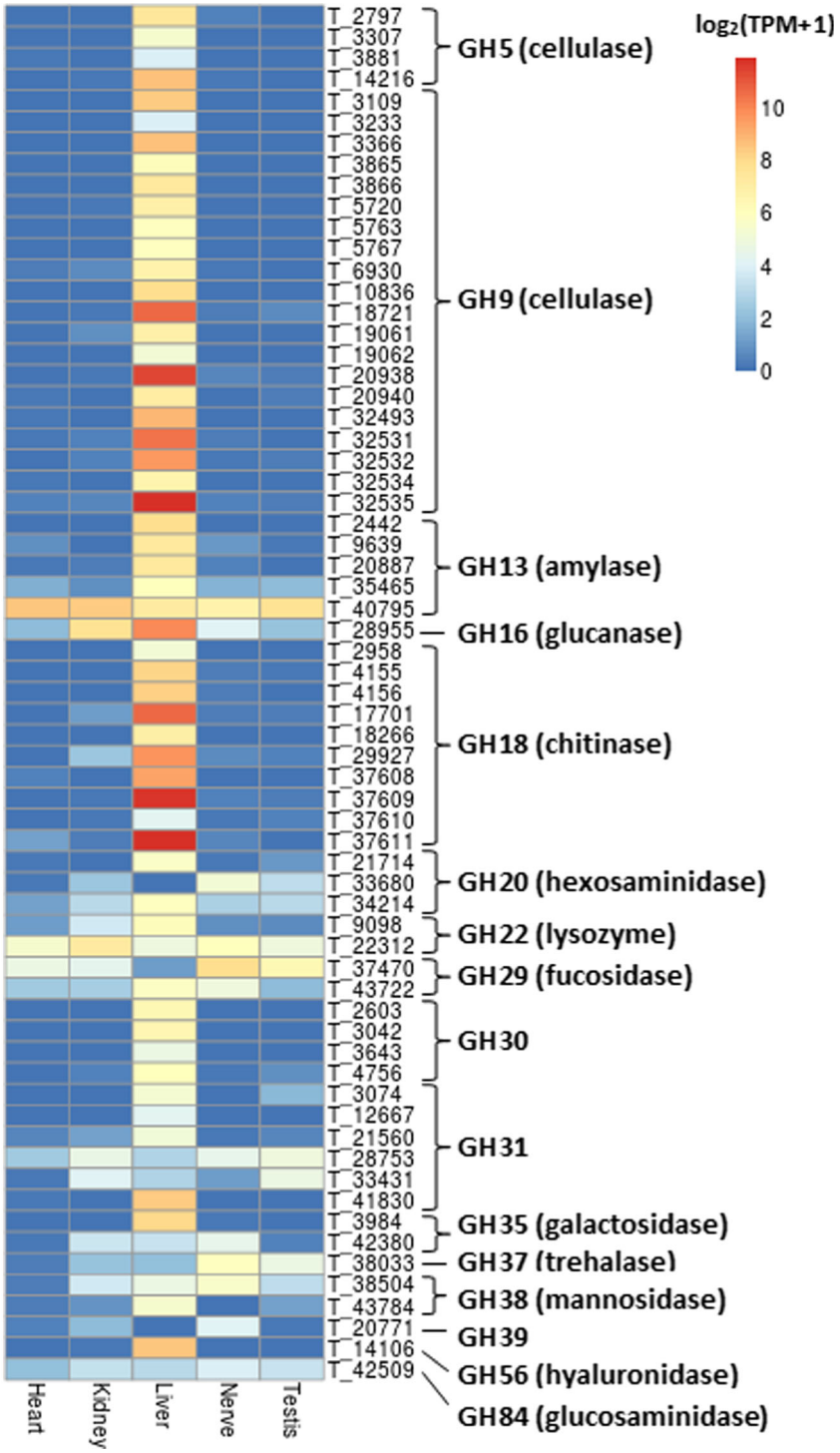


Fig. 3 Expression of 65 transcripts categorized into 16 GH families measured in $\log_2$ of (TPM + 1) values. The descriptions of GH families were assigned based on InterPro annotations of transcripts in each family (Online Resource 3), providing higher confidence of its putative function. In contrast, GH30, GH31, and GH39 do not have descriptions as their corresponding transcripts do not possess detectable domains or signatures that point to any enzymatic functions

length 17,352aa] and the *Drosophila melanogaster* TITIN protein [Swiss-Prot: Q9I7U4, length 18,141 aa]. Multiple immunoglobulin (Ig) and immunoglobulin-like signatures [InterPro: IPR013783, IPR007110, IPR013098, IPR003599, IPR003598] were also detected throughout the translated protein (Online Resource 1).

Transcriptome sequencing of multiple tissues improves transcript representation of *C. quadricarinatus*

The abundance of transcripts was represented by TPM values calculated from counts of reads mapped against the transcripts. The percentage of reads mapped to the total transcripts

for heart, kidney, hepatopancreas, nerve, and testis tissues were 50.81, 60.30, 60.36, 61.64, and 55.66 %, respectively. From this analysis, it was apparent that a higher diversity of genes was expressed in the nerve (38,001 genes), testis (37,985 genes), and kidney (37,401 genes) compared to hepatopancreas (34,649 genes) and heart (30,936 genes). The top ten most expressed transcripts (excluding mitochondrial transcripts) in each tissue are listed in Online Resource 1. It is noteworthy that, compared to other tissues, there were more highly expressed transcripts in the kidney, nerve, and testis that did not match any protein sequences in the NR or Swiss-Prot databases.

The overlap of transcripts between tissues expressed at a TPM threshold of 100 is summarized in a Venn diagram in Fig. 2b. A total of 203 highly expressed transcripts, 162 (79.80 %) of which were annotated to known protein functions or domains, were found in all five tissues and listed in Online Resource 1, representing the core transcriptome expressed in all tissues. A significant portion of these core transcripts matched to 60S (19.70 %) and 40S (13.30 %) ribosomal proteins, while most of the remaining transcripts were associated with proteins involved in energy production, cell regulation, and maintenance or were encoded on the mitochondrion. The bar charts in Online Resource 1 show the distribution of tissue-specific transcripts according to the top ten second-level GO categories at a TPM threshold of 100.

Comparison of top ten GO term distributions between tissue-specific transcripts and transcripts shared among all tissues expressed above TPM thresholds of 100 showed the differences in GO profiles for various tissues. The major GO terms found in all tissue-specific transcripts include cell, cell part, and cellular process (see bar chart (f) in Online Resource 1). Transcripts found specific to the hepatopancreas tissue appeared to be more functionally related to catalytic activity and metabolic process compared to other tissues (see bar chart (c) in Online Resource 1) while transcripts specific to the heart or nerve tissue showed a wide variety of GO distributions since a large portion of transcripts (44 % in both tissues) were associated with GO terms other than the listed top ten ontologies as seen in Online Resource 1 (see bar chart (d)).

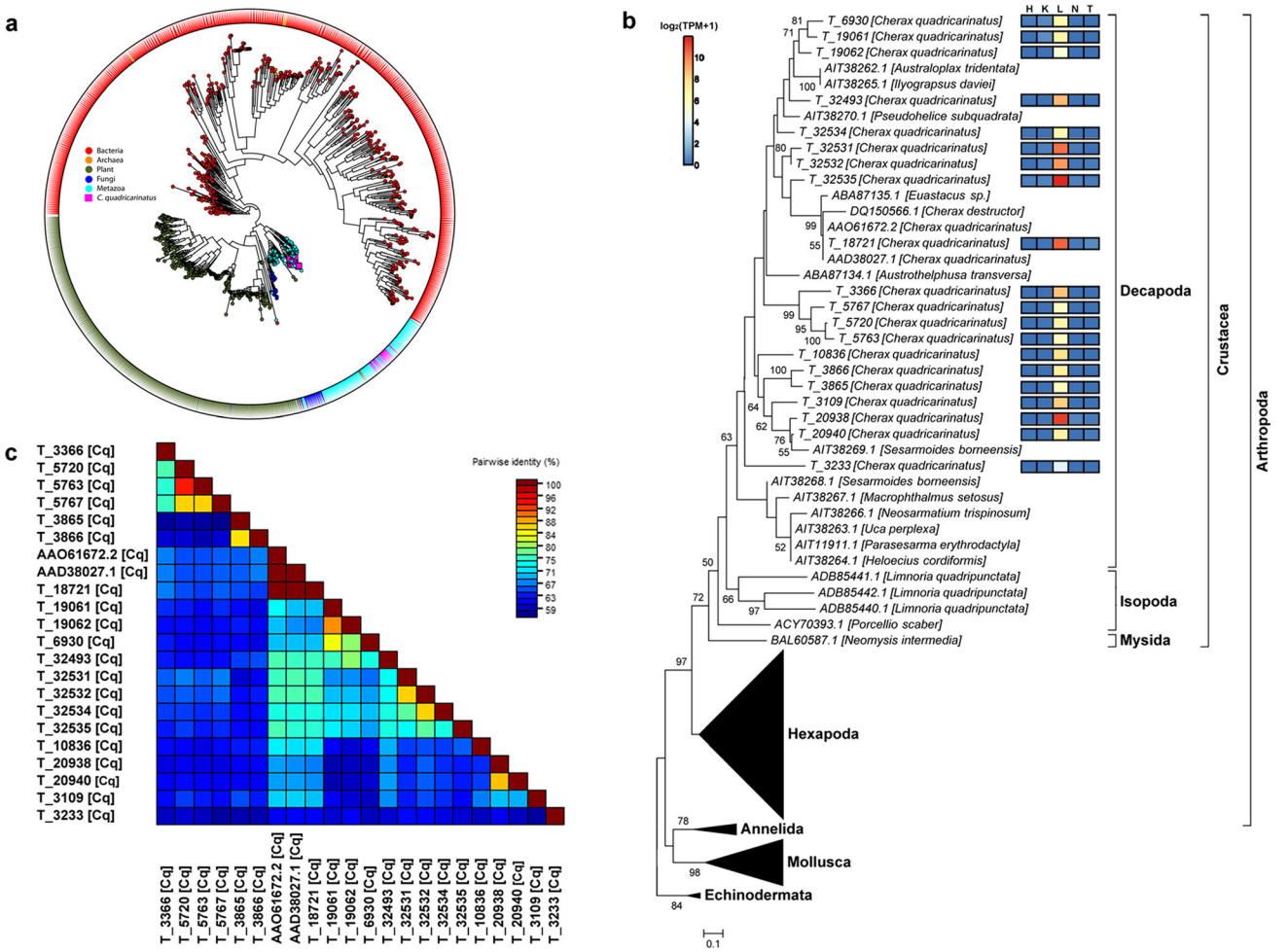
Expressions of transcripts according to second- and third-level GO terms in each tissue are illustrated in Fig. 2c, d. Overall, several GO terms such as cellular process [GO: 0009987], metabolic process [GO: 0008152], single-organism process [GO: 0044699], cell component [GO: 0005623], macromolecular complex component [GO: 0032991], organelle component [GO: 0043226], binding function [GO: 0005488], and catalytic activity function [GO: 0003824] showed relatively high expression across all tissues. In addition, some GO terms were expressed in relatively higher or lower values in specific tissues that are consistent with the respective tissue functions. For example, out of all tissues, the kidney contained higher expression of transcripts assigned to respond to stimulus [GO: 0050896] and antioxidant activity [GO: 0016209] while the nerve showed higher expression of transcripts associated with synapse [GO: 0045202] and channel regulator activity [GO: 0016247].

C. quadricarinatus expresses several putative CAZy transcripts with predominant abundance in the hepatopancreas

A total of 65 transcripts were categorized into 16 different glycoside hydrolase (GH) families (Fig. 3). InterPro annotations for each representative protein sequence are available in Online Resource 3 and mostly support the enzymatic functions listed in the figure. Most noteworthy are the cellulase (GH9, GH5) and chitinase (GH18), with high expression of transcripts predominantly in the hepatopancreas. Protein sequences are available in Online Resource 4. The diversity of putative cellulases related to the GH9 and GH5 families were examined more thoroughly.

Phylogenetic analysis incorporating 20 putative *C. quadricarinatus* GH9 cellulase and 1300 sequences from the CAZy database shows that most GH9 proteins cluster monophyletically mainly based on taxonomic domains with a few outliers (Fig. 4a). As expected, our putative crayfish GH9 proteins (Online Resource 4) group with other metazoan sequences (available in Newick format as Online Resource 5). A second maximum likelihood (ML) phylogenetic tree that was constructed using metazoan-derived GH9 sequences and rooted with members of phylum Echinodermata (Fig. 4b) shows that the putative *C. quadricarinatus* GH9 cellulases cluster with sequences from other decapod species. The tree also shows the relative abundance of sequences reported for every phylum or subphylum, with Hexapoda having the largest representation. Nodal support for this tree is relatively weak, with many nodes having bootstrap values of less than 50 %. One of the putative GH9 proteins (T_18721) forms a tight group with two public *C. quadricarinatus* cellulases [GenBank: AAO61672.2, AAD38027.1] with high

Fig. 4 Putative *Cherax quadricarinatus* GH9 cellulases and the evolution of GH9 proteins. **a** Circular phylogram shows the clustering of GH9 proteins from three domains: Bacteria, Archaea, and Eukarya (plants, fungi, and metazoans only). Markers are colored according to taxonomic domains with the putative GH9 proteins colored in *magenta* (see legend). **b** Maximum-likelihood phylogenetic tree for metazoan GH9 proteins rooted with Echinodermata. Only bootstrap values greater than 50 % are shown at nodes. Expression levels in different tissues (*H:* heart, *K:* kidney, *L:* liver/hepatopancreas, *N:* nerve, *T:* testis) for individual *C. quadricarinatus* GH9 transcripts are also shown. **c** Color-coded pairwise identity matrix generated from 22 *C. quadricarinatus* proteins based on amino acid similarity. **d** Partial alignment of multiple GH9 proteins displaying conserved active sites as seen in Davison and Blaxter (2005). Putative GH9 sequences can be found in Online Resource 4, and complete sequence alignments is contained in Online Resource 6



d

Superkingdom/ Kingdom	Phylum	Species	Accession	252 258 261 268	599 604 611	649 654 663 666
Bacteria	Proteobacteria	*Myxobacter sp.*	AAF19168.1	LTGGWYDAGDHVKFGLP	PPKHPHHRTAHGS	DDSYRDETNDYVSMEVAI
Bacteria	Actinobacteria	*Cellulomonas fimi*	P26225.1	LTGGWYDAGDHVKFGFP	PPTAPHHRTAHGS	NDAYTDSRQDYVAMEVAT
Fungi	Neocallimastigomycota	*Piromyces sp.*	AAM81967.1	LTGGYYDAGDNVKFNFP	SPKAVHHRAASGT	KDEYTDSRKNYEMNEVAL
Amoebozoa	-	*Dictyostelium discoideum*	P22699.1	LSGGYFDAGDGVKFGLP	YPINPHHRAAHHS	NDEYTDDRTDYISNEVAT
Metazoa	Annelida	*Metaphire hilgendorfi*	BAH22180.1	LTGGWYDAGDHVKFGFP	PPQQPHHRSSSCP	NDNYNDARSDYISNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	AAO61672.2	LTGGYYDAGDHVKFGFP	PPTRPHHRSSSCP	DGSYNDDRQDYQHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	AAD38027.1	LTGGYYDAGDHVKFGFP	PPTRPHHRSSSCP	DGSYNDDRQDYQHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_10836	LTGGYYDAGDHVKFGFP	PPQRPHHRSSSCP	DGTYSDDRNDYVHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_18721	LTGGYYDAGDHVKFGFP	PPTRPHHRSSSCP	DGSYNDDRQDYQHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_19061	LTGGYYDAGDHVKFGFP	PPQKPHHRSSSCP	DGSYNDDRGDYQHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_19062	LTGGYYDAGDHVKFGFP	PPEKPHHRSSSCP	DGNYTDDRGDYQHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_20938	LTGGYYDAGDHVKFGFP	PPTRPHHRSSSCP	GDGYTDVRTDYQHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_20940	LTGGYYDAGDHVKFGFP	PPTRPHHRSSSCP	GDGYTDVRTDYQHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_3109	LTGGYYDAGDHVKFGFP	PPQRPHHRSSSCP	DDSYTDDRMDYTHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_3233	LTGGYFDAGDFVKFLLP	APERVHAAASCP	NDVFVDDREDYVHNEIAL
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_32493	LTGGYYDAGDHVKFGFP	PPQKPHHRSSSCP	DGSYNDDRGDYQHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_32531	LTGGYYDAGDHVKFGFP	PPQRPHHRSSSCP	DGTYNDDRQDYQHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_32532	LTGGYYDAGDHVKFGFP	PPQRPHHRSSSCP	SGDYNDDRQDYQHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_32534	LTGGYYDAGDHVKFGFP	PPQRPHHRSSSCP	DGSYTDDRHDYQHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_32535	LTGGYYDAGDHVKFGFP	PPEKPHHRSSSCP	DGSYVDDRQDYQHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_3366	LSGGYYDAGDLVKFGFP	YPERPHHRPSSCP	SGTFTDDRTLYEQNEVGL
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_3865	LTGGYYDAGDHVKFGFP	PPKRPHHRSSSCP	DDTYTDDRIDYIHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_3866	LTGGYYDAGDHVKFGFP	YPKRPHHRSSSCP	NDEYTDDRMDYVHNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_5720	LTGGYYDADDFVKFGFP	YPERPHHRSSSCP	NGDYEDDRSNYVQNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_5763	LTGGYYDADDFVKFGFP	YPERPHHRASSCP	NGDYEDDRSNYEQNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_5767	LTGGYYDAGDLVKYGFP	YPERPHHRASSCP	NGDYEDDRSNYEQNEVAC
Metazoa	Arthropoda	*Cherax quadricarinatus*	T_6930	LSGGYYDAGDRVKFGFP	PPVRPCHGSSACP	DGNYNDNRTDWQHNEVAC
Metazoa	Arthropoda	*Nasutitermes takasagoensis*	BAA33708.1	LTGGYFDAGDFVKFGFP	PPTRPHHRSSSCP	NDNYVDDRSDYVHNEVAT
Metazoa	Arthropoda	*Nasutitermes walkeri*	BAA33709.1	LTGGYFDAGDFVKFGFP	PPTRPHHRSSSCP	NDNYVDDRSDYVHNEVAT
Metazoa	Arthropoda	*Reticulitermes speratus*	BAA31326.1	LTGGYYDAGDFVKFGFP	PPVRPHHRSSSCP	NDSYTDARSDYISNEVAT
Metazoa	Arthropoda	*Panesthia cribrata*	AAF80584.1	LTGGYYDAGDFVKFGFP	YPTHESHRSSSCP	NDDYEDLRSDYVHNEVAD
Metazoa	Echinodermata	*Mesocentrotus nudus*	BAF62178.1	LTGGWYDAGDHVKFGLP	PPQRPHHRSSSCP	NDNYTDERSDYISNEVAC
Metazoa	Mollusca	*Haliotis discus*	BAC67186.1	LTGGWYDAGDHVKFSLP	YPRNPHHRSASCP	DDSYKDNREDYVHNEVAC

pairwise identity (>99 %) as shown in Fig. 4c. Pairwise identity scores show a large range in the degrees of amino acid sequence similarity (56.8–99.8 %) for putative GH9 homologs identified in *C. quadricarinatus*. Finally, the alignment of the 20 putative cellulases and GH9 cellulases from several species from other kingdoms and phyla revealed several conserved sites, and these are highlighted in the partial alignment in Fig. 4d (see Online Resource 6 for the complete alignment). Out of the 20 putative GH9 cellulases, 19 have the expected conserved amino acid at all five active sites. The remaining homolog (T_6930) has the expected conserved amino acid at only four sites but differs at position 604 of the alignment with a cysteine (C) instead of a histidine (H). The cockroach, *Panesthia cribrata*, also has a different amino acid, serine (S), at this position.

Homologs of putative cellulases belonging to GH5 families were also identified and are available in Online Resource 4. Four protein sequences were identified from the transcriptome data as putative GH5 cellulases. Combined with sequences available in the CAZy database, a total of 3447 protein sequences were used to plot the unrooted circular phylogram shown in Fig. 5a (available in Newick format as Online Resource 7). The circular phylogram contains sequences from 49 GH5 subfamilies, annotated according to the assignment in the CAZy database. Putative GH5 sequences identified from the crayfish transcriptome were clustered with proteins from the GH5_10 subfamily, which contains sequences from bacterial and other metazoan (arthropods and molluscs) sources. An additional ML analysis for members of the GH5_10 subfamily resulted in the unrooted phylogenetic tree in Fig. 5b which shows the clustering of metazoan sequences into a monophyletic group. All four putative *C. quadricarinatus* proteins form a monophyletic group together with the publicly available *C. quadricarinatus* partial mannanase protein sequence [GenBank: AIN40245.1], and one homolog (T_2797) appears to be highly similar to the aforementioned public sequence.

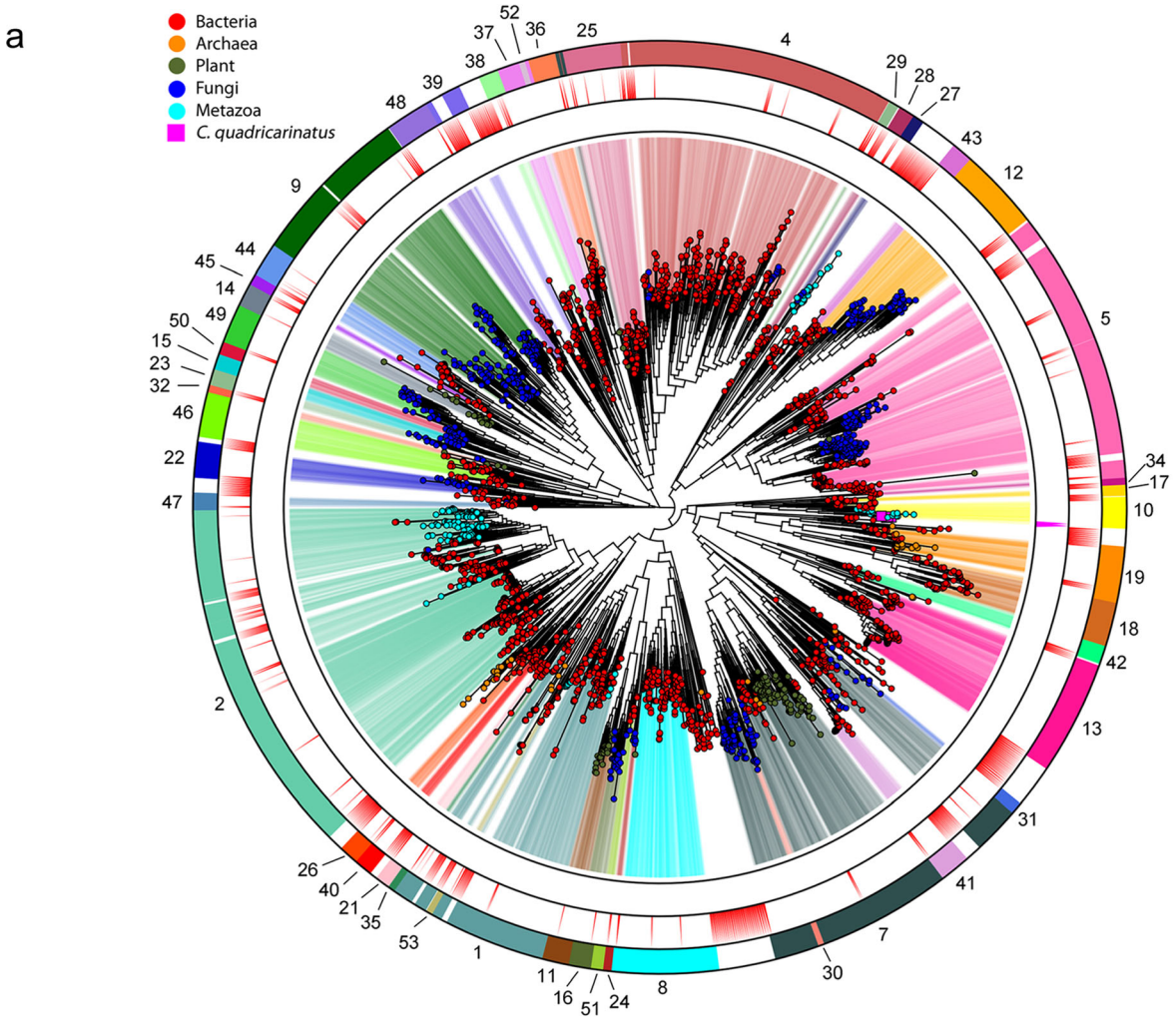
Fig. 5 Putative *Cherax quadricarinatus* GH5 cellulases and the evolution of GH5 proteins. **a** Circular phylogram shows the clustering of GH5 proteins from three domains: Bacteria, Archaea, and Eukarya (plants, fungi, and metazoa only) and the distribution of 49 GH5 subfamilies. Markers are colored according to taxonomic domains with the putative GH5 proteins colored in *magenta* (see legend). From inside: First ring marks the position of putative *C. quadricarinatus* GH5 proteins clustered in subfamily GH5_10. Second ring marks GH5 proteins not assigned to any subfamily in the CAZy database. Third ring shows the distribution of subfamilies in different colors. **b** Unrooted maximum-likelihood phylogenetic tree for proteins in the GH5_10 subfamily. Only bootstrap values greater than 50 % are shown at nodes. Expression levels in different tissues (*H:* heart, *K:* kidney, *L:* liver/hepatopancreas, *N:* nerve, *T:* testis) for individual *C. quadricarinatus* GH5 transcripts are also shown. Putative GH5 sequences can be found in Online Resource 4

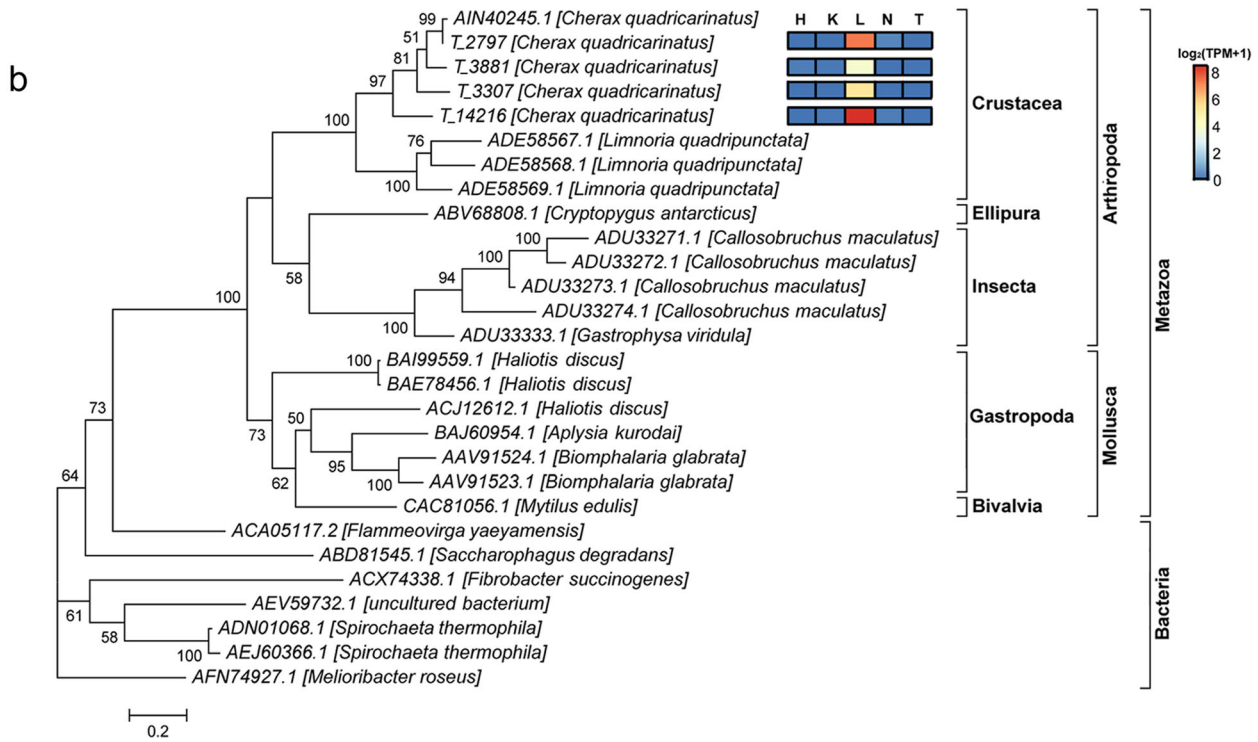
Discussion

Deep sequencing of transcriptomes from multiple distinct *C. quadricarinatus* tissues adds to the small but growing collection of high-quality transcriptomic and genomic resources for decapods and crustaceans more generally. While this is not the first *C. quadricarinatus* transcriptome sequenced (Glazer et al. 2013), our dataset is almost eightfold larger (based on total transcriptome size) and more comprehensive through higher coverage and the sampling of multiple tissues. Both transcriptomes were assessed based on percent completeness reported by CEGMA, taking into consideration that most genes represented in the CEGMA gene set are highly expressed basic housekeeping genes (Akashi and Eyre-Walker 1998; Parra et al. 2007). However, even with this caveat, a more comprehensive transcriptome assembly is expected to include highly expressed conserved genes. In this context, the *C. quadricarinatus* transcriptome reported in this study is estimated to contain 94.76 % complete genes (235 genes) out of the CEGMA 248 gene set, considerably surpassing the transcriptome of Glazer et al. (2013) at only 24.19 % (60 complete genes).

The sequencing of the transcriptome from a range of tissues also allowed us to estimate the core crayfish proteome as well as make comparisons between the transcriptomic profiles of different tissues. From this analysis, we discovered that the highly expressed core crayfish transcripts were better annotated compared to tissue-specific transcripts (see bar charts in Online Resource 1). This observation is not surprising since transcripts shared among all tissues are expected to be genes that are more conserved and are thus usually better characterized in other organisms as well. In addition, unannotated transcripts from the core transcript set may also have been transcribed from crayfish- or decapod-specific genes, which have not been extensively studied and are not present in the protein databases. A comparison of expression profiles of multiple tissues also reveals a lack of annotation information for transcripts expressed in some tissues, specifically the kidney, nerve, and testis. The significance of this issue is highlighted from finding that only one out of the top ten highly expressed transcripts in the testis tissue was successfully annotated (Online Resource 1).

In this study, we identified an abundance of enzymes involved in various glycoside hydrolase functions (Fig. 3), some of which have been found in crayfishes or other crustaceans (Huang et al. 2010; Zhang et al. 2014; Linton et al. 2015; Ong et al. 2015). One of the significant findings includes the discovery of 11 putative crayfish chitinases belonging to GH18, which is consistent with other studies that have identified multiple chitinases in other crustaceans such as shrimps and crabs (Huang et al. 2010; Zhang et al. 2014), further extending the known diversity of the chitinase family in crustaceans.


a
Bacteria
Archaea
Plant
Fungi
Metazoa
C. quadricarinatus


b
H K L N T
log2(TPM+1)
AIN40245.1 [Cherax quadricarinatus]
T_2797 [Cherax quadricarinatus]
T_3881 [Cherax quadricarinatus]
T_3307 [Cherax quadricarinatus]
T_14216 [Cherax quadricarinatus]
ADE58567.1 [Limnoria quadripunctata]
ADE58568.1 [Limnoria quadripunctata]
ADE58569.1 [Limnoria quadripunctata]
ABV68808.1 [Cryptopygus antarcticus]
ADU33271.1 [Callosobruchus maculatus]
ADU33272.1 [Callosobruchus maculatus]
ADU33273.1 [Callosobruchus maculatus]
ADU33274.1 [Callosobruchus maculatus]
ADU33333.1 [Gastrophysa viridula]
BAI99559.1 [Haliotis discus]
BAE78456.1 [Haliotis discus]
ACJ12612.1 [Haliotis discus]
BAJ60954.1 [Aplysia kurodai]
AAV91524.1 [Biomphalaria glabrata]
AAV91523.1 [Biomphalaria glabrata]
CAC81056.1 [Mytilus edulis]
ACA05117.2 [Flammeovirga yaeyamensis]
ABD81545.1 [Saccharophagus degradans]
ACX74338.1 [Fibrobacter succinogenes]
AEV59732.1 [uncultured bacterium]
ADN01068.1 [Spirochaeta thermophila]
AEJ60366.1 [Spirochaeta thermophila]
AFN74927.1 [Melioribacter roseus]
0.2
Crustacea
Ellipura
Insecta
Gastropoda
Bivalvia
Arthropoda
Mollusca
Metazoa
Bacteria

Chitinase may facilitate the digestion of chitinous foods (Zhang et al. 2014), the degradation of endogenous chitin in the gut peritrophic membrane prior to molting (Tan et al. 2000; Zhang et al. 2014), or even other roles not related to molting, but their exact role will need to be validated since chitinases have a wide variety of functions. In addition, we also found a single GH16 glucanase sequence in *C. quadricarinatus* that contains sequence regions conserved in various glucanases isolated from decapod crustaceans (Linton et al. 2015), suggesting the ubiquity and an essential role of GH16 glucanase in decapod carbohydrate metabolism (T_28955). The high expression of CAZy-coding transcripts in the hepatopancreas tissue supports their function as digestive enzymes and more specifically a direct role in the digestion of plant-derived material (Byrne et al. 1999). The diverse CAZy profile in *C. quadricarinatus* (Fig. 3) is likely to be encountered in other *Cherax* spp. and freshwater crayfishes given their similar ecological niche (Crawford et al. 2004). However, some CAZy proteins may play a more general physiological role in *C. quadricarinatus* given their uniform transcript expression across all tissues e.g., T_40795 and T_22312 (Fig. 3 and Online Resource 4).

A major outcome of this study was finding an unexpected diversity of cellulase genes in GH5 and GH9 families. Byrne et al. (1999) first reported a cellulase from the GH9 family in *C. quadricarinatus*, and it was further characterized by Crawford et al. (2004). Notably, we found this family to be highly diverse in our crayfish sample, represented by 20 forms with varying degrees of divergence (Fig. 4c). Some of these forms are likely to be of significant evolutionary antiquity as they cluster with homologs from quite divergent taxa. This is supported by the protein sequences from the GH9 family which cluster according to their taxonomic domains while forming monophyletic clades for plants, fungi, and metazoans (Fig. 4a). Similarly, the freshwater crayfish sequences cluster with other crustaceans (Fig. 4b), supporting an endogenous origin for this gene family in the crayfishes and most likely in other metazoan groups as opposed to a recent or bacterial origin. Davison and Blaxter (2005) also suggested a single, ancient origin of the eukaryotic GH9 gene. In addition, most of our reported putative GH9 homologs identified from the transcriptome appear to be "true" cellulases as they contain the expected amino acid residues at the five active sites which are almost always invariant across the Eukarya as documented by Davison and Blaxter (2005).

In a recent study, Dammannagoda et al. (2015) characterized the GH5 mannanase in *C. quadricarinatus*, which has also been found in related species of crayfishes and other decapod crustaceans (Linton et al. 2006). We identified transcripts identical to their reported genes and three additional closely related forms, all predominantly expressed in the hepatopancreas, supporting an endogenous function in crayfishes. In the ML tree for subfamily GH5_10 (Fig. 5b), all four putative GH5 proteins form a common clade with the publicly available *C. quadricarinatus* partial mannanase protein (Dammannagoda et al. 2015), possibly representing multiple forms of the protein. The unrooted circular phylogram in Fig. 5a for GH5 clearly shows groups of cellulase sequences in their expected subfamilies as also demonstrated by Aspeborg et al. (2012), with our putative GH5 proteins placed in the GH5_10 subfamily. Figure 5a also shows the grouping of cellulases first into separate subfamilies and then within most subfamilies, where they form rather distinct monophyletic groups according to taxonomic domains. This observation points to the likelihood of a vertical descent of GH5 genes as opposed to a horizontal transfer of the gene as suggested by Aspeborg et al. (2012). In addition, our update of the GH5 phylogenetic tree also places unclassified GH5 proteins into possible subfamilies based on their phylogenetic positions as indicated on the second ring around the phylogram in Fig. 5a.

Our identification of putative endogenous cellulases of both GH families in *C. quadricarinatus* is consistent with findings in other studies (Byrne et al. 1999; Crawford et al. 2004; Dammannagoda et al. 2015). Our findings on cellulase diversity have important implications in understanding the nutritional biology and trophic roles of crayfishes, which are of considerable interest as a number of species have been shown to have major ecological impacts in freshwater systems (Nyström et al. 1996; Whitledge and Rabeni 1997), to be excellent invaders of aquatic environments worldwide (Ahyong and Yeo 2007; Saoud et al. 2013) or are subject to significant aquaculture industries in many parts of the world (Huner 1994).

Traditionally, crayfishes have been assumed to be detritivores or herbivores based on stomach content analysis and field observations (Momot 1995; Roth et al. 2006). Aquaculturists widely assume that detritus is a nutritional requirement for crayfishes for successful pond culture (Mitchell et al. 1995; Salame and Rouse 2000; Huner 1994). In a provocative paper, Momot (1995) opined that the view of crayfishes as primarily detritivores or herbivores is flawed as stomach content analysis seriously underestimates the significance of soft-bodied animal prey relative to the consumption of plant and detrital material, which are retained for longer in the stomach. He concluded that freshwater crayfishes are better considered as keystone predators and occasionally facultative herbivores that play a major role in the transfer of energy and materials in aquatic food webs. Several subsequent studies have essentially supported his position based on studies of both northern hemisphere (Weinländer and Füreder 2011; Nyström et al. 1996; Roth et al. 2006) and southern hemisphere (Parkyn et al. 2001; Johnston et al. 2011) crayfish confirming that they mostly operate as predators but can also consume plant material from time to time.

Our discovery that *C. quadricarinatus* has a rich diversity of cellulase genes of significant evolutionary antiquity

suggests that the now prevailing view of crayfishes as primarily carnivores requires reassessment, especially from an evolutionary viewpoint. This position is supported by the recent study of Johnston et al. (2011), who demonstrated considerable variation between species from the same crayfish community ranging from primarily carnivorous species (*Euastacus bispinosus*) to primarily herbivorous species (*Gramastacus insolitus*) and to species that had either mixed diets or switched between plants and animals at different sites (*Geocharax falcata* and *Cherax destructor*). Thus, part of the conflicting views in the literature may relate to studies of crayfishes from different families and inhabiting very different environments (i.e., slow flowing, organically rich tropical streams versus clear temperate lakes) and insufficient sampling in space and time and limited comparisons of co-occurring species (Johnston et al. 2011). It will be of great interest to extend the findings of this study to examine cellulase diversity in other species of crayfishes with contrasting trophic status (e.g., *Euastacus* and *Gramastacus*) and from the northern hemisphere families and to study cellulase gene expression in controlled feeding trials. A possibility, therefore, that remains to be further explored, is that the diversification of cellulase genes we have observed in this study may be part of a suit of adaptive responses that allowed ancestral crayfishes to exploit freshwater environments after they split from their marine ancestors over 250 million years ago (Martin et al. 2008; Porter et al. 2005). From a nutrition viewpoint, freshwater systems are markedly different from marine systems in that the predominant organic inputs are from terrestrial plant sources (Fisher and Likens 1973; Vannote et al. 1980; Tank et al. 2010).

Conclusion

Our study significantly contributes to the limited genomic resource available for crayfishes and more generally for crustacean species, thereby providing important baseline data for comparative transcriptomic studies of this important and threatened group of organisms (Richman et al. 2015). In addition, an investigation of carbohydrate-active enzymes revealed an unexpected diversity, including a proliferation of cellulose metabolizing proteins, with phylogenetic affinities across divergent taxa, suggesting an ancient evolutionary origin for some of these forms. These results have important implications for the current understanding of the nutrition biology of freshwater crayfishes, which may be less dependent on animal sources of protein than currently thought. The high diversity of carbohydrate-active enzymes may reflect the historical adaptation of crayfishes to the exploitation of plant-based nutrient sources in the course of their evolution into freshwater environments from their ancestral marine environments.

Acknowledgments Funding for this study was provided by the Monash University Malaysia (MUM) Tropical Medicine and Biology Multidisciplinary Platform.

Conflict of interest The authors declare that they have no conflict of interest.

Authors' contributions HMG and CMA conceived and designed the study. HMG, HYG, and YPL performed the sequencing of transcriptomes, and MHT analyzed the data. MHT, HMG, CMA, LJC, MBS, and ADM discussed the results and wrote the manuscript. All authors read and approved the final manuscript.

References

Abdu, U., Davis, C., Khalaila, I., & Sagi, A. (2002). The vitellogenin cDNA of *Cherax quadricarinatus* encodes a lipoprotein with calcium binding ability, and its expression is induced following the removal of the androgenic gland in a sexually plastic system. *General and Comparative Endocrinology, 127*(3), 263–272. doi:10.1016/S0016-6480(02)00053-9.

Ahyong, S., & Yeo, D. J. (2007). Feral populations of the Australian Red-Claw crayfish (*Cherax quadricarinatus* von Martens) in water supply catchments of Singapore. *Biological Invasions, 9*(8), 943–946. doi:10.1007/s10530-007-9094-0.

Akashi, H., & Eyre-Walker, A. (1998). Translational selection and molecular evolution. *Current Opinion in Genetics & Development, 8*(6), 688–693. doi:10.1016/S0959-437X(98)80038-5.

Alföldi, J., & Lindblad-Toh, K. (2013). Comparative genomics as a tool to understand evolution and disease. *Genome Research, 23*(7), 1063–1068.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403–410. doi:10.1016/S0022-2836(05)80360-2.

Amemiya, C. T., Alfoldi, J., Lee, A. P., Fan, S., Philippe, H., MacCallum, I., et al. (2013). The African coelacanth genome provides insights into tetrapod evolution. [Article]. *Nature, 496*(7445), 311–316. doi:10.1038/nature12027. **http://www.nature.com/nature/journal/v496/n7445/abs/nature12027.html#supplementary-information**.

Aspeborg, H., Coutinho, P., Wang, Y., Brumer, H., & Henrissat, B. (2012). Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evolutionary Biology, 12*(1), 186.

Austin, C. (1996). Systematics of the freshwater crayfish genus *Cherax* Erichson (Decapoda: Parastacidae) in Northern and Eastern Australia: electrophoretic and morphological variation. *Australian Journal of Zoology, 44*(3), 259–296. doi:10.1071/ZO9960259.

Austin, C., & Knott, B. (1996). Systematics of the freshwater crayfish genus *Cherax* Erichson (Decapoda: Parastacidae) in South-Western Australia: electrophoretic, morphological and habitat variation. *Australian Journal of Zoology, 44*(3), 223–258. doi:10.1071/ZO9960223.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. doi:10.1093/bioinformatics/btu170.

Bracken-Grissom, H. D., Ahyong, S. T., Wilkinson, R. D., Feldmann, R. M., Schweitzer, C. E., Breinholt, J. W., et al. (2014). The emergence of lobsters: phylogenetic relationships, morphological evolution and divergence time comparisons of an ancient group (Decapoda: Achelata, Astacidea, Glypheidea, Polychelida). *Systematic Biology*. doi:10.1093/sysbio/syu008.

Byrne, K. A., Lehnert, S. A., Johnson, S. E., & Moore, S. S. (1999). Isolation of a cDNA encoding a putative cellulase in the red claw

crayfish *Cherax quadricarinatus*. *Gene, 239*(2), 317–324. doi:10.1016/S0378-1119(99)00396-0.

Chen, R. (2015). A paradigm shift in biomass technology from complete to partial cellulose hydrolysis: lessons learned from nature. *Bioengineered*. doi:10.1080/21655979.2014.1004019.

Chen, H., & Boutros, P. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics, 12*(1), 35.

Clarke, T., Garb, J., Hayashi, C., Haney, R., Lancaster, A., Corbett, S., et al. (2014). Multi-tissue transcriptomics of the black widow spider reveals expansions, co-options, and functional processes of the silk gland gene toolkit. *BMC Genomics, 15*(1), 365.

Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., et al. (2011). The ecoresponsive genome of daphnia pulex. *Science, 331*(6017), 555–561. doi:10.1126/science.1197761.

Consortium, i. K. (2013). The i5K initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *Journal of Heredity, 104*(5), 595–600. doi:10.1093/jhered/est050.

Crawford, A. C., Kricker, J. A., Anderson, A. J., Richardson, N. R., & Mather, P. B. (2004). Structure and function of a cellulase gene in redclaw crayfish, *Cherax quadricarinatus*. *Gene, 340*(2), 267–274. doi:10.1016/j.gene.2004.06.060.

Criscuolo, A., & Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology, 10*(1), 210.

Dammannagoda, L. K., Pavasovic, A., Prentis, P. J., Hurwood, D. A., & Mather, P. B. (2015). Expression and characterization of digestive enzyme genes from hepatopancreatic transcripts from redclaw crayfish (*Cherax quadricarinatus*). *Aquaculture Nutrition*. doi:10.1111/anu.12211.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics, 27*(8), 1164–1165. doi:10.1093/bioinformatics/btr088.

Davidson, N., & Oshlack, A. (2014). Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biology, 15*(7), 410.

Davison, A., & Blaxter, M. (2005). Ancient origin of glycosyl hydrolase family 9 cellulase genes. *Molecular Biology and Evolution, 22*(5), 1273–1284. doi:10.1093/molbev/msi107.

Eaton, D. A. R., & Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology, 62*(5), 689–706. doi:10.1093/sysbio/syt032.

Fang, D., Wang, Q., Wang, J., He, L., Liu, L., & Wang, Y. (2011). A novel DDX5 gene in the freshwater crayfish *Cherax quadricarinatus* is highly expressed during ontogenesis and spermatogenesis. *Genetics and Molecular Research, 10*(4), 3963–3975.

Fernández, M. S., Bustos, C., Luquet, G., Saez, D., Neira-Carrillo, A., Corneillat, M., et al. (2012). Proteoglycan occurrence in Gastrolith of the crayfish Cherax quadricarinatus (Malacostraca: Decapoda). *Journal of Crustacean Biology, 32*(5), 802–815. doi:10.1163/193724012x649804.

Fisher, S. G., & Likens, G. E. (1973). Energy flow in bear brook, New Hampshire: an integrative approach to stream ecosystem metabolism. *Ecological Monographs, 43*(4), 421–439. doi:10.2307/1942301.

Gan, H. M., Tan, M. H., & Austin, C. M. (2014). The complete mitogenome of the red claw crayfish *Cherax quadricarinatus* (Von Martens, 1868) (Crustacea: Decapoda: Parastacidae). *Mitochondrial DNA, 0*(0), 1–2, doi:doi:10.3109/19401736.2014.895997.

Glazer, L., Tom, M., Weil, S., Roth, Z., Khalaila, I., Mittelman, B., et al. (2013). Hemocyanin with phenoloxidase activity in the chitin matrix of the crayfish gastrolith. *The Journal of Experimental Biology, 216*(10), 1898–1904. doi:10.1242/jeb.080945.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology, 29*(7), 644–652. doi:10.1038/nbt.1883. **http://www.nature.com/nbt/journal/v29/n7/abs/nbt.1883.html#supplementary-information**.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology, 52*(5), 696–704. doi:10.1080/10635150390235520.

Hayakijkosol, O., La Fauce, K., & Owens, L. (2011). Experimental infection of redclaw crayfish (Cherax quadricarinatus) with Macrobrachium rosenbergii nodavirus, the aetiological agent of white tail disease. *Aquaculture, 319*(1–2), 25–29. doi:10.1016/j.aquaculture.2011.06.023.

Hayward, A., Cornwallis, C. K., & Jern, P. (2015). Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. *Proceedings of the National Academy of Sciences, 112*(2), 464–469. doi:10.1073/pnas.1414980112.

Huang, Q.-S., Yan, J.-H., Tang, J.-Y., Tao, Y.-M., Xie, X.-L., Wang, Y., et al. (2010). Cloning and tissue expressions of seven chitinase family genes in Litopenaeus vannamei. *Fish & Shellfish Immunology, 29*(1), 75–81. doi:10.1016/j.fsi.2010.02.014.

Huner, J. V. (1994). *Freshwater crayfish aquaculture in North America, Europe, and Australia: Families Astacidae, Cambaridae, and Parastacidae*. New York: Food products press.

James, J., Slater, F. M., Vaughan, I. P., Young, K. A., & Cable, J. (2014). Comparing the ecological impacts of native and invasive crayfish: could native species' translocation do more harm than good? *Oecologia*, 1–8, doi:10.1007/s00442-014-3195-0.

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., et al. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science, 346*(6215), 1320–1331. doi:10.1126/science.1253451.

Johnston, K., Robson, B. J., & Fairweather, P. G. (2011). Trophic positions of omnivores are not always flexible: evidence from four species of freshwater crayfish. *Austral Ecology, 36*(3), 269–279. doi:10.1111/j.1442-9993.2010.02147.x.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics, 30*(9), 1236–1240. doi:10.1093/bioinformatics/btu031.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution, 30*(4), 772–780. doi:10.1093/molbev/mst010.

Kolde, R. (2012). pheatmap: Pretty heatmaps. *R package version 0.6, 1*.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. [Brief Communication]. *Nature Methods, 9*(4), 357–359. doi:10.1038/nmeth.1923. **http://www.nature.com/nmeth/journal/v9/n4/abs/nmeth.1923.html#supplementary-information**.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology, 10*(3), R25.

Larson, E. R., & Olden, J. D. (2012). Using avatar species to model the potential distribution of emerging invaders. *Global Ecology and Biogeography, 21*(11), 1114–1125. doi:10.1111/j.1466-8238.2012.00758.x.

Li, B., & Dewey, C. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics, 12*(1), 323.

Li, W., Jaroszewski, L., & Godzik, A. (2001). Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics, 17*(3), 282–283. doi:10.1093/bioinformatics/17.3.282.

Linton, S. M., Greenaway, P., & Towle, D. (2006). Endogenous production of endo-β-1,4-glucanase by decapod crustaceans. *Journal of*

Comparative Physiology B, 176(4), 339–348. doi:10.1007/s00360-005-0056-5.

Linton, S. M., Cameron, M. S., Gray, M. C., Donald, J. A., Saborowski, R., von Bergen, M., et al. (2015). A glycosyl hydrolase family 16 gene is responsible for the endogenous production of β-1,3-glucanases within decapod crustaceans. *Gene, 569*(2), 203–217. doi:10.1016/j.gene.2015.05.056.

Liu, H. P., Chen, R. Y., Zhang, Q. X., Peng, H., & Wang, K.-J. (2011). Differential gene expression profile from haematopoietic tissue stem cells of red claw crayfish, *Cherax quadricarinatus*, in response to WSSV infection. *Developmental & Comparative Immunology, 35*(7), 716–724. doi:10.1016/j.dci.2011.02.015.

Lodge, D. M., Deines, A., Gherardi, F., Yeo, D. C. J., Arcella, T., Baldridge, A. K., et al. (2012). Global introductions of crayfishes: evaluating the impact of species invasions on ecosystem services. *Annual Review of Ecology, Evolution, and Systematics, 43*(1), 449–472. doi:10.1146/annurev-ecolsys-111511-103919.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research, 42*(D1), D490–D495. doi:10.1093/nar/gkt1178.

Lv, J., Liu, P., Wang, Y., Gao, B., Chen, P., & Li, J. (2013). Transcriptome analysis of *Portunus trituberculatus* in response to salinity stress provides insights into the molecular basis of osmoregulation. *PLoS ONE, 8*(12), e82155. doi:10.1371/journal.pone.0082155.

Ma, H., Ma, C., Li, S., Jiang, W., Li, X., Liu, Y., et al. (2014). Transcriptome analysis of the mud crab (*Scylla paramamosain*) by 454 deep sequencing: assembly, annotation, and marker discovery. *PLoS ONE, 9*(7), e102668. doi:10.1371/journal.pone.0102668.

Martin, A. J., Rich, T. H., Poore, G. C. B., Schultz, M. B., Austin, C. M., Kool, L., et al. (2008). Fossil evidence in Australia for oldest known freshwater crayfish of Gondwana. *Gondwana Research, 14*(3), 287–296. doi:10.1016/j.gr.2008.01.002.

McCormack, J. E., Harvey, M. G., Faircloth, B. C., Crawford, N. G., Glenn, T. C., & Brumfield, R. T. (2013). A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS ONE, 8*(1), e54848. doi:10.1371/journal.pone.0054848.

Mitchell, B. D., Anderson, T., De Silva, S. S., Collins, R. O., Chavez, J. R., Jones, P. L., et al. (1995). A conceptual production model for freshwater crayfish pond culture incorporating detrital forage. *Aquaculture Research, 26*(2), 117–127. doi:10.1111/j.1365-2109.1995.tb00891.x.

Momot, W. T. (1995). Redefining the role of crayfish in aquatic ecosystems. *Reviews in Fisheries Science, 3*(1), 33–63. doi:10.1080/10641269509388566.

Muhire, B. M., Varsani, A., & Martin, D. P. (2014). SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS ONE, 9*(9), e108277. doi:10.1371/journal.pone.0108277.

Nguyen, K. Y., Sakuna, K., Kinobe, R., & Owens, L. (2014). Ivermectin blocks the nuclear location signal of parvoviruses in crayfish, Cherax quadricarinatus. *Aquaculture, 420–421*, 288–294. doi:10.1016/j.aquaculture.2013.11.022.

Nyström, P. E. R., Brönmark, C., & Granéli, W. (1996). Patterns in benthic food webs: a role for omnivorous crayfish? *Freshwater Biology, 36*(3), 631–646. doi:10.1046/j.1365-2427.1996.d01-528.x.

Oakley, T. H., Wolfe, J. M., Lindgren, A. R., & Zaharoff, A. K. (2013). Phylotranscriptomics to bring the understudied into the fold: monophyletic ostracoda, fossil placement, and pancrustacean phylogeny. *Molecular Biology and Evolution, 30*(1), 215–233. doi:10.1093/molbev/mss216.

Ong, S.-S., Bhassu, S., Kwong, Q. B., Mather, P., Simarani, K., & Othman, R. Y. (2015). Identification of a putative cellulase gene in the giant freshwater prawn, Macrobrachium rosenbergii (De Man, 1879). *Aquaculture Research*. doi:10.1111/are.12818.

Pallavicini, A., Canapa, A., Barucca, M., Alfoldi, J., Biscotti, M., Buonocore, F., et al. (2013). Analysis of the transcriptome of the Indonesian coelacanth *Latimeria menadoensis*. *BMC Genomics, 14*(1), 538.

Pamuru, R. R., Rosen, O., Manor, R., Chung, J. S., Zmora, N., Glazer, L., et al. (2012). Stimulation of molt by RNA interference of the molt-inhibiting hormone in the crayfish Cherax quadricarinatus. *General and Comparative Endocrinology, 178*(2), 227–236. doi:10.1016/j.ygcen.2012.05.007.

Parkyn, S. M., Collier, K. J., & Hicks, B. J. (2001). New Zealand stream crayfish: functional omnivores but trophic predators? *Freshwater Biology, 46*(5), 641–652. doi:10.1046/j.1365-2427.2001.00702.x.

Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics, 23*(9), 1061–1067. doi:10.1093/bioinformatics/btm071.

Peng, Y., Leung, H. C. M., Yiu, S.-M., Lv, M.-J., Zhu, X.-G., & Chin, F. Y. L. (2013). IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. *Bioinformatics, 29*(13), i326–i334. doi:10.1093/bioinformatics/btt219.

Porter, M. L., Pérez-Losada, M., & Crandall, K. A. (2005). Model-based multi-locus estimation of decapod phylogeny and divergence times. *Molecular Phylogenetics and Evolution, 37*(2), 355–369. doi:10.1016/j.ympev.2005.06.021.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE, 5*(3), e9490. doi:10.1371/journal.pone.0009490.

Richman, N. I., Böhm, M., Adams, S. B., Alvarez, F., Bergey, E. A., Bunn, J. J. S., et al. (2015). Multiple drivers of decline in the global status of freshwater crayfish (Decapoda: Astacidea). *Philosophical Transactions of the Royal Society, B: Biological Sciences, 370*(1662), 20140060. doi:10.1098/rstb.2014.0060.

Roth, B. M., Hein, C. L., & Vander Zanden, M. J. (2006). Using bioenergetics and stable isotopes to assess the trophic role of rusty crayfish (Orconectes rusticus) in lake littoral zones. *Canadian Journal of Fisheries and Aquatic Sciences, 63*(2), 335–344. doi:10.1139/f05-217.

Rubin, C.-J., Megens, H.-J., Barrio, A. M., Maqbool, K., Sayyab, S., Schwochow, D., et al. (2012). Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences, 109*(48), 19529–19536. doi:10.1073/pnas.1217149109.

Sagi, A., & Khalaila, I. (2001). The Crustacean androgen: a hormone in an isopod and androgenic activity in decapods. *American Zoologist, 41*(3), 477–484. doi:10.1093/icb/41.3.477.

Sahoo, P. K., Kar, B., Mohapatra, A., & Mohanty, J. (2013). De novo whole transcriptome analysis of the fish louse, *Argulus siamensis*: first molecular insights into characterization of Toll downstream signalling molecules of crustaceans. *Experimental Parasitology, 135*(3), 629–641. doi:10.1016/j.exppara.2013.09.018.

Salame, M. J., & Rouse, D. B. (2000). Forage-based feeding in commercial red claw ponds in Ecuador. *Journal of Applied Aquaculture, 10*(3), 83–90. doi:10.1300/J028v10n03_07.

Saoud, I. P., Ghanawi, J., Thompson, K. R., & Webster, C. D. (2013). A review of the culture and diseases of Redclaw crayfish *Cherax quadricarinatus* (Von Martens 1868). *Journal of the World Aquaculture Society, 44*(1), 1–29. doi:10.1111/jwas.12011.

Scientists G. C.o. (2014). The Global Invertebrate Genomics Alliance (GIGA): developing community resources to study diverse invertebrate genomes. *Journal of Heredity, 105*(1), 1–18. doi:10.1093/jhered/est084.

Shechter, A., Aflalo, E. D., Davis, C., & Sagi, A. (2005). Expression of the reproductive female-specific vitellogenin gene in endocrinologically induced male and intersex Cherax quadricarinatus crayfish. *Biology of Reproduction, 73*(1), 72–79. doi:10.1095/biolreprod.104.038554.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics, 22*(21), 2688–2690. doi:10.1093/bioinformatics/btl446.

Tan, S. H., Degnan, B. M., & Lehnert, S. A. (2000). The Penaeus monodon Chitinase 1 gene is differentially expressed in the Hepatopancreas during the molt cycle. *Marine Biotechnology, 2*(2), 126–135. doi:10.1007/s101269900016.

Tank, J. L., Rosi-Marshall, E. J., Griffiths, N. A., Entrekin, S. A., & Stephen, M. L. (2010). A review of allochthonous organic matter dynamics and metabolism in streams. *Journal of the North American Benthological Society, 29*(1), 118–146. doi:10.1899/08-170.1.

Toon, A., Pérez-Losada, M., Schweitzer, C. E., Feldmann, R. M., Carlson, M., & Crandall, K. A. (2010). Gondwanan radiation of the Southern Hemisphere crayfishes (Decapoda: Parastacidae): evidence from fossils and molecules. *Journal of Biogeography, 37*(12), 2275–2290. doi:10.1111/j.1365-2699.2010.02374.x.

Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., & Cushing, C. E. (1980). The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences, 37*(1), 130–137. doi:10.1139/f80-017.

Ward, L. D., & Kellis, M. (2012). Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Science, 337*(6102), 1675–1678. doi:10.1126/science.1225057.

Weinländer, M., & Füreder, L. (2011). Crayfish as trophic agents: effect of Austropotamobius torrentium on zoobenthos structure and function in small forest streams. *Knowledge and Management of Aquatic Ecosystems, 401*.

Whitledge, G. W., & Rabeni, C. F. (1997). Energy sources and ecological role of crayfishes in an Ozark stream: insights from stable isotopes and gut analysis. *Canadian Journal of Fisheries and Aquatic Sciences, 54*(11), 2555–2563. doi:10.1139/f97-173.

Yang, Y., & Smith, S. A. (2014). Orthology inference in non-model organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution*. doi:10.1093/molbev/msu245.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., & Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research, 40*(Web Server issue), W445–W451. doi:10.1093/nar/gks479.

You, X., Bian, C., Zan, Q., Xu, X., Liu, X., Chen, J., et al. (2014). Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes. [Article]. *Nat Commun, 5*, doi:10.1038/ncomms6594.

Yudkovski, Y., Glazer, L., Shechter, A., Reinhardt, R., Chalifa-Caspi, V., Sagi, A., et al. (2010). Multi-transcript expression patterns in the gastrolith disk and the hypodermis of the crayfish *Cherax quadricarinatus* at premolt. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 5*(2), 171–177. doi:10.1016/j.cbd.2010.03.010.

Zhang, S., Jiang, S., Xiong, Y., Fu, H., Sun, S., Qiao, H., et al. (2014). Six chitinases from oriental river prawn Macrobrachium nipponense: cDNA characterization, classification and mRNA expression during post-embryonic development and moulting cycle. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 167*, 30–40. doi:10.1016/j.cbpb.2013.09.009.

Zhernakova, A., Elbers, C. C., Ferwerda, B., Romanos, J., Trynka, G., Dubois, P. C., et al. (2010). Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. *The American Journal of Human Genetics, 86*(6), 970–977. doi:10.1016/j.ajhg.2010.05.004.