A crustacean annotated transcriptome (CAT) database

Wenyan Nong¹, Zacary Y. H. Chai², Xiaosen Jiang³, Jing Qin⁴, Ka Yan Ma², King Ming Chan⁵, Ting Fung Chan⁶, Billy K. C. Chow⁷, Hoi Shan Kwan⁵, Chris K. C. Wong⁸, Jian-Wen Qiu⁸, Jerome H. L. Hui¹* and Ka Hou Chu²

Abstract

Background: Decapods are an order of crustaceans which includes shrimps, crabs, lobsters and crayfish. They occur worldwide and are of great scientific interest as well as being of ecological and economic importance in fisheries and aquaculture. However, our knowledge of their biology mainly comes from the group which is most closely related to crustaceans – insects. Here we produce a de novo transcriptome database, crustacean annotated transcriptome (CAT) database, spanning multiple tissues and the life stages of seven crustaceans.

Description: A total of 71 transcriptome assemblies from six decapod species and a stomatopod species, including the coral shrimp Stenopus hispidus, the cherry shrimp Neocaridina davidi, the redclaw crayfish Cherax quadricarinatus, the spiny lobster Paralithodes ornatus, the coconut crab Birgus latro, and the zebra mantis shrimp Lysiosquillina maculata, were generated. Differential gene expression analyses within species were generated as a reference and included in a graphical user interface database at http://cat.sls.cuhk.edu.hk/. Users can carry out gene name searches and also access gene sequences based on a sequence query using the BLAST search function.

Conclusions: The data generated and deposited in this database offers a valuable resource for the further study of these crustaceans, as well as being of use in aquaculture development.

Background

The Arthropoda is a phylum containing the largest number (nearly 85%) of described living species in the world. For various historical reasons, most of our knowledge of their biology comes from insects, particularly fruit flies Drosophila. Crustacea (including shrimps, lobsters, crayfish, crabs) forms a large subphylum of arthropods now proven to be the closest relatives of Insecta. In the past decade, a substantial number of insect genomes have been sequenced across the different groups (e.g. beetle, wasp, bee, aphid, butterfly, and moth), especially in the course of the on-going 5000 insect genome project (i5k Consortium). By contrast, the genomic resources of crustaceans are relatively scarce, and are limited to a few species (e.g. [1–6]). Carcinology, or the study of crustaceans, benefits both basic science and the aquaculture industry, presently the fastest growing animal food-producing sector worldwide. Here, we generated a user-friendly database, the crustacean annotated transcriptome (CAT) database, which enables users to search for the annotated gene name as well as gene sequences based on sequence query. This database contains newly generated crustacean transcriptomic data from different developmental stages and the tissues of seven crustacean species, including a stomatopod mantis shrimp, two decapod shrimps, a crayfish, a lobster, and two anomuran crabs (Fig. 1).
shop and maintained for over 2 weeks as mating pairs in separate 10-L seawater tanks at an ambient indoor temperature (20–26 °C) with diurnal lighting and environmental enrichments of moss and wood, and were fed with aquarist shrimp feed. Tissue samples were collected from a single adult female at the intermolt stage, while “whole body” samples were obtained from 50 to 100 early (no eye spot) and late (with eye spot) stage eggs obtained from two females separately.

Cherry shrimp (Decapoda: Caridea: Neocaridina davidi) were purchased from an aquarium shop in Hong Kong. Again, they were kept in 10-L freshwater tanks at an ambient indoor temperature with diurnal lighting, and fed with aquarist shrimp feed. Tissue samples were collected from a single female adult at the intermolt stage, while “whole-body” samples were obtained from 50 to 100 early (no eye spot) and late (with eye spot) stage eggs obtained from two females separately.

Red claw crayfish (Decapoda: Astacidea: Parastacidae: Cherax quadricarinatus) at different life history stages were sourced from a breeder in Queensland, Australia. The juvenile (~7–10 cm in length) and adult (15–18 cm in length) crayfish were acclimated for over 2 weeks in 100-L freshwater tanks at an ambient indoor temperature with diurnal lighting and enriched with hiding nets, and were fed aquarist shrimp feed. Tissue samples were collected from a single adult female at the intermolt stage, while “whole-body” samples were obtained from a 15-day-old juvenile, as well as from ~20 early (no eye spot) and late (with eye spot) stage eggs (~6 eggs per replicate) from two females separately.

Spiny lobsters (Decapoda: Achelata: Palinuridae: Panulirus ornatus) were purchased from a fish market in Hong Kong, and acclimated for 2 weeks in 500-L tanks in an outdoor enclosure at 25–30 °C and fed with live clams. Tissue samples were collected from a single adult female at the intermolt stage.

Adult male coconut crabs (Decapoda: Anomura: Coenobitiidae: Birgus latro) were purchased and imported from a fish market in Okinawa, Japan. The crabs were fed a diet of coconut meat and boiled root vegetables while acclimating for 2 weeks in a controlled environment in a large outdoor enclosure at 25–30 °C. The enclosure was enriched with damp straw, reptile lights on diurnal control and a pool of running fresh water, and a humidifier maintained a relative humidity of 70–80%. Tissue samples were collected from a single individual.

Adult male king crabs (Decapoda: Anomura: Lithodidae: Paralithodes camtschaticus) were imported from Alaska and fed with live clams while acclimating for 2 weeks in 100-L seawater tanks kept at 4 °C in a dark room. Tissue samples were collected from a single individual.

Zebra mantis shrimp (Stomatopoda: Lysiosquillidae: Lysiosquillina maculata) were purchased from a fish market in Hong Kong and acclimated for 2 weeks in 100-L seawater tanks at ambient indoor temperature with diurnal lighting and 20 cm of sand, and were fed with live fish. Tissue samples were collected from a single adult female at the intermolt stage.

Tissue samples of gill, eye stalk, ovary (female only), hepatopancreas, and muscle were obtained from adults of all target species and juveniles of crayfish. Gill tissues were dissected, pooled and homogenised. Tissue from eyestalks were dissected, avoiding the pigmented retina and discarding the exoskeleton. Ovary tissues were collected from mature females. Hepatopancreas tissues
### Table 1 Transcriptomes generated in this study

| Species                        | Samples ID | No. of Raw Reads | No. of Clean Reads | No. of Clean Bases | Accession No. |
|-------------------------------|------------|------------------|--------------------|--------------------|---------------|
| **Coral shrimp** S. hirudina |            |                  |                    |                    |               |
| Early Eggs                    | Early Eggs | 113,624,220      | 111,779,862        | 110,951,378        | SAMN14600695  |
| Eye Stalk                     | Eye Stalk  | 170,784,308      | 169,325,278        | 168,006,598        | SAMN14600695  |
| Gill                          | Gill        | 126,534,189      | 124,468,190        | 123,291,500        | SAMN14600695  |
| Hepatopancreas                | Hepatopancreas | 119,247,605       | 117,325,278        | 116,130,742        | SAMN14600695  |
| Larvae                        | Larvae     | 123,880,566      | 121,631,326        | 120,426,538        | SAMN14600695  |
| Late Eggs                     | Late Eggs  | 140,726,264      | 138,690,312        | 137,474,522        | SAMN14600695  |
| Muscle                        | Muscle     | 119,851,122      | 118,105,593        | 116,951,378        | SAMN14600695  |
| Ovaly                        | Ovaly      | 95,053,662       | 93,854,916         | 92,749,688         | SAMN14600695  |
| **Cherry crab** N. davidi   |            |                  |                    |                    |               |
| Early Eggs                    | Early Eggs | 117,111,332      | 115,820,738        | 114,581,985        | SAMN14600695  |
| Eye Stalk                     | Eye Stalk  | 176,951,368      | 175,618,494        | 174,354,700        | SAMN14600695  |
| Gill                          | Gill        | 126,534,189      | 124,468,190        | 123,291,500        | SAMN14600695  |
| Hepatopancreas                | Hepatopancreas | 119,247,605       | 117,325,278        | 116,130,742        | SAMN14600695  |
| Larvae                        | Larvae     | 123,880,566      | 121,631,326        | 120,426,538        | SAMN14600695  |
| Late Eggs                     | Late Eggs  | 120,688,840      | 119,204,154        | 117,951,304        | SAMN14600695  |
| Muscle                        | Muscle     | 118,371,584      | 116,874,096        | 115,727,868        | SAMN14600695  |
| **Redclaw crayfish** C. quadricarinatus |            |                  |                    |                    |               |
| Brown Eggs                    | Brown Eggs | 117,111,332      | 115,820,738        | 114,581,985        | SAMN14600695  |
| Eye Stalk                     | Eye Stalk  | 176,951,368      | 175,618,494        | 174,354,700        | SAMN14600695  |
| Gill                          | Gill        | 126,534,189      | 124,468,190        | 123,291,500        | SAMN14600695  |
| Hepatopancreas                | Hepatopancreas | 119,247,605       | 117,325,278        | 116,130,742        | SAMN14600695  |
| Larvae                        | Larvae     | 123,880,566      | 121,631,326        | 120,426,538        | SAMN14600695  |
| Muscle                        | Muscle     | 118,371,584      | 116,874,096        | 115,727,868        | SAMN14600695  |
| Ovaly                        | Ovaly      | 95,053,662       | 93,854,916         | 92,749,688         | SAMN14600695  |
| **Spiny lobster** P. ornatus  |            |                  |                    |                    |               |
| Early Eggs                    | Early Eggs | 108,244,990      | 106,345,078        | 105,229,494        | SAMN14600695  |
| Eye Stalk                     | Eye Stalk  | 134,407,510      | 132,431,672        | 131,305,888        | SAMN14600695  |
| Gill                          | Gill        | 123,616,922      | 121,555,503        | 120,440,924        | SAMN14600695  |
| Muscle                        | Muscle     | 116,359,450      | 114,212,768        | 113,087,432        | SAMN14600695  |
| **Coconut crab** B. latro   |            |                  |                    |                    |               |
| Eye Stalk                     | Eye Stalk  | 127,480,250      | 125,491,759        | 124,348,180        | SAMN14600695  |
| Gill                          | Gill        | 131,660,764      | 129,600,212        | 128,454,638        | SAMN14600695  |
| Muscle                        | Muscle     | 115,823,126      | 113,849,262        | 112,724,478        | SAMN14600695  |
| **Red king crab** P. paradoxus |            |                  |                    |                    |               |
| Eye Stalk                     | Eye Stalk  | 121,406,956      | 119,332,566        | 118,207,968        | SAMN14600695  |
| Gill                          | Gill        | 131,660,764      | 129,600,212        | 128,454,638        | SAMN14600695  |
| Muscle                        | Muscle     | 134,693,083      | 132,585,835        | 131,470,220        | SAMN14600695  |
| **Zebra mussel shrimp** M. trossulus |            |                  |                    |                    |               |
| Eye Stalk                     | Eye Stalk  | 123,669,922      | 121,675,158        | 120,540,938        | SAMN14600695  |
| Gill                          | Gill        | 131,660,764      | 129,600,212        | 128,454,638        | SAMN14600695  |
| Muscle                        | Muscle     | 134,693,083      | 132,585,835        | 131,470,220        | SAMN14600695  |
were taken at distant tubules from the midgut caeca to avoid heavy bacterial contamination. Muscle was isolated from the abdomen from all the shrimp and crayfish species (including stomatopod) and from the large chela of crabs. Duplicate biological samples were collected. Tissue samples from adult and the "whole body" of juvenile animals, larvae, and eggs were frozen in liquid nitrogen and then stored at −80 °C before total RNA extraction.

**RNA extraction and sequencing**

Total RNAs were isolated using the miRVana microRNA Isolation Kit (Thermo Fisher Scientific). RNA concentration and quality were assessed by a NanoDrop Fluorospectrometer (Thermo Scientific). At least 5 μg of total RNA for each sample were enriched by ribo-reduction using Ribo-Zero rRNA removal kits (Epicentre). Trans criptome libraries were created using TrueSeq Stranded RNA Library Prep Kit v2 (Illumina) by Theragen Bio Institute in Korea, followed by 150 bp paired-end sequencing on an Illumina HiSeq 4000 platform to obtain at least 51 million clean reads (after filtering and trimming).

**Transcriptome assembly and annotation**

Raw sequencing reads from 71 transcriptomes were pre-processed with quality trimmed by trimmomatic (v0.33 with parameters “ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW:4:5 LEADING:5 TRAILING:5 MINLEN:25”, [7]), followed by de novo transcriptome assembly using Trinity (v2.4.0, [8, 9]) with the options “--SS_lib_type RF --normalize_reads” and other default parameters. All biological duplicates were combined to carry out the de novo assembly and estimation of transcript abundance using the script “align_and_estimate_abundance.pl” of the Trinity software with “--est_method RSEM --aln_method bowtie” (v1.1.2, [10]). Coding regions within transcripts were annotated using TransDecoder (v5.0.2 [11]), and functional annotation and analyses were carried out using Trinotate (v3.1.1, [12]). A summary of the assembled transcriptomes is shown in Table 1.

**Utility and discussion**

**Website construction**

The Crustacean Annotated Transcriptome (CAT) database is available at http://cat.sls.cuhk.edu.hk/. It was built using Codelgniter Web Framework. Codelgniter (https://www.codelgniter.com/) is a powerful PHP framework with a tiny footprint. The website provides researchers with several tools for transcriptome visualization, gene search, and gene blast.

**Transcriptome visualisation**

Gene expression data of various samples in each species can be visualised through the Degust (https://github.com/Victorian-Bioinformatics-Consortium/degust) toolset [13]. It allows the comparison of gene expression between different tissues of the same species. The users can browse differentially expressed genes (DEGs) between samples within the same species, perform their own DEG analysis, or analyse expression profiles using the inbuilt server.

**Gene sequence search**

The database contains 462,877 pieces of gene annotation information (coral shrimp: 57240, cherry shrimp: 92956, red claw crayfish: 99100, spiny lobster: 28805, coconut crab: 72729, red king crab: 73144, zebra mantis shrimp: 38903). The users can search a gene of a certain species by querying the “gene id” or “gene name” and selecting the species in the gene search section. After users submit their request, the results will be displayed in a table. The number of results will be shown at the top of the table. The table will list the general information of all matched genes, including the gene id, gene name and species information. Clicking on the “gene id” or “gene name” will bring users to a detailed information page of the gene. The nucleic acid sequence derived from de novo assembly, protein sequence deduced from assembled transcripts, and the expression of the gene in each sample can be viewed on the page.

**Gene blast**

The user can input or upload query sequence(s) in fasta format, select the corresponding species database and the blast type to perform the gene blast. Hits will be listed in the result table. The users can browse the detailed information of the hit genes by clicking on the hit IDs.

**Conclusions**

Carcinology benefits both the basic science and the aquaculture industry. We have here generated a platform (CAT) in hosting 71 new transcriptomes generated for seven species of decapod crustaceans and a stomatopod. CAT is constructed in a way aiming to facilitate research on this important branch of life, and will continue to be updated, to host more crustacean genomic resources in the future.

**Abbreviations**

BLAST: Basic Local Alignment Search Tool; CAT: Crustacean annotated transcriptome database; DEGs: Differentially expressed genes; IDs: Identifications; RNA: Ribonucleic acid; rRNA: Ribosomal ribonucleic acid

**Acknowledgements**

The authors would like to thank David Wilmshurst, former academic editor at CUHK, for editing the English of the manuscript. The authors would also thank Clive Jones of James Cook University, Australia for his kind assistance in procuring the crayfish samples; Henry So of The Chinese University of Hong Kong for providing the cherry shrimp picture; and Tin-Yam Chan of National Taiwan Ocean University and Peter Ng of National University of Singapore for providing other pictures in Fig. 1.
Authors’ contributions
WN performed transcriptome assemblies and functional annotation. ZYHC and KYM extracted RNA. WN, ZYHC, KYM, XJ, JQ designed and constructed the website. KMC, TFC, BKC, HSK, CKCW, JWQ, JHLH, KHC conceived the study design. WN, JHLH, KHC wrote the manuscript. All authors critically reviewed and approved the manuscript.

Funding
The work presented in this paper was supported by a grant from the Collaborative Research Fund (project no. C4042-14G), Research Grants Council, Hong Kong SAR, China. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials
The transcriptome data were deposited in NCBI under BioProjects PRJNA562428.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1School of Life Sciences, Simon F.S. Li Marine Science Laboratory, State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Hong Kong, China. 2School of Life Sciences, Simon F.S. Li Marine Science Laboratory, The Chinese University of Hong Kong, Hong Kong, China. 3School of Future Technology, University of Chinese Academy of Sciences, Hong Kong, China. 4School of Pharmaceutical Sciences (Shenzhen), Sun Yat-sen University, Shenzhen, China. 5School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China. 6School of Life Sciences, State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Hong Kong, China. 7School of Biological Sciences, The University of Hong Kong, Hong Kong, China. 8Department of Biology, Hong Kong Baptist University, Hong Kong, China.

Received: 3 October 2019 Accepted: 26 December 2019
Published online: 09 January 2020

References
1. Colbourne JK, et al. The ecoresponsive genome of Daphnia pulex. Science. 2011;331(6017):555–61.
2. Kenny NJ, et al. Genome sequence and experimental tractability of a new decapod shrimp model, Neocaridina denticulata. Marine Drugs. 2014;12(3):1419–37.
3. Song L, et al. Draft genome of the Chinese mitten crab, Eriocheir sinensis. Gigascience. 2016;5:5.
4. Kao D, et al. The genome of the crustacean Parhyale hawaiensis, a model for animal development, regeneration, immunity and lignocellulose digestion. Elife. 2016;16.5.
5. Gutknust J, et al. Clonal genome evolution and rapid invasive spread of the marbled crayfish. Nat Ecol Evol. 2018;2(3):567–73.
6. Zhang X, et al. Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. Nat Commun. 2019;10(1):356.
7. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
8. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Maucella E, Hacothen N, Girke T, Rihn N, Di Palma F, Birnbaum MC, Linblad-Toh K, Friedman N, Regev A. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 2011;29(4):145–50.
9. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, I B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494.
10. Langmead B. Aligning short sequencing reads with bowtie. Curr Protoc Bioinformatics. 2010;32(1):11–7.
11. Haas, B., Papanicolaou, A. (2012). Transdecoder. http://transdecoder.github.io/.
12. Haas, B. J. (2015). Trinotate: transcriptome functional annotation and analysis. https://github.com/Trinotate/Trinotate.github.io/wiki
13. Powell D. Degust: visualize, explore and appreciate RNA-seq differential gene-expression data. In: COMBINE RNA-seq workshop; 2015.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.