Comparative transcriptome analysis reveals osmotic-regulated genes in the gill of Chinese mitten crab (Eriocheir sinensis)

PLOS One

Junyu Zhou[1]

[1]Key Laboratory of Freshwater Aquatic Genetic Resources, Ministry of Agriculture, Shanghai Ocean University, Shanghai, China

dx.doi.org/10.17504/protocols.io.v4ve8w6

Junyu Zhou

ABSTRACT

To reveal the genes and pathways involved in osmoregulation of Chinese mitten crab, adult male crabs were acclimated for 144 h in freshwater (FW, 0 ppt) or seawater (SW, 25 ppt). Changes in the transcriptome of crab gills were then analysed by RNA-Seq.

EXTERNAL LINK

https://doi.org/10.1371/journal.pone.0210469

PROTOCOL STATUS

Working

MATERIALS TEXT

1. healthy male Chinese mitten crabs (body weight = 110 ± 5 g)
2. six tanks (80 cm ´ 40 cm ´ 40 cm)
3. Brine with salinity of 100ppt
4. Handheld refractive salinometer MASTER-S28M（ATAGO , Japan）
5. Freezing point osmotic apparatus Gonotec OSMOMAT 3000（Gonotec, Germany）
6. Water quality analyzer Haiheng XZ-0178 78(Haiheng, China)
7. ABI 7500 Real-Time PCR System (Applied Biosystems, America)
8. eppendorf centrifuge 5417r (eppendorf, Germany)
9. Agilent 2100（Agilent, America）
10. Thermo NanoDrop 2000/2000c(Thermo, America)
11. TBS-380 Microfluorometer(turner biosystems, America)
12. Cluster Generation (cBot) (Illumina, America)
13. HiSeq4000 platform (Illumina, America)
14. transfer liquid gun (eppendorf, Germany)
15. disposable tip 0.5-10μL, 2-200μL, 100-1000μL (Axygen)
16. centrifuge tube 0.2mL, 1.5mL (Axygen)
17. NEBNext（Ultra（RNA（Library（Prep（Kit（for（Illumina (NEB)
18. NEBNext PolyA mRNA Magnetic Isolation Module (NEB)
19. NEBNext Multiplex Oligos for Illumina (Index Primers Set1) (NEB)
20. NEBNext Multiplex Oligos for Illumina (Index Primers Set2) (NEB)
21. RNA 6000 Pico chip (Agilent)
22. High sensitivity DNA Assay Kit (Agilent)
23. RNA 6000 Nano chip (Agilent)
24. QUBIT RNA BR ASSAY KIT (Invitrogen)
25. QUBIT DNA BR ASSAY KIT (Invitrogen)
26. QUBIT DNA HS ASSAY KIT (Invitrogen)
27. KAPA SYBR FAST Master Mix Universal 2X qPCR Master Mix (KAPA Biosystem)
28. DNA Quantification Standards and Primer Premix Kit (KAPA Biosystem)
29. TruSeq Rapid Duo cBot Sample Loading Kit (Illumina)
30. TruSeq Rapid PE Cluster Kit (Illumina)
31. TruSeq Rapid SBS Kit (200 cycle)(Illumina)
32. 1 PhiX Control Kit v3 (10 lanes) (Illumina)

Crabs acclimation

1 Chinese mitten crabs (body weight 110±5g) were obtained from the Chongming Research Base of Shanghai Ocean University and kept in a freshwater tank for 1 week for acclimation.

NOTE
Cleaning the tank, oxygenating, changing water everyday

Osmotic pressure of crabs under salinity stress

2 After a week of temporary cultivation, crabs with the same specifications were selected and transferred to a rectangular glass tank (length, width, and height: 80cm×40cm×40cm) for salinity experiment. A salinity group (ppt 25) and a control group (ppt 0) were set up, three parallel crabs were placed in each group, and 10 crabs were placed in each parallel group. The different salinity aquaculture water was prepared by brine and filtered tap water, and salinity was tested by hand-held refractive salinity meter. Samples were taken at 0, 3, 6, 12, 24, 48, 72, 96 and 144 hours of the experiment. At each sampling, one crab was taken from each parallel group, and three crabs were taken from each experimental group. Crabs were placed on ice dishes for15 to 30 minutes under freeze anesthesia. Hemolymph samples were extracted from the base joints of the third or fourth foot with a 2 ml syringe. Serum osmotic pressure was measured by freezing point osmotic apparatus Gonotec OSMOMAT 3000.

Comparative transcriptome analysis of crabs under salinity stress

3 After a week of temporary cultivation, crabs with the same specifications were selected and transferred to a rectangular glass tank (length, width, height: 80cm×40cm×40cm) for salinity experiment. A salinity group (ppt 25) and a control group (ppt 0) were set up, three parallel crabs were placed in each group, and 10 crabs were placed in each parallel group. After salinity stress for 144 h, six crabs per tank were treated with tricaine methane sul fonate (MS 222, 200 mg/L) and posterior gill tissues were harvested immediately and stored at -80°C for RNA extraction, and the gill tissues were polled for RNA-Seq.

Total RNA Sample QC

4 Total RNA was extracted from gill tissues using TRIzol (Invitrogen) according to the manufacturer's instructions. All samples need to pass through the following three steps before library contamination
1 Nanodrop: Preliminary quantitation
2 Agarose Gel Electrophoresis: tests RNA degradation and potential contamination
3 Agilent 2100: checks RNA integrity and quantitation

Library Construction

5 After the QC procedures, mRNA from eukaryotic organisms is enriched using oligo(dT) beads. For prokaryotic samples, rRNA is removed using a specialized kit that leaves the mRNA. The mRNA from either eukaryotic or prokaryotic sources is then fragmented randomly in fragmention buffer, followed by cDNA synthesis using random hexamers and reverse transcriptase. After first-strand synthesis, a custom second-starand synthesis buffer (Illumina) is added with dNTPs, RNase H and Escherichia coli polymerase to generate the second strand by nick-translation. The final cDNA library is ready after a round of purification, terminal repair, A-tailing, ligation of sequencing adapters, size selection and PCR enrichment.

Library QC

6 Library concentration was first quantified using a TBS-380 Microfluorometer, and then diluted to 1 ng/μl before checking insert size on an Agilent 2100 and quantifying to greater accuracy by quantitative PCR (Q-PCR) (library activity >2 nM)

Sequancing

7 Libraries are fed into Hiseq machines according to activity and expected data volume

Raw data

8 The original raw data from Illumina HiSeqTM are transformed to Sequenced Reads by base calling. Raw data are recorded in a FASTQ file, contains sequence information (reads) and corresponding sequencing quality information.

A/T/G/C Content Distribution

9 The error rate for each base can be transformed by the the Phred score as in equation1(Qphred=-10log10(e)). The relationship between Phred quality scores Q and base-calling error e is given below:

10 GC content distribution is evaluated to detect potential AT/GC separation, which affects subsequent gene expression quantification. Theoretically, G should equal C, and A should equal T throughout the whole sequencing process for non-stranded libraries, whereAT/GC separation is normally observed in stranded libraries, For DGE (Digital Gene Expression) libraries, a large variation of sequencing error in the first 6-7 bases is allowed due to the use of random primers in library construction.

Data Filtering

11 Raw reads are filtered to remove reads containing adapters or reads of low quality, so that downstream analyses are based on clean reads. The filtering process is as follows:
(1) Discard reads with adaptor contamination.
(2) Discard reads when uncertain nucleotides constitute more than 10 percent of either read(N>10
(3) Discard reads when low quality nucleotides (base quality less than 20) constitute more than 50 percent of the read.

Mapping to a Reference Genome

12 Mapped regions can be classified as exons, introns, or intergenic regions. Exon-mapped reads should be the most abundant type of read when the reference genome is well-annotated. Intron-reads may be derived from pre-mRNA contamination or intron-retention events from alternative splicing. Reads mapped to intergenic regions are mainly because of weak annotation of the reference genome.

Expression Quantification

13 In order for the gene expression levels estimated from different genes and experiments to be comparable, the FPKM is used. In RNA-Seq, FPKM, short for the expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced, is the commonest method of estimating gene expression levels, which takes into account the effects of both sequencing depth and gene length on counting of fragments. HTSeq software was used to analyze the gene expression levels in this experiment, using the union mode. The result files present the munber of genes with different expression levels and the expression level of single genes. In general, an FPKM value of 0.1 or 1 is set as the threshold for determining whether the gene is expressed or not.

Comparison between Gene Expression Levels

14 To compare gene expression levels under different conditions, an FPKM distribution diagram and violin Plot are used. For biological replicates, the final FPKM would be the mean value.

RNA-Seq Correlation

15 Biological replicates are necessary for any biological experiment, including those involving RNA-seq technology. In RNA-seq, replicates have a two-fold purpose. Fiest, they demonstrate whether the experiment is repeatable, and secondly, they can reveal differences in gene expression between samples. The correlation between samples is an important indicator for the similarity of the samples. ENCODE suggests that the square of the pearson correlation coefficient should be larger than 0.92, under ideal experimental conditions. In this project, the R2 should be larger than 0.8. Heat maps of the correlation coefficient between samples are shown.

Differential Gene Expression Analysis List of Differentially Expressed Genes

16 The input data for differential gene expression analysis are readcounts from gene expression level analysis. Volcano plots are used to infer the overall distribution of differentially expressed genes. For experiments without biological replicates. Cluster analysis is used to find genes with similar expression patterns under various experimental conditions. By clustering genes with similar expression patterns, it is possible to discern unknown functions of previously characterized genes or functions of unknown genes. in hierarchical clustering, areas of different colors denote different groups of genes, and genes within each cluster may have similar functions or take part in the same biological process. In addition to the FPKM cluster, the H-cluster, K-means and SOM are also used to cluster the log2(ratios). Genes within the same cluster exhibit the same trends in expression leves under different condition. The Venn diagram presents the number of genes that are uniquely expressed within each group, with the overlapping regions showing the number of genes that are expressed in two or more groups.

GO Enrichment Analysis of DEGs

17 Gene Ontology (GO, http://www.geneontology.org/) is a major bioinformatics initiative to unify the presentation of gene and gene product attributes across all species. DEGs refer to differentially expressed genes. GO enrichment analysis is used by GOseq which is based on Wallenius non-central hyper-geometric distribution. Its characteristics are: the probability of drawing an individual from a certain category is

different from that of drawing it from outsides of the category, and this difference is obtained from estimating of the preference of gene length.

KEGG Pathway Enrichment Analysis of DEGs

18 The interactions of multiple genes may be involved in certain biological functions. KEGG(Kyoto Ebcyclopedia of Genes and Genomes) is a collection of manually curated databases dealing with genomes, biological pathways, diseases, drugs, and chemical substances. Pathways enrichment analysis identifies significantly enriched metabolic pathways or signal transduction pathways associated with differentially expressed genes compared with the whole genome background. Scatter diagram is a graphical display way of KEGG enrichment analysis results. In this plot, enrichment degree of KEGG can be measured through Rich factor, Qvalueand genes counts enriched to this pathway. Rich factor is the ratio of DEGs counts tothis pathway in the annotated genes counts. The more the Rich factor is, the higher is the degree of enrichment. Top 20 most significant enriched pathways are chosen in KEGG scatter plot, and if the enriched pathways counts is less than 20, then put all of them into the plot.