Transcriptome *Penaeus monodon*

9/22/2021

The samples used for this analysis were:

|  |  |
| --- | --- |
| Group | Sample ID |
| knockdown\_PmSTAT | 1\_S1 |
| knockdown\_PmSTAT | 2\_S2 |
| knockdown\_PmSTAT | 3\_S3 |
| WSSV\_infection | 4\_S4 |
| WSSV\_infection | 5\_S5 |
| WSSV\_infection | 6\_S6 |
| WSSV\_infection | 7\_S7 |
| WSSV\_infection | 8\_S8 |
| knockdown\_PmSTAT\_WSSV\_infection | 9\_S9 |
| knockdown\_PmSTAT\_WSSV\_infection | 10\_S10 |
| knockdown\_PmSTAT\_WSSV\_infection | 11\_S11 |
| knockdown\_PmSTAT\_WSSV\_infection | 12\_S12 |

## Quality pre-treatment

The raw data were filtered using *Trimmomatic* with this parameters:

1. Remotion of adapters and N's  
2. Average quality of reads > Q20

In the next table and graph we can observe the number of reads before (1.Raw\_data) and after (2.Quality\_filtered) the quality pre-treatment.

|  |  |  |  |
| --- | --- | --- | --- |
| Sample | Raw data | Quality filtered | % after quality |
| 1\_S1 | 16 202 030 | 16 097 983 | 99.35 |
| 2\_S2 | 12 890 555 | 12 788 438 | 99.20 |
| 3\_S3 | 839 588 | 830 354 | 98.90 |
| 4\_S4 | 14 467 820 | 14 371 003 | 99.33 |
| 5\_S5 | 13 316 992 | 13 245 638 | 99.46 |
| 6\_S6 | 43 615 302 | 43 404 612 | 99.51 |
| 7\_S7 | 13 142 443 | 13 070 733 | 99.45 |
| 8\_S8 | 13 871 940 | 13 791 480 | 99.41 |
| 9\_S9 | 10 403 660 | 10 335 646 | 99.34 |
| 10\_S10 | 14 615 414 | 14 500 097 | 99.21 |
| 11\_S11 | 14 428 705 | 14 355 271 | 99.49 |
| 12\_S12 | 11 557 788 | 11 479 499 | 99.32 |

![Read count before and after the quality filtering](data:application/pdf;base64,)

Read count before and after the quality filtering

## *De novo* assembly with Trinity

With the quality filtered data, we made a *de novo* assembly with Trinity. These are the *Trintity.fasta* stats:

|  |  |
| --- | --- |
| **Counts of transcripts, etc.** |  |
| Total trinity ‘genes’ | 586 928 |
| Total trinity transcripts | 778 680 |
| Percent GC | 37,7 |
| - | - |
| **Stats based on ALL transcript contigs** |  |
| Contig N10 | 3 271 |
| Contig N20 | 1 700 |
| Contig N30 | 799 |
| Contig N40 | 574 |
| Contig N50 | 472 |
| Median contig length | 331 |
| Average contig | 467,55 |
| Total assembled bases | 364 068 637 |
| - | - |
| **Stats based on ONLY LONGEST ISOFORM per ‘GENE’** |  |
| Contig N10 | 1 672 |
| Contig N20 | 727 |
| Contig N30 | 565 |
| Contig N40 | 479 |
| Contig N50 | 417 |
| Median contig length | 325 |
| Average contig | 409,1 |
| Total assembled bases | 240110795 |

To determine the representativeness of our samples in the assembly, each of them was realigned using bowtie2.

|  |  |
| --- | --- |
| Sample | % realigned |
| 1\_S1 | 76.34 |
| 2\_S2 | 90.59 |
| 3\_S3 | 87.12 |
| 4\_S4 | 89.03 |
| 5\_S5 | 88.54 |
| 6\_S6 | 91.74 |
| 7\_S7 | 89.92 |
| 8\_S8 | 85.38 |
| 9\_S9 | 92.13 |
| 10\_S10 | 84.76 |
| 11\_S11 | 91.25 |
| 12\_S12 | 78.82 |

![Percentage of reads realigned to the assembly](data:application/pdf;base64,)

Percentage of reads realigned to the assembly

## Differential expression with DESeq2

We used the DESeq2 with *Log Fold Change*>2 and *p-value*<0.05.

### Considering genes

|  |  |
| --- | --- |
| WSSV infection vs knockdown PmSTAT | DE genes |
| knockdown PmSTAT-UP | 56 |
| WSSV infection-UP | 813 |

|  |  |
| --- | --- |
| WSSV infection vs knockdown PmSTAT WSSV infection | DE genes |
| knockdown PmSTAT WSSV infection-UP | 138 |
| WSSV infection-UP | 675 |

|  |  |
| --- | --- |
| knockdown PmSTAT vs knockdown PmSTAT WSSV infection | DE genes |
| knockdown PmSTAT-UP | 0 |
| knockdown PmSTAT WSSV infection-UP | 0 |

![Correlation between samples considering genes](data:application/pdf;base64,)

Correlation between samples considering genes

![Hierarchical clustering considering genes](data:application/pdf;base64,)

Hierarchical clustering considering genes

### Considering isoforms

|  |  |
| --- | --- |
| WSSV infection vs knockdown PmSTAT | DE isoformsd |
| knockdown PmSTAT-UP | 92 |
| WSSV infection-UP | 496 |

|  |  |
| --- | --- |
| WSSV infection vs knockdown PmSTAT WSSV infection | DE isoforms |
| knockdown PmSTAT WSSV infection-UP | 474 |
| WSSV infection-UP | 1242 |

|  |  |
| --- | --- |
| knockdown PmSTAT vs knockdown PmSTAT WSSV infection | DE isoforms |
| knockdown PmSTAT-UP | 0 |
| knockdown PmSTAT WSSV infection-UP | 0 |

![Correlation between samples considering all isoforms](data:application/pdf;base64,)

Correlation between samples considering all isoforms

![Hierarchical clustering considering all isoforms](data:application/pdf;base64,)

Hierarchical clustering considering all isoforms