Title option 1:
Title option 2:
Title option 3:

Abstract

Introduction

Methods

Experiment

Transcriptome Assembly (Trinity)
        Sequencing reads from all six libraries were assembled using Trinity (version 2.0.6). As part of the Trinity package reads were quality trimmed (Trimmomatic) and normalized prior to assembly with minimum kmer coverage of 2. Assessment of transcriptome quality and completeness was performed using Transrate. To further assess completeness of the transcriptome contigs were compared to ???(Cgigas and Moriera Mytilus)???? The Transdecoder package that is part of Trinity was used to predict putative corresponding proteins.

Transcriptome Annotation (Trinity)
        Resulting contigs were first compared to the NCBI nucleotide (nt) database to identify any non target taxa sequences (ie bacteria) and these were removed from further analysis. Contigs were annotated by comparing contiguous sequences to the UniProtKB/Swiss-Prot database. Comparisons were made using the BLASTx algorithm with a 1.0E-5 e-value threshold. Genes were classified according to Swiss-Prot Gene Ontology (GO) associations, as well as respective parent categories (GO Slim). Annotation analyses and data are published

Differential Expression

Long non-coding RNA identification (CLC + online tools)

Results

Experiment

Transcriptome Assembly (Trinity)

Following quality trimming, 792,714,472 (99%) of reads were assembled into 184834 transcripts corresponding to 110408 genes. Trinity fasta file(184834 seqs)
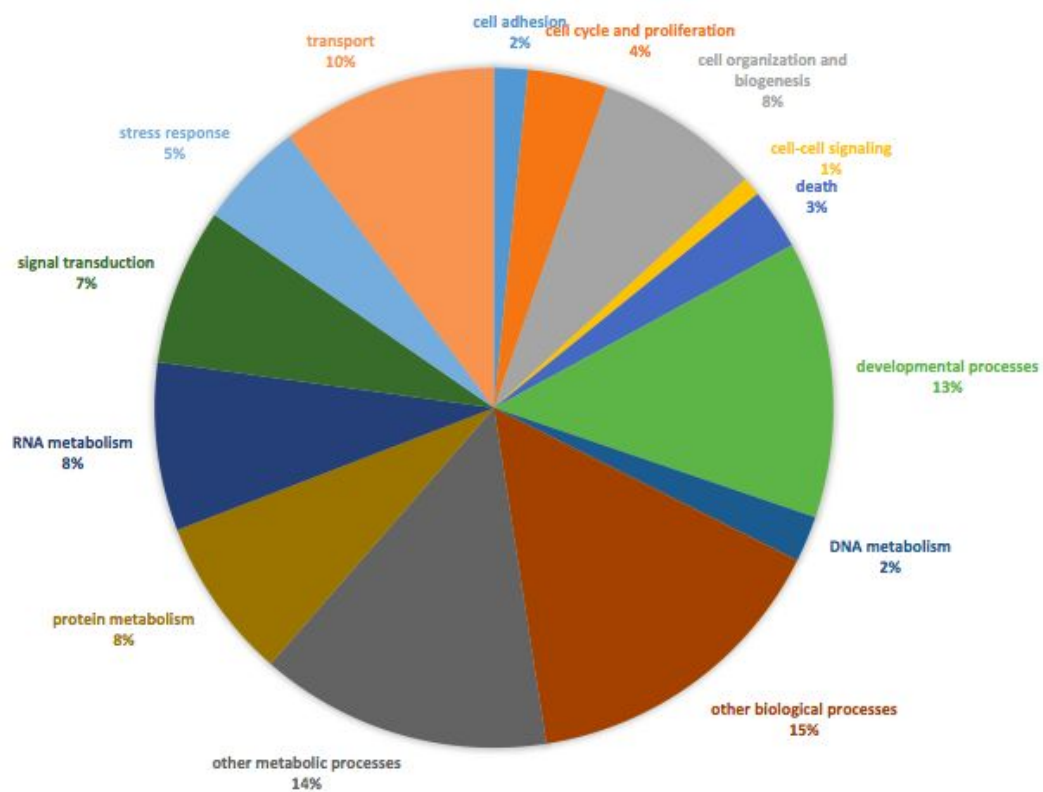
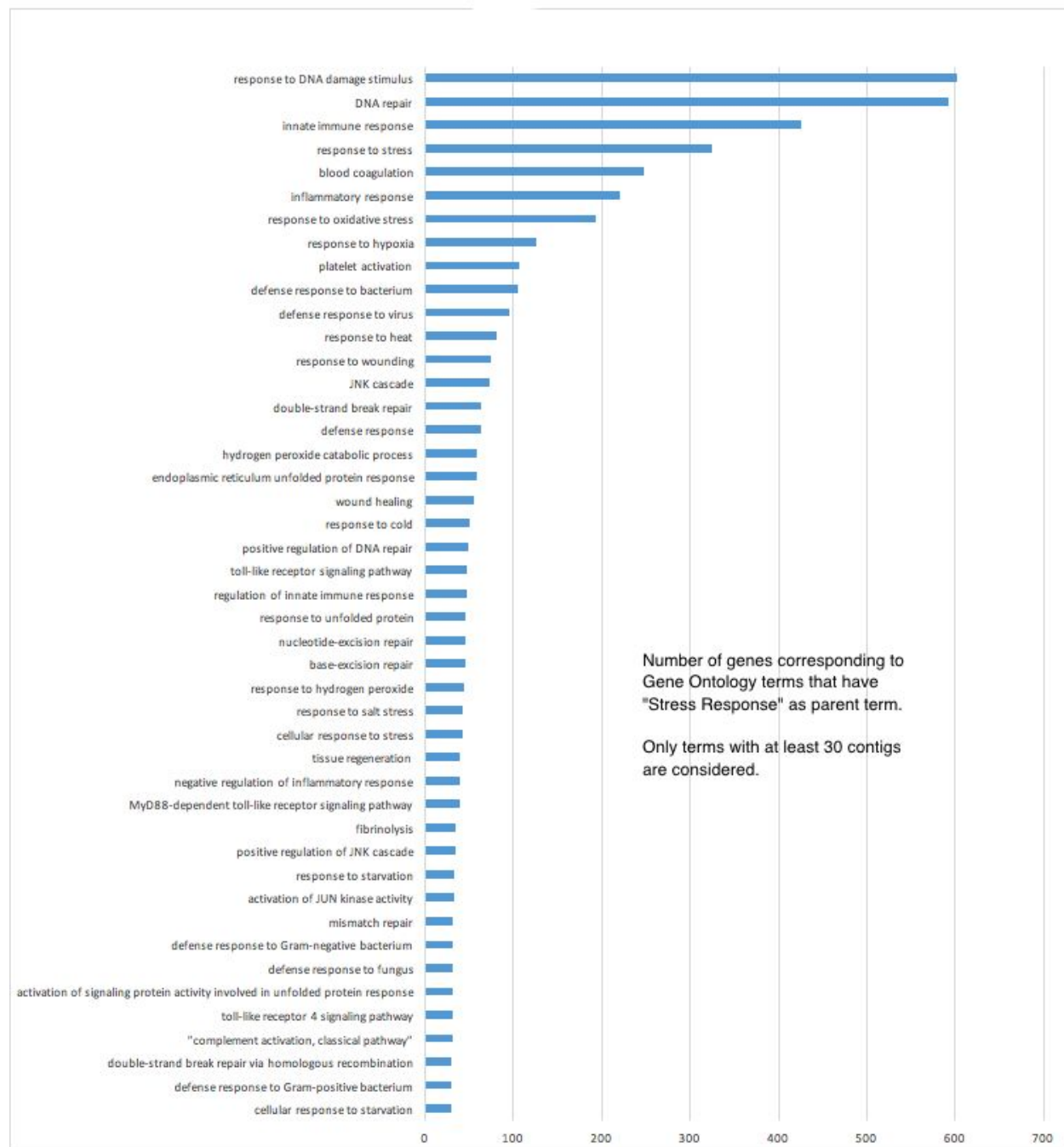After removing sequences (668) with significant matches to not Eukaryota taxa 184,166 contigs remained. Fasta file.

Transcriptome Comparison

assembly	CLC	Trinity
n_seqs	138883	184166
smallest	54	224
largest	12602	21486
n_bases	72930134	134270562
mean_len	481.7897	729.07356
n_under_200	38556	0
n_over_1k	16777	38651
n_over_10k	9	7
n_with_orf	22343	42761
mean_orf_percent	55.68005	52.36376
n90	311	303
n70	558	603
n50	1013	1077
n30	1839	1745
n10	5601	3161
gc	0.33684	0.33357
gc_skew	0.0018	0.01875
at_skew	0.001	0.0111
cpg_ratio	1.43512	1.41025
bases_n	4	0
proportion_n	0	0
linguistic_complexity	0.09772	0.1295

Transcriptome Annotation (Trinity)

Comparison with Uniprot Swiss-Prot database resulted in 29,486 contigs with annotations.

Number of genes corresponding to Gene Ontology terms that have "Stress Response" as parent term.

Only terms with at least 30 contigs are considered.

Differential Expression

Long non-coding identification (CLC + online tools)

Table I. Long non-coding sequences after data filtering steps.

Item	Number of obtained sequences	Number of discarded sequences
De novo assembly	138833	0
Coverage (average coverage of contigs > 50)	38609	100224
ORF identification (sequence with ORF > 200 were discarded)	23492	15117
Coding potential (CPAT)	22308	1148
Contig length (> 250 bp)	16012	6296
blastX against mollusca proteins	12714	3298
Conserved Domains Search		
blastn against nr genbank database		

Discussion