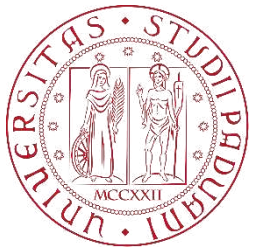


UNIVERSITÀ DEGLI STUDI DI PADOVA

Department of Comparative Biomedicine and Food Science

School of Agricultural Sciences and Veterinary Medicine

Second Cycle Degree (MSc)

In Biotechnologies for Food Science

Comparative analysis of RNA-seq protocols: a case study on heat stress response in *Ruditapes philippinarum*

Supervisor

Prof. Luca Peruzza

Submitted by

Miriam Quarzago

Student n. 2082170

ACADEMIC YEAR 2023 – 2024

ABSTRACT

RNA-sequencing is a technique used to study transcriptome (defined as the sum of all mRNA molecules expressed by genes) within cells, tissues or organisms. It gives information on which genes are more expressed by comparing two (or more) conditions and provides a set of tools to retrieve biological information of genes (i.e. their involvement in pathways/processes). A plethora of tools are available to perform such tasks.

In order to perform such comparative analysis, a dataset available on NCBI was used. In this case the RNA was extracted from the digestive gland of *Ruditapes philippinarium* exposed to heat stress, 30°C for 30 days and compared to controls maintained at 20°C. Sequencing data, stored on NCBI in SRA format, were downloaded using SRA toolkit and then sequences were converted in FASTQ format. Quality control analysis was performed using FastQC and trimmed using Fastp. STAR tool was used for generating genome indexes based on reference genome and its annotation in GTF format and then used for read mapping. FeatureCounts program was used for quantifying read alignment to genomic features.

This thesis investigated the impact of different RNA-seq approaches on the number of differentially expressed genes (DEGs) and the subsequent biological impacts. The first aspect was related to the different impact of two alternative settings on STAR. These setting diverge in their output filtering option and alignments strategies. The two different setting showed some variation in mapping rates.

The second was to compare between two different approaches for data normalization: log2(CPM+c) transformation and RUVSeq method. For log2(CPM + c) transformation was used by iDEP(integrated differential expression and pathway analysis), a web application. RUV(Removal of Unwanted Variation) used factor analysis to remove unwanted variation in particular it was employed RUVseq that used negative control samples. It helped to eliminate systematic artefact and improving the detection of true biological signals.

After normalization, differential expression analysis was conducted using DESeq2 for both methods. Gene Set Enrichment Analysis (GSEA) was used for pathways analysis on iDEP. GSEA was also implemented after RUVseq using clusterProfile package.

The two methods, iDEP and RUVseq+DESeq2 showed some differences in the number of differentially expressed genes (785 with RUVSeq+DESeq2 and 659 with iDEP), and also

some differences in the functional results obtained. Only 19 pathways were shared among the two methods out of a total of 111. The main biological processes that were highlighted were: cilium motility and structure, sperm motility and structure, alteration of non-canonical Wnt signalling, while mitochondria were enhanced by heat stress.

Such comparative analyses are useful in order to reveal differences in performances among RNAseq pipelines which may lead to slight differences in the results that are obtained.

RIASSUNTO

L'RNA-sequencing è una tecnica usata per studiare il trascrittoma (definito come la somma di tutte le molecole di mRNA espresse dai geni) all'interno delle cellule, tessuti o organismi. Ci da informazioni su quali geni sono più o meno espressi durante la comparazione di due (o più) condizioni e offre una serie di strumenti per ottener informazioni biologiche sui geni (ad esempio, il loro coinvolgimento in vie/pathways biologiche). Esiste una moltitudine di strumenti disponibili per svolgere tali compiti.

Per compiere questa analisi comparativa è stato usato un dataset su NCBI. In questo caso l'RNA è stato estratto dalla ghiandola digestiva di diversi esemplari di vongola Filippina *Ruditapes philippinarium* esposti ad uno stress termico corrispondente a 30°C per 30 giorni mentre gli animali di controllo sono stati mantenuti a 20°C.

I dati di sequenziamento archiviati su NCBI con il formato SRA, sono stati scaricati usando SRA toolkit, poi le sequenze sono state convertite in formato FASTQ. Il controllo qualità delle sequenze è stato fatto usando fastQC e il taglio delle sequenze è stato fatto usando Fastp. L'indice genomico è stato creato usando STAR basandosi su un genoma di riferimento in formato GTF ed è stato utilizzato per quantificare e allineare le sequenze al genoma.

In questa tesi è stato esaminato l'impatto di diversi approcci per l'RNA-seq sul numero dei geni diversamente espressi e sul conseguente impatto biologico. Il primo aspetto era relativo ai diversi effetti di due diversi settaggi su STAR per il mapping, Le due impostazioni hanno mostrato alcune variazioni nei tassi di mapping. Il secondo aspetto è stato comparare due diversi approcci per la normalizzazione dei dati: la trasformazione log2(CPM+c) e il metodo RUVSeq sono stati utilizzati. La trasformazione log2(CPM+c) è stata applicata tramite un'applicazione web, iDEP ("integrated Differential Expression and pathway Analysis"), RUV ("Removal of Unwanted Variation") ha utilizzato l'analisi dei fattori per rimuovere le variazioni indesiderate, in particolare è stato usato RUVSeq, che ha utlizzato campioni di controllo negativi. Questo ha aiutato ad eliminare errori sistematici e migliorare il rivelamento dei segnali biologici reali.

Dopo la normalizzazione, l'analisi per i geni differentemente espressi è stata fatta usando DESeq2 per entrambi i metodi. GSEA ("Gene Set Enrichment Analysis") è stata usata per

l’analisi delle pathways su iDEP. GSEA è stata anche implementata dopo RUVseq usando il pacchetto clusterProfile.

I due metodi, iDEP e RUVseq+DESeq2 presentano alcune differenze nel numero di geni diversamente espressi (785 con RUVSeq+DESeq2 e 659 con iDEP), anche alcune differenze nei risultati funzionali ottenuti. Solo 19 pathways in comune sono state trovate tra i due metodi su un totale di 111. I principali processi biologici che sono stati evidenziati sono: motilità e struttura delle cilia, motilità e struttura degli spermatozoi, alterazione delle vie di segnalazione Wnt, e l’attività dei mitocondri.

Questa analisi comparativa è stata utile per rivelare le differenze tra le pipeline di RNA-seq, che possono portare a lievi differenze nei risultati ottenuti.

1. INTRODUCTION 7
2. MATERIALS AND METHODS 13
2.1 Data source 13
2.2 Experimental conditions 13
2.3 SRA Toolkit 13
2.4 FastQC 14
2.5 Fastp 14
2.6 Genome index 14
2.7 Mapping 15
2.8 FatureCounts 15
2.9 iDEP Workflow for Differential Gene Expression Analysis 16
2.10 DEGs Analysis with RUVSeq and DESeq2 in R 16
2.11 Pathway analysis using iDEP 17
2.12 Pathways analysis using Cluster Profile 17
3. RESULTS 19
3.1 Quality check and trimming 19
3.2 Differential gene expression analysis 20
4. DISCUSSION 32
4.1 Comparative analysis 32
4.2 Pathways analysis 33
4.2.1 Cilium motility and structure downregulation: 33
4.2.2 Sperm motility and structure downregulation: 34
4.2.3 Mitochondrial gene expression 35
4.2.4 Non-canonical Wnt signalling 36
5. CONCLUSION 38
6. REFERENCES 39
7. ACKNOWLEDGEMENTS 46

1. INTRODUCTION

Climate change is a long-term shift in global or regional climate patterns. Often, climate change refers specifically to the rise in global temperatures from the mid-20th century to present day. Since the pre-industrial period, human activities are estimated to have increased Earth's global average temperature by about 1 degree Celsius. Climate change may cause weather patterns to be less predictable and is also associated with more frequent and more intense hurricanes, floods, downpours and heat waves [1].

Excessive global warming is a consequence of human activities, particularly the burning of fossil fuels like coal, oil, and natural gas, which are the main drivers of modern climate change. Those combustibles release greenhouse gases, such as carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O), in the atmosphere. These gases absorb heat, solar radiation, including infrared, and re-emit some of it back towards the Earth's surface. As a result, the heat remains trapped, contributing to the overall warming of the planet [2]. Greenhouse effect is a natural phenomenon; the increasing of gases released by anthropogenic intensify this effect contributing to global warming.

The oceans are a major sink for atmospheric CO_2 after forest partially mitigating temperature increases since photosynthetic organism fix CO_2 to generate organic matter [3]. The ocean absorbs most of the excess heat from greenhouse gas emissions, leading to rising ocean temperatures. Leading ocean temperatures affect marine species and ecosystems. Rising ocean temperatures also affect the benefits humans derive from the ocean, threatening food security, increasing the prevalence of diseases, causing more extreme weather events, and the loss of coastal protection [4]. Food production from fisheries and shellfish aquaculture in some oceanic regions has been negatively affected.

As ocean temperatures continue rising, marine heatwaves are expected to be more frequent and intense. This had an impact on global water cycle, including its variability, global monsoon precipitation. In the ocean, ice sheets are melting causing global sea level to increase. Ocean acidification, ocean deoxygenation and global mean sea level will continue to increase in the 21st century, at rates dependent on future emissions increasing frequency of marine heatwaves will increase risks of biodiversity loss in the oceans, including mass mortality events [2]. A marine heat wave is usually defined as a consistent area of exceptionally warm sea surface temperature that lasts for days to months. MHWs

have been observed in all of the main ocean basins over the past ten years, also depend on the location and season of occurrence. These mechanisms can be attributing to large and small-scale oceanic forcing, atmospheric forcing or a mix of both, and the dominant mechanisms also depend on the location and season of occurrence. The reported biological impacts include: adjustments in the geographic distribution of species, widespread changes in species composition, harmful algal blooms, mass stranding's of mammals and mass mortalities of particular species. Mass mortality of marine species [5].

In the study Garrabou et al. [6] authors highlight the consequences of these extreme events focused on Mediterranean Sea. Marine heatwaves have increased significantly in intensity and frequency over the past 20 years. Heatwaves has been associated with the increase of mass mortality events. Mediterranean area is sensitive to this phenomenon, effecting more than 40 species from various taxa (e.g., Porifera, Cnidaria, Bivalvia, Bryozoa, Ascidiacea) around the coastal regions, the two severe mass mortalities occurred in 1999 and 2003. Using high-resolution, satellite-derived sea surface temperature (SST) and subsurface coastal seawater temperature data (from 5 to 40 m depth), they were able to discover that the Mediterranean Sea experienced a period of exceptional thermal conditions (2015–2019); during this period, some species exhibited a mortality rate of up 80%. Quantified mass mortality data were obtained from T-MEDNet, a temperature database used for long-term collaborative initiative focused to building a pan-Mediterranean observation network on climate change effects in marine ecosystems. High temperatures during this period showed some regional differences with the Aegean and Levantine seas that saw warming rates higher than the other regions (1.4°C and 1.3°C).

Some species are more vulnerable to these events, such as cnidarinas, like gorgonian coral; this group accounts for 54% of the overall mass mortality records. Correlation between mass mortality and marine heatwave was confirmed, however it remains a complex phenomenon since the biological response to heat stress is not fully understood. Improving monitoring cold be useful for safeguarding the resilience of Mediterranean ecosystem and its ability to provide services, such as habitat for commercially valuable species, to human societies [6].

Among this species there are bivalves, marine bivalves are found as key components of coastal habitats and have several important roles, such as serving as a food source. In fact, as the world's population continues to grow, bivalve aquaculture is expected to increase in importance as a means of addressing demands for animal protein; however, their growth may be compromised by unfavourable climatic conditions [7]. Bivalves have soft body

protected by the shell made of proteins and calcium carbonate. The nervous system is simple and organized in ganglia connected by nerve cords, the circulatory system is open with heart that pumps haemolymph in the body. Bivalves also have gills that have multiple roles, including respiration, feeding, and are involved in the movement of water through the mantle cavity [8]. Gonads are represented by a diffused tissue closely linked to the digestive system [9].

An example of commercial valuable species is *Ruditapes philippinarum. R. philippinarum* is originally from the Indo-Pacific region, it has been introduced in many countries for fisheries and aquaculture, including estuarine environments along Atlantic and Mediterranean European coasts [10]. Is widely distributed in China, Korea, and Japan. It is a bivalve animal with high economic value, and it is subjected to stresses from the environment [11]. *R. philippinarum* has an important ecological role and socio-economic value, from the Atlantic and Mediterranean to the Indo-Pacific region [9]. *R. philippinarum* is gonochoric, sexually reproducing species in which individuals have one of at least two distinct sexes [12]. Reproduction takes place at different times depending on geographical area, the release of eggs and sperm usually occurs between 20-25°C and two spawning events may occur in the same season. Temperature and feeding are two primary factors influencing gametogenesis. The stages of life cycle in bivalves are represented in Fig. 1

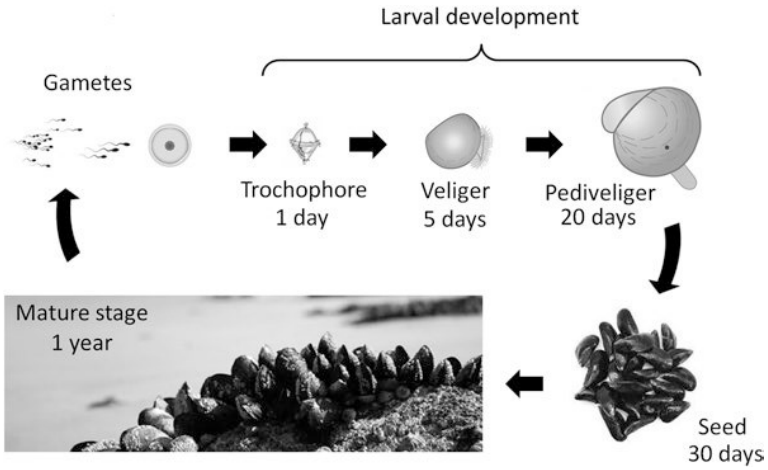


Figure 1. Life cycle of a representative bivalve. Figure taken from Sorgeloos (1999).

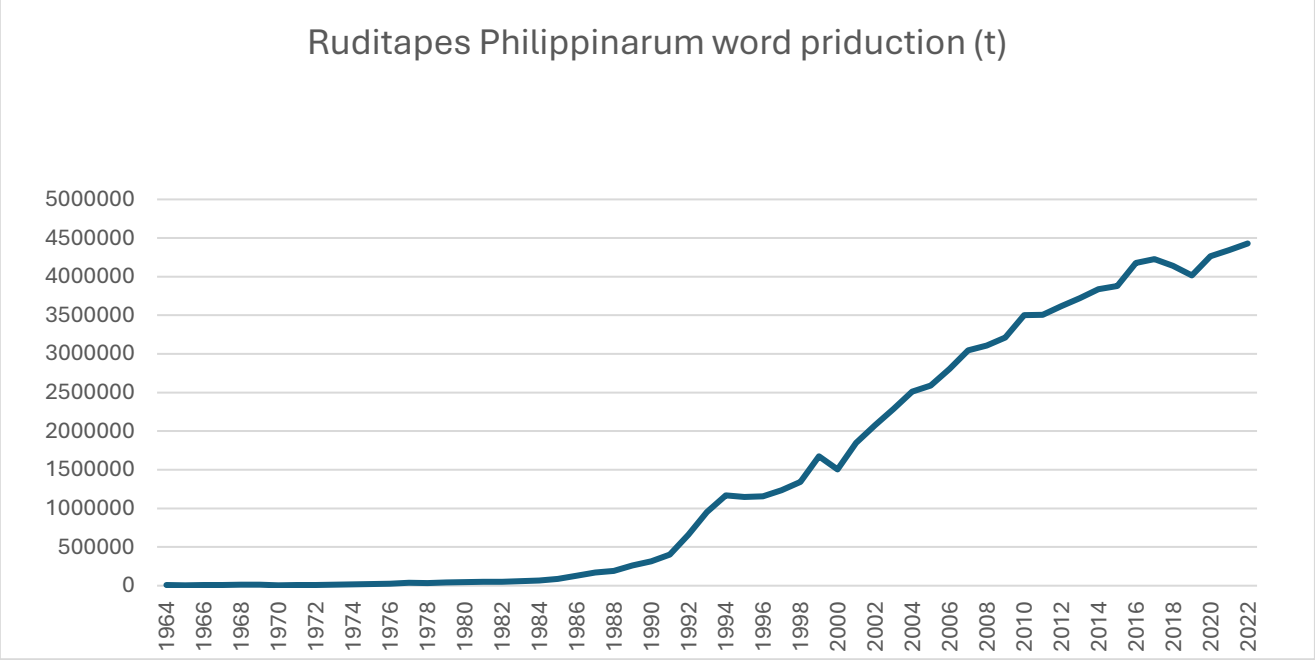


FIGURE 2. *R. Philippinaurm* production during the years in tonnes (FAO)

Based on FAO statistics, the global production of *R.* philippinarum increased during the years as depicted in Fig. 2 [13]. However, in the last years heatwave have became a major problem that is affecting *R. philippinarum* production, as reported by Ponti et al. [17], a study is focused on *R. philippinarum* in the Pialassa Baiona lagoon in the northern Adriatic Sea (Ravenna). *R. philippinarum* was introduced in Venice's lagoons in 1983 for aquaculture and spread in all Mediterranean Sea. It replaced in many areas the native species *Ruditapes decussatus*. Between 2002 and 2003 the population of *R. philippinarum* drastically declined. Stock estimates a reduction from 36,800 kg in July 2002 to 10,300 kg in October 2003. The decline of stock could be due to overfishing and summer heat wave occurred in 2003, which may have reduced larval survival. The heat wave was caused by anomalous atmospheric high-pressure conditions, the water temperature was higher than 31°C corresponding to the limit of tolerance of *R. philippinarum*, also the water conditions were worsened by eutrophication in lagoons [14].

In general, climate change has different effects on aquatic species, including reproductive disruption, immune system compromission, increased diseases and parasite infections due to warmer temperatures that boost parasite metabolism, altered development and growth [15] [16] [17]. Given the vast range of effects on animal biology, it is important to investigate these phenomena, RNA sequencing could be a method for identifying pathways and genes

influenced by stress due to climate change. To explore this effect, at molecular level is essential to consider the transcriptome, which is the complete set of RNA transcripts that are produced by the genome, under specific circumstances or in a specific cell [18]. RNA sequencing (RNA-seq) is a technique used to determine the presence and abundance of RNA/transcripts in a biological sample at a specific time, revealing which genes are expressed and more generally what genomic regions are transcribed. RNA-seq studies, referred to as massively parallel RNA sequencing or RNA high-throughput sequencing, subsequently use next-generation sequencing (NGS), a cost-effective technology that sequences millions of DNA fragments in parallel, providing high-depth sequenced reads at a relatively quick speed. For NGS purposes, high depth implies that each mRNA molecule is sequenced several times to decrease detection errors, and thus resulting in increased accuracy [19]. Gene expression is the fundamental level at which the results of various genetic and regulatory programs are observable [20]. Gene expression studies were used to get a snapshot of the RNA molecules present in a biological system, which dictates what cells are doing or are capable of [20].

RNA-seq has become ubiquitous for studying the gene expression and transcript identification of cells or organisms. Transcriptomics methods are still under active development, and computational theories, analysis pipelines, and integrated databases are continuously proposed [19]. RNA sequencing (RNA-seq) provides a quantitative and open system for profiling transcriptional outcomes on a large scale and therefore facilitates a large diversity of applications, including basic science studies, but also agricultural or clinical situations [20]. RNA-seq is driven by its large number of applications. One of the main application areas is gene regulation. RNA-seq enables the comparison of gene/transcript/exon expression between different tissues, cell types, genotypes, stimulation conditions, time points, disease states, growth conditions, and so on. Ultimately, the goal of such comparisons is to identify the genes that change in expression to understand the molecular pathways that are used or altered or the regulatory components that are utilized [20].

As the analysis of RNA-seq data is complex, it has prompted a large amount of research on algorithms and methods. This has resulted in a substantial increase in the number of options available at each step of the analysis. Consequently, there is no clear consensus about the most appropriate algorithms and pipelines that should be used to analyse RNA-seq data [21]. In this thesis two different setting of mapping using STAR were compared, analysing

the different alignment matrix provided in the bam files including, number and percentage of uniquely mapped reads, reads mapped to multiple loci, and unmapped reads. Additionally, two different gene expression analysis protocols were compared: iDEP and RUVSeq+DESeq. The analysis was based on a publicly available dataset focussed on the effect of heat stress on *R. philippinarum* in which the treatment condition was 30°C for 30 days for the heatwave treated animals, while it was 25°C for the control animals. This dataset allowed for a comparison of the two methods. Many differentially expressed genes were shared between iDEP and RUVSeq+DESeq2, among the unique genes not shared between the methods, the functions were similar despite the genes being different. This highlights that, while the overall biological conclusions are still unchanged between the two methods, changing the normalization method can lead to minor differences in the results that may in turn provide a slight different point of view on some aspects of the results generated via RNA-seq.

For pathways analysis the main difference between methods was identified since only 19 pathways were shared among the method; this could be due different parameters in the algorithms used to calculate GSEA or in slight differences in which the null distribution is calculated when estimating p-values.

2. MATERIALS AND METHODS

2.1 Data source

In this thesis, RNA sequencing data was used. Data files were sourced from the Sequence Read Archive. SRA is a database in which there are stored raw sequencing data and alignment information to enhance reproducibility and facilitate new discoveries. This archive is one of the many that can be found on NCBI website, National Center for Biotechnology Information [22].

Specifically, the data were derived from BioProject PRJNA90652 [23]. For library preparation, the QuantSeq 3' mRNA-Seq Library Prep Kit FWD was employed (Lexogen GmbH, Vienna), which is specifically designed to generate Illumina compatible libraries of sequences close to the 3' end of polyadenylated RNA. The sequencing was performed with Illumina NovaSeq 6000; this instrument is able to generate up to 20 billion reads per run with a maximum length of 250 bp. In this case, it was set up for generating single-end 75 bp.

2.2 Experimental conditions

R. philippinarum were purchased from the hatchery SATMAR (France). HW clams were exposed to a simulated Marine Heatwave (MHW) of 30 °C for 30 days, while CTRL clams were kept at 20 °C for 30 days. Each treatment was replicated on 4 independent tanks. A total of 240 clams were used. The temperature was constantly monitored, the water was regularly changed, and the clams were adequately fed to ensure optimal condition. RNA was extracted, after 30 days of heat stress exposure from digestive gland of *R. philippinarum.* This organ is a key organ that regulates many fundamental processes involved in the digestion and absorption of nutrients and storage of lipids and carbohydrates. This organ in *R. philippinarum* is surrounded by a thin layer of reproductive tissue, which extends further in the body mass of the animals [23].

2.3 SRA Toolkit

The SRA Toolkit from NCBI is a collection of tools for accessing data in the INSDC (International Nucleotide Sequence Database Collaboration). Within the toolkit is possible to find prefetch and fastq-dumb tools [24]. The prefetch tool downloads the sequencing files; it was possible to download multiple sra files starting from the list of accession

numbers. fastq-dump tool converts these files into FASTQ. format. FASTQ format is a text-based file format for storing sequences and their corresponding quality scores. Both the sequence letter and quality score are each encoded with a single ASCII character for brevity. This file has four lines: the first is for the sequence title and optional description, the second is for raw sequence, the third is the separator, and the fourth is the quality score. Once retrieved, fastq files were then analysed with the use of fastQC [25].

2.4 FastQC

FastQC was used to conduct quality control checks on the raw sequence data, helping to detect problematic areas within the reads. This tool is able to identify numerous issues, including adaptor sequences, overrepresented sequences, GC content imbalances, low quality regions, high N content, and sequence length distribution irregularities [26]. In this case, there was always the presence of adaptors that needed to be trimmed, and the quality needed to be improved; in some cases, the GC content was not regular. To “resolve” issues raised via QC analysis, trimming was performed. Trimming with Fastp was focused on the problems individuated by FasQC report. After the trimming, another FastQC report was created to be sure the issues were resolved.

2.5 Fastp

Fastp is a highly efficient tool designed for FASTQ data preprocessing, combining multiple functions such as quality control, adapter trimming, and quality filtering in a single operation [27]. In this case, fastp was used to trimmed the adapter sequence “AAAAAAAAAA”. By default, Fastp automatically applies those other settings: trimming reads with Phred score below 15 with a percentage of 40% of bases that can be unqualified. These options were sufficient for most of the reads since from previous fastQC report adapter and quality score were the principal issues. Additionally, the tool generates an HTML report and provides an informative overview of the data before and after trimming.

2.6 Genome index

A genome index was generated using the STAR (version 2.7.11b). The index was built from the reference genome file in the FASTA format. The option was: genomeFastaFiles. GTF file containing gene annotations was used to extract known splice junctions and then builds “spliced” sequences by deleting intron sequences, i.e., joining the sequences of the exons. This is highly recommended since it allows for a more accurate mapping of the spliced reads [28]. The option used for this purpose was sjdbGTFfile. Other options used were genomeDir

Index, which specifies the directory where the genome index files were stored; jdbOverhang was setting at 75 because the length of the reads was 76.

2.7 Mapping

STAR (Spliced Transcripts Alignment to a Reference) is an RNA-seq mapper that performs highly accurate spliced sequence alignment at an ultrafast speed. STAR alignment algorithm can be controlled by many user-defined parameters [28].

In this thesis, two different settings were chosen; For run# 1 specific options were taken from Lexogen's website, run2# had more default settings. The basic settings used for the alignment were the same for both Run#1 and Run#2 . The runThreadN option, specifying the number of CPU threads to be utilized for parallel processing. The genomeDir parameter defined the directory where the genome indices were stored, allowing STAR to efficiently map the reads to the reference genome. RNA-seq reads, provided in FASTQ format, were specified using the readFilesIn option, while the sjdbGTFfile parameter was used to input the path to the GTF file containing annotated transcripts. This GTF file helps to improve the accuracy of spliced alignments by incorporating known junctions. The sjdbOverhang value, set to the length of the reads minus one (75), was used to optimize splice junction detection by defining the number of bases to be included on each side of a junction. For run# 1 specific options were taken from Lexogen's website, run2# had more default settings.

From bam file were extracted information for see the difference between the two settingincluding the number and percentage of uniquely mapped reads, reads mapped to multiple loci, and unmapped reads, were extracted. Unmapped reads were further categorized based on specific reasons, such as excessive mismatches or insufficient read length.

2.8 FatureCounts

FeatureCounts is a read summarization program suitable for counting reads generated from RNA sequencing [29]. The data input were BAM files (Binary Alignment/Map) obtained after mapping. The options utilized were: -a, -o, -M and -fraction.

-a: indicates the genome annotation file.

-o: specifies the name of the output file, which includes the read counts.

-M (countMultiMappingReads): this option was used to count multi-mapping reads.

-Fraction: split the count of multi-mapped reads based on how many locations read maps to. If a read mapped in 3 different places, it was not counted 3 times but was divided among locations to obtain a more proportional count. FeatureCount with these options ensures that also the read that maps to multiple locations were maintained, keeping that important library information. The count table obtained from this was used for differential expression analysis.

2.9 iDEP Workflow for Differential Gene Expression Analysis

To identify differentially expressed genes (DEGs), two methods were used: iDEP [30], a web-based application, and the R packages RUVSeq and DESeq2. The steps used with the iDEP (integrated Differential Expression and Pathway analysis) are briefly explained here. Initially the expression matrix was loaded; this file was made from the count tables derived from FeatureCount step. Then the experimental design file was loaded; this file recapitulates the relationship between samples and treatments. In addition to those two files, iDEP offers the possibility to load annotation files for non-model species. Since *R. philippinarum* is not present on iDEP's library, a custom ".gmt" (Gene Matrix Transposed file format) file was used. This annotation file is a custom file that describes the gene sets of *R. philippinarum.*

The second step of iDEP was the pre-process; in this step, the data were normalized: only genes with minimal 1 count per million (CPM) in at least 20 libraries were kept. EdgeR: log2(CPM+c) was selected as the method for transform counts data for clustering and PCA. Pseudocount c=4 was chosen based on the suggestion reported on iDEP; this pseudocount was added to all counts before log transformation. The bigger this number is, the less sensitive it is to noise from low counts. 20 libraries were chosen since there were 20 control and 20 treatment replicates. iDEP provided principal component analysis (PCA). It was useful to project samples into two-dimensional space; data inspection via PCA was used to assess the presence of outlier samples in the dataset and, in case, remove them. In this context, one sample, sample CTRL_77 was removed and the whole RNA-seq pipeline was restarted after its removal. Finally, iDEP offers the possibility to determine the differentially expressed genes (DEGs) via the DESeq2 library [30]. The FDR (false discovery rate) cutoff was set to 0.1, and the minimum fold change was |2|. The comparison performed was heatwave-treated vs control samples.

2.10 DEGs Analysis with RUVSeq and DESeq2 in R

To compare the results obtained via iDEP, a similar approach was undertaken with the use of the library RUVseq (Remove Unwanted Variation from RNA-seq Data)[31] which was

used for normalization. Following the normalisation, differential gene expression analysis was executed with DEGseq2 R package [32]. For this approach the RStudio software was utilized. After loading the gene counts, lowly expressed (and non-expressed) genes were filtered out by maintaining those with more than five reads in at least two samples. To visualize sample quality before RUVseq action, RLE (Relative Log Expression) and PCA (Principal Component Analysis) plots were generated using the plotRLE, plotPCA functions from EDASeq package [33]. After this, RUVseq was applied. in particular RUVs that used replicate samples for which the covariates (HW vs CTRL) of interest are constant; in this case, K = 1 was used to account for a single factor of unwanted variation. The K value was chosen after testing values from 1 to 8. A value of 1 was sufficient to clearly separate the treated samples from the controls. After this, the RLE and PCA plots were created to verify the effectiveness of the RUVs correction. Differential expression analysis was performed using the library (DESeq2). A volcano plot was created using EnhancedVolcano library[38] and also two bar plots; one with the total number of up and down expressed genes and the other with the log fold change distribution, were generated using ggplot2 package[34]. To compare the DEGs resulted from iDEP with those obtained via RUVs+DESeq2, a Venn diagram was constructed using an online tool from Bioinformatics & Evolutionary Genomics [35].

2.11 Pathway analysis using iDEP

Using iDEP, pathway analyses were done using fold-change values returned by DESeq2. Pathway analysis was performed using GSEA (gene set enrichment analysis) method [36]. The options were set to perform the comparison “treatment vs control”. The significance cutoff (FDR) was set at 0.05 and was used to determine statistically significant pathways; the gene set size was chosen between 5 and 500 genes. The custom .gmt file with the annotation was used to identify pathways based on DEGs.

2.12 Pathways analysis using Cluster Profile

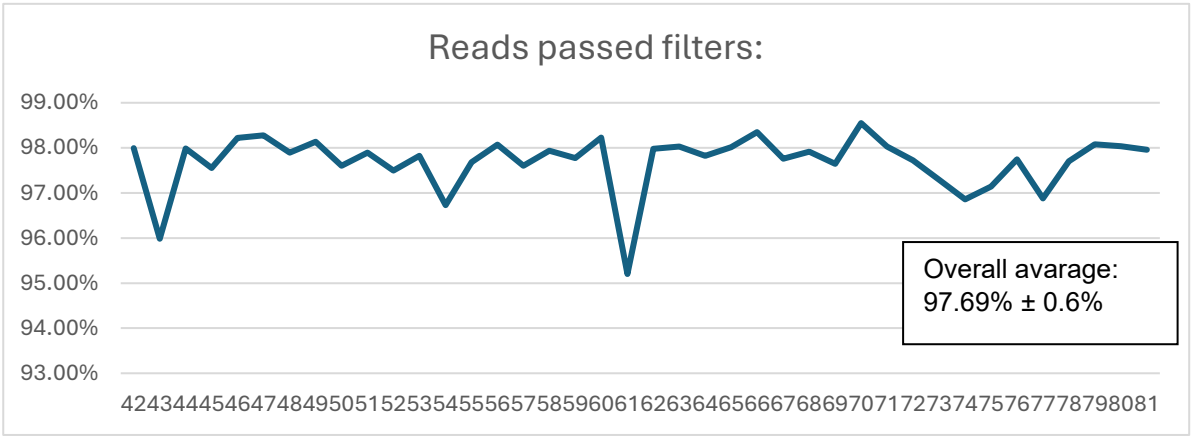
Pathway analysis using the approach RUVs+DESeq2 was performed with the library ClusterProfiler [37]. The results derived from the previous differential expression analysis were loaded and the custom .gmt annotation file was uploaded in R Studio. Also in this case, the analysis was performed using Gene Set Enrichment Analysis (GSEA) approach implemented in the clusterProfiler package. The settings were: exponent value of 1; adjusted p-value (the BH method was used to adjust for multiple testing) cutoff < 0.05; the epsilon (eps) value was set to 0; the analysis was performed with the fgsea method implemented in

the fgsea package; finally the gene set size was between 5 and 500 [38]. To compare the significant pathways obtained with iDEP and RUVs+DESeq2, two Venn diagrams were made, one with the upregulated pathways and the other with the downregulated pathways, using an online tool from Bioinformatics & Evolutionary Genomics [35].

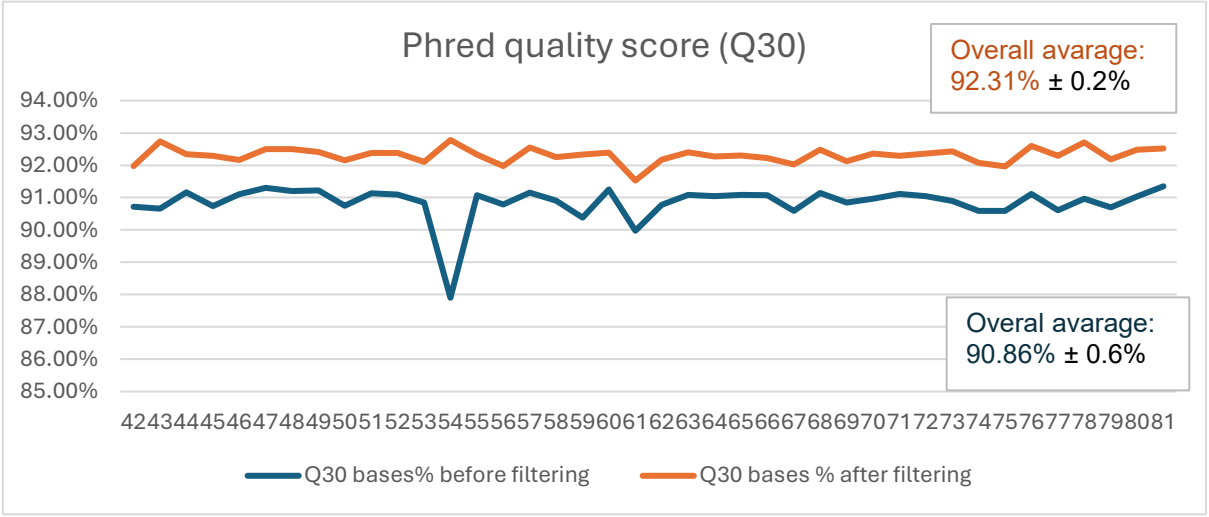
3. RESULTS

3.1 Quality check and trimming

Quality control of sequenced reads is a crucial step for ensuring solid data. Fastp and fastQC were used for this purpose. Two factors were taken in consideration to evaluate the overall quality of the reds after filtering: length and Phred quality score. The overall percentage of read for all samples that passed the trimming step was 97.69% with the smallest value of 95.20% (Fig. 3A). The Phred quality score, is important for determining the reliability of sequencing data. Q30 represents an accuracy of 99.9% [39]. Trimming with fastp improved Q30 from 90.86% to 92.31% (Fig. 3B); the overall scores obtained are indicative of high quality sequencing data suitable for downstream analyses.



A



B

FIGURE 3: Quality check results of the RNAseq reads pre- and post-trimming. A) Percentage of reads that passed the trimming step for each RNAseq library. B) Quality score (Q30) pre- and post-trimming for each RNAseq library. In each plot, on the Y-axis, the percentage mean for each library is plotted. Sample names are written on the X-axis; only the last two digits of their full names were kept. Quality score post-trimming is visualized in orange, while the quality average pre-trimming is shown in blue.

3.2 Differential gene expression analysis

The choice of parameters for mapping can impact the quality of alignment. A first run (i.e. #1) with a more strict set of parameters, particularly in regards to the "outFilterMismatchNoverLmax" was performed. Subsequently, a second run (i.e. #2) with more "relaxed" parameters was performed. A summary of the changes between the two runs is provided in Table 1.

Table 1: Summary of all STAR options considered. *In this table, in the first column, all STAR options that were used. There are two columns represented each one different setting. In the last column, the comparison between the two options is illustrated.*

Options	Run #1	Run #2	comparison
outFilterMultimapNmax:	20	10	This option limits the maximum number of multiple alignments a read can have to the desired value.
alignSJoverhangMin:	8	5	Specifies the minimum overhang for spliced alignments (i.e., junctions). Run #2 is less stringent
outFilterMismatchNmax:	999	10	Limits the maximum number of mismatches per read. Setting it to 999 essentially disables the limit on mismatches. Run #2 is more strict in forcing more high quality alignments
outFilterMismatchNoverLmax:	0.1	0.3	Restricts the maximum allowed ratio of mismatches over the read length to X (i.e., X% of the read length can be mismatches). Run #1 is more stringent.

A comparative table of the results of the mapping is reported in Table 2. Overall, run #2 was more permissive in accepting reads with mismatches and short reads with an overall mean lower value than run #1 (i.e. the average uniquely mapped reads were 62.27% ± 2% in contraposition with 62.82% ± 2% in run #1 and #2, respectively).

Table 2: Comparison Results of the Two Mapping Runs. *The two mapping setting results were summarized. Numbers represent the average percentages across samples ± SD.*

Mapping statistics	Uniquely mapped reads %		% of reads mapped to multiple loci		% of reads unmapped: too many mismatches		% of reads unmapped: too short		% of reads unmapped: other	
Option run	#1	#2	#1	#2	#1	#2	#1	#2	#1	#2
Mean	62.27% ± 2%	62.82% ± 2%	9.44% ± 1%	9.23% ± 1%	0.04% ±0%	0.00% ±0%	28.02 %±2%	27.46± %2%	0.07% ±0%	0.07% ±0%

The PCA results obtained using iDEP (Fig. 4A) showed that PC1 captured 9% of the total variance, distinctly separating control and treated groups. The second principal component (PC2) explained 5% of the variance. On the Y axis, the separation was less sharp; it could explain other sources of variation like biological or technical factors.
In figure (4B), PCA plot showed PC1 that explains 6.98% of the variance and PC2 4.61%. Like the result obtained with iDEP, PC1 separated the control and the treated groups, confirming the primary variation was due to the exposure to the heat wave.

The comparison of PCA results from iDEP and RUVSeq demonstrated that both normalizations separated the control and the treated groups. RUVSeq shows a slightly lower variance explained by both PC1 and PC2. Technical biases were minimized. However, separation remained consistent in both methods, confirming the reliability of the observed treatment effects.

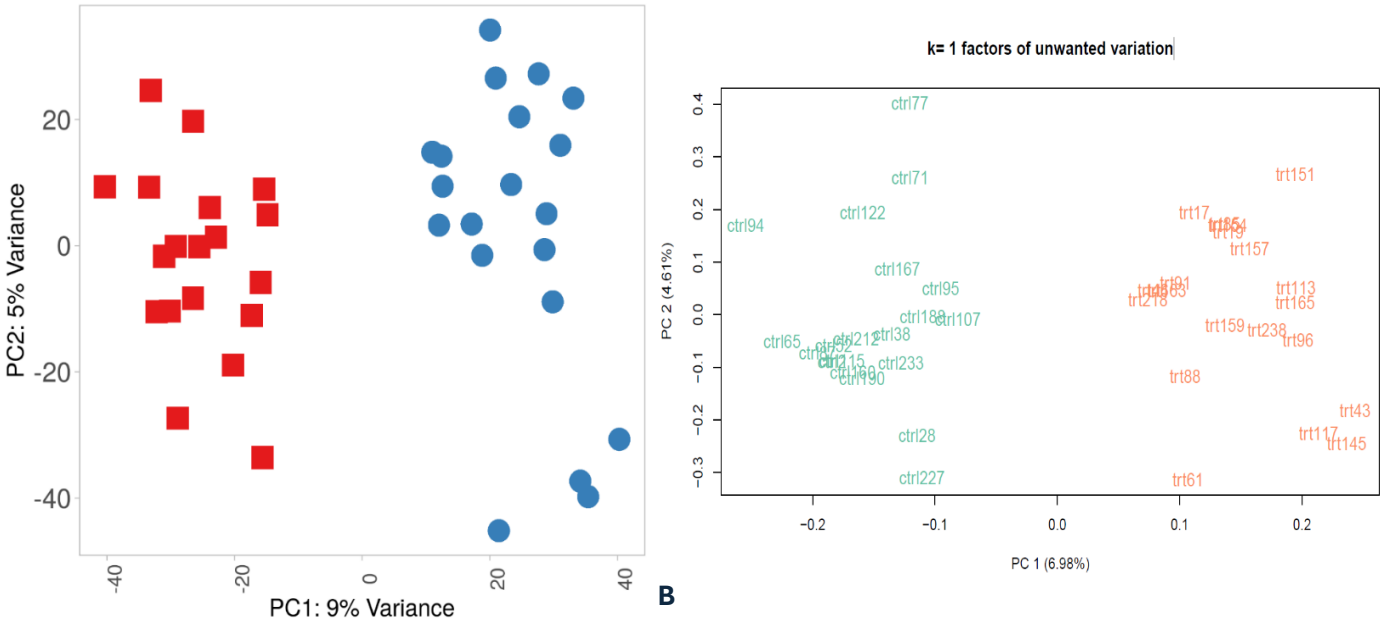
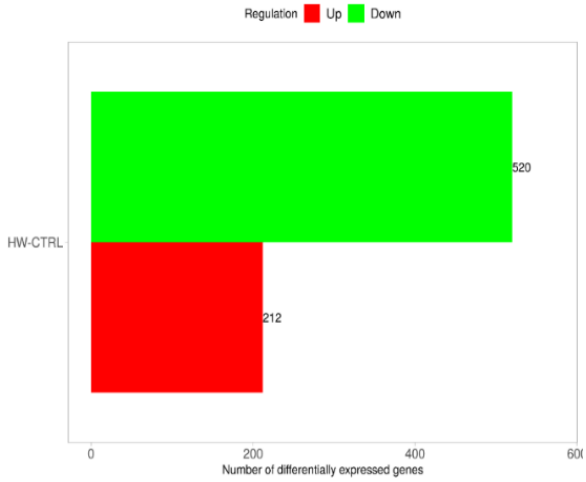


***FIGURE 4*: Comparison of Principal Component Analysis (PCA) Between iDEP and RUVs Normalization Methods.** *PCA on the X axis explains the main variance between samples (PC1). The Y axis explains secondary variations (PC2). In A) the red squares represented the control samples, and the blue circles represented the treated sample exposed to heat stress. A shows the result derived from iDEP, while B is derived from RUVs normalization with k = 1. In figure B, the names of the samples are plotted instead of the shapes like in the iDEP PCA.*

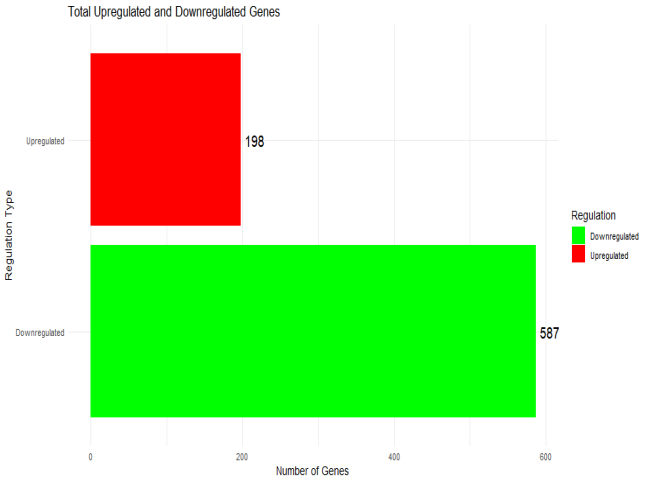
Following differential expression analysis, results were visualized by means of bar plots (Fig. 5A). The results showed that 520 upregulated and 212 downregulated were found with iDEP. On the other hand, with the DESeq2 method 198 upregulated genes and 587 downregulated genes were obtained (Fig. 5B).

Additionally, the log2 fold change distribution of the differentially expressed genes was plotted for both methods (Fig. 5C, D). For both iDEP and DESeq2, the log fold change distribution was concentrated in the intervals from -2 to -1 and from 1 to 2.

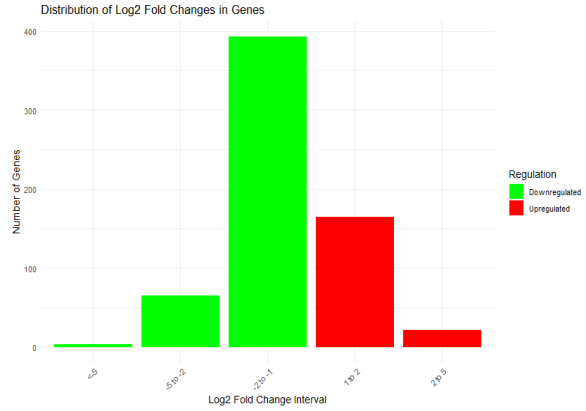
A) BAR PLOT IDEP



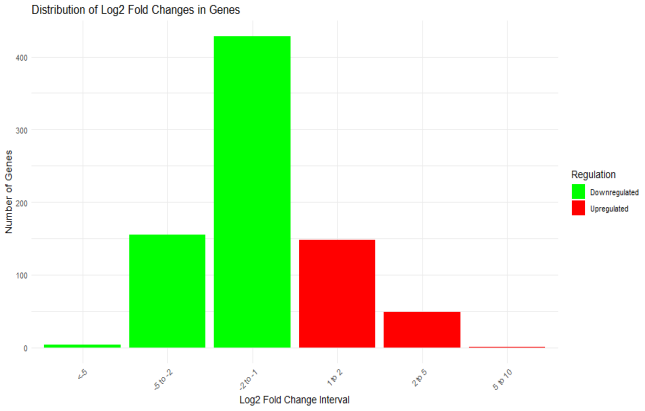
B) BAR PLOT RUVS



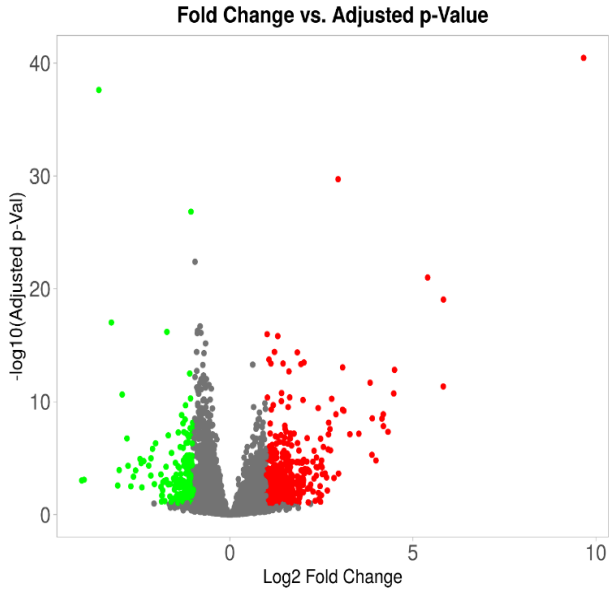
C) DISTRIBUTION OF LOGFC IDEP



D) DISTRIBUTION OF LOGFC RUVS



E) VULCANO PLOT IDEP



F) VULCANO PLOT RUVS

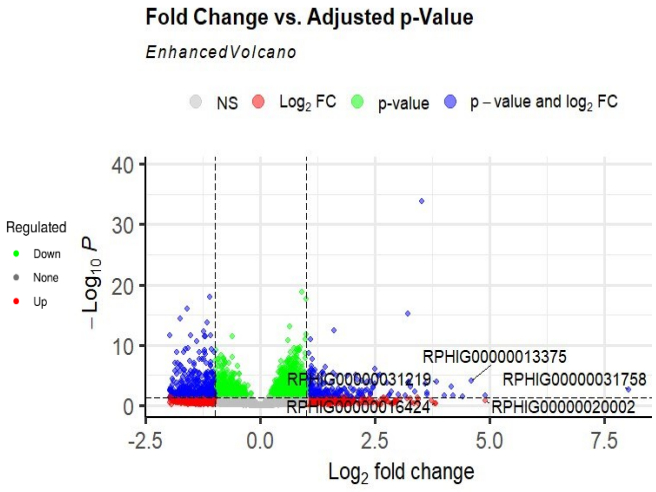


***FIGURE 5: Differentially genes expression plots**: In plots A and B, the green bars represent the number of downregulated genes, while the red bars reflected the upregulated genes. In figures C and D, there is the distribution of log2 fold change among the differential expressed genes grouped in intervals (X axis). In green, it was possible to find downregulated genes (negative fold change); in red, upregulated genes are displayed. On the Y axis, the numbers of genes are plotted.*

E plot: in grey all the genes that didn't pass the threshold FDR value (≤0.05) and fold change (> |1|) are shown. Red and gree dots refer to significantly upregulated and downregulated genes, respectively. In the Y axis is plot -log10(adjusted p-value); it is -log to give a better graphical visualization because all p-values were near 0, so higher in the graphic we can find more statistically significant genes because they have smaller p-adjusted values. On the X-axis, the Log2 fold change is displayed.

F plot: In green, there are presented genes with adequate p-values, but the fold change was out of the setting -1 or +1, so they were not considering differentially expressed genes. In red, it is possible to find genes that pass the fold change setting, but they were not statistically significant. In grey, there are located genes outside the significant region of the plot. The blue represents the significant genes, that pass all the setting (fold change -1, +1 and p-value <=0.05). On the left, there were downregulated genes and, on the right, the up-regulated genes.

A comparative Venn diagram was utilized to evaluate the overlapping genes from iDEP and DESeq2. The diagram revealed that 219 genes were exclusively identified by iDEP, 355 by DESeq2, and an overlap of 430 genes was observed between the two approaches (Fig. 6).

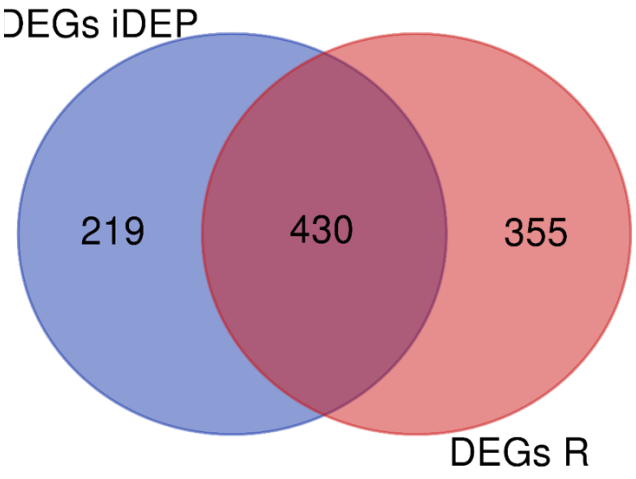


***FIGURE 6: Venn plot for differential expressed genes**: in the blue area there are present unique differential expressed genes found with iDEP; on the red zone are show the unique DEGs resulted from DEGSeq2 library. In the overlap between the blue and red regions, there are the number of genes common between the two different methods.*

The list of differentially expressed genes in common between the two methods is depicted in Table 3 and 4. Among the downregulated genes, it was found to be significant: "Tubulin alpha-1A chain" (RPHIG00000006377) with a fold change of -5.83, and "Serine/threonine-protein kinase/endoribonuclease IRE2" (RPHIG00000016986) with a fold change of -5.83. For the upregulated genes in Table 4 emerged "Heat shock protein beta-1" (RPHIG00000023892) with a fold change of 3.24, and "Collagen alpha-1(XIV) chain" (RPHIG00000024465) with a fold change of 2.94. In addition, there was a consistent number of hypothetical proteins (with no annotation) among the differentially expressed genes.

Table 3: Common genes between iDEP and DEGSeq2 library. *In table 3 is possible to find the list of the first 10 downregulated genes in common between iDEP and RUVSeq+DESEq2 based on Log fold change and False discovery rate (FDR).*

More significant common downregulated genes base on fold change:			
Genes	**Products**	**LogFC**	**FDR**
RPHIG00000014317	hypothetical protein	-9.66	3.46E-41
RPHIG00000032224	hypothetical protein	-5.40	9.98E-22
RPHIG00000006377	Tubulin alpha-1A chain"; uniprot_id "P6837	-5.84	8.97E-20
RPHIG00000004026	hypothetical protein	-4.50	1.51E-13
RPHIG00000016986	Serine/threonine-protein kinase/endoribonuclease IRE2"; uniprot_id "Q76MJ5	-5.83	4.46E-12
RPHIG00000001975	hypothetical protein	-4.48	1.88E-11
RPHIG00000005278	Tetratricopeptide repeat protein 25"; uniprot_id "Q7ZU45	-4.19	1.30E-09
RPHIG00000020128	Serine/threonine-protein kinase SCH9"; uniprot_id "P11792	-4.15	3.18E-09
RPHIG00000014872	Tubulin beta-4B chain"; uniprot_id "P68372	-4.19	1.46E-08
RPHIG00000001404	hypothetical protein	-4.32	4.66E-08

Table 4: Common genes between iDEP and DEGSeq2 library. *In table 4 is possible to find the list of the first 10 upregulated genes in common between iDEP and RUVSeq+DESEq2 based on Log fold change.*

More significant common upregulated genes base on fold change:			
Genes	**Products:**	**LogFC**	**FDR**
RPHIG00000031843	Protein lethal(2)essential for life"; uniprot_id "P82147	3.58	2.40E-38
RPHIG00000023892	Heat shock protein beta-1; uniprot_id "Q00649	3.24	9.53E-18
RPHIG00000024465	Collagen alpha-1(XIV) chain"; uniprot_id "Q80X19"	2.94	2.31E-11
RPHIG00000009669	Eukaryotic initiation factor 4A; uniprot_id "P27639	2.81	1.79E-07
RPHIG00000010612	hypothetical protein	2.45	1.19E-05
RPHIG00000014071	hypothetical protein	2.78	4.79E-05
RPHIG00000023697	hypothetical protein	3.02	1.15E-04
RPHIG00000022744	hypothetical protein	2.58	1.23E-04
RPHIG00000004538	hypothetical protein	4.05	9.49E-04
RPHIG00000027211	hypothetical protein	2.70	3.15E-03

A total of 79 pathways were identified using iDEP and 22 using Cluster Profile, both utilizing Gene Set Enrichment Analysis (GSEA). Among the downregulated pathways, 26 were identified by iDEP (Table 5) and 10 by Cluster Profile (Table 6), with no overlap between the two methods. In terms of upregulated pathways, 19 were shared, with 53 unique to iDEP (Table 7) and 3 unique to Cluster Profile (Table 8). The shared and unique pathways are visualized in Figure 7 through Venn diagrams.

Table 5: Downregulated iDEP pathways. *Downregulated pathways identified using iDEP, ranked by False Discovery Rate (FDR).*

Downregulated pathways iDEP			
ID	**Description**	**ES**	**FDR**
GO:0035082	axoneme assembly	-0.57	1.50E-04
GO:0001539	cilium or flagellum-dependent cell motility	-0.54	2.10E-04
GO:0060285	cilium-dependent cell motility	-0.54	2.10E-04
GO:0007018	microtubule-based movement	-0.41	2.10E-04
GO:0003341	cilium movement	-0.49	3.10E-04
GO:0001578	microtubule bundle formation	-0.50	1.30E-03
GO:0044782	cilium organization	-0.39	2.90E-03
GO:0060271	cilium assembly	-0.39	5.00E-03
GO:0030031	cell projection assembly	-0.36	6.40E-03
GO:0060294	cilium movement involved in cell motility	-0.52	8.40E-03
HSA04540	Gap junction	-0.60	9.00E-03
GO:0030317	flagellated sperm motility	-0.51	9.20E-03
GO:0097722	sperm motility	-0.51	9.20E-03
GO:0120031	plasma membrane bounded cell projection assembly	-0.36	9.30E-03
GO:0001964	startle response	-0.79	2.30E-02
GO:0019934	cGMP-mediated signaling	-0.79	2.50E-02
GO:2000051	negative regulation of non-canonical Wnt signaling pathway	-0.59	2.60E-02
GO:0070286	axonemal dynein complex assembly	-0.59	3.20E-02
GO:0016052	carbohydrate catabolic process	-0.45	3.30E-02
GO:0120322	lipid modification by small protein conjugation	-0.59	3.40E-02
GO:0120323	lipid ubiquitination	-0.59	3.40E-02
GO:0140042	lipid droplet formation	-0.59	3.60E-02
GO:0019932	second-messenger-mediated signaling	-0.42	3.60E-02
GO:2000050	regulation of non-canonical Wnt signaling pathway	-0.54	3.70E-02
GO:0090083	regulation of inclusion body assembly	-0.69	4.20E-02

GO:0043500	muscle adaptation	-0.55	4.40E-02

Table 6: Downregulated Cluster Profile pathways. *Downregulated pathways identified using Cluster profile, ranked by False Discovery Rate (FDR).*

Downregulated pathway Cluster profile R			
ID	**Description**	**ES**	**FDR**
GO:0090596	sensory organ morphogenesis	-0.46	1.92E-04
GO:0048592	eye morphogenesis	-0.47	1.02E-03
GO:0046348	amino sugar catabolic process	-0.78	2.76E-03
GO:0048593	camera-type eye morphogenesis	-0.48	5.61E-03
GO:0060042	retina morphogenesis in camera-type eye	-0.54	1.33E-02
hsa04974	Protein digestion and absorption	-0.52	1.77E-02
GO:0006040	amino sugar metabolic process	-0.69	2.63E-02
GO:0060041	retina development in camera-type eye	-0.46	3.89E-02
GO:1901071	glucosamine-containing compound metabolic process	-0.78	4.03E-02
GO:0006577	amino-acid betaine metabolic process	-0.63	4.38E-02

Table 7: Upregulated iDEP pathways. *Upregulated pathways identified using iDEP, ranked by False Discovery Rate (FDR).*

Upregulated pathway iDEP			
ID	**Description**	**ES**	**FDR**
GO:0022613	ribonucleoprotein complex biogenesis	0.52	1.40E-31
GO:0042254	ribosome biogenesis	0.59	1.40E-31
GO:0006364	rRNA processing	0.60	2.50E-24
GO:0034660	ncRNA metabolic process	0.46	1.00E-23
GO:0034470	ncRNA processing	0.50	5.10E-23
GO:0016072	rRNA metabolic process	0.55	1.60E-21
HSA03010	Ribosome	0.56	1.70E-11
GO:0140053	mitochondrial gene expression	0.54	9.60E-09
GO:0032543	mitochondrial translation	0.56	1.50E-08
GO:0042274	ribosomal small subunit biogenesis	0.57	2.80E-08

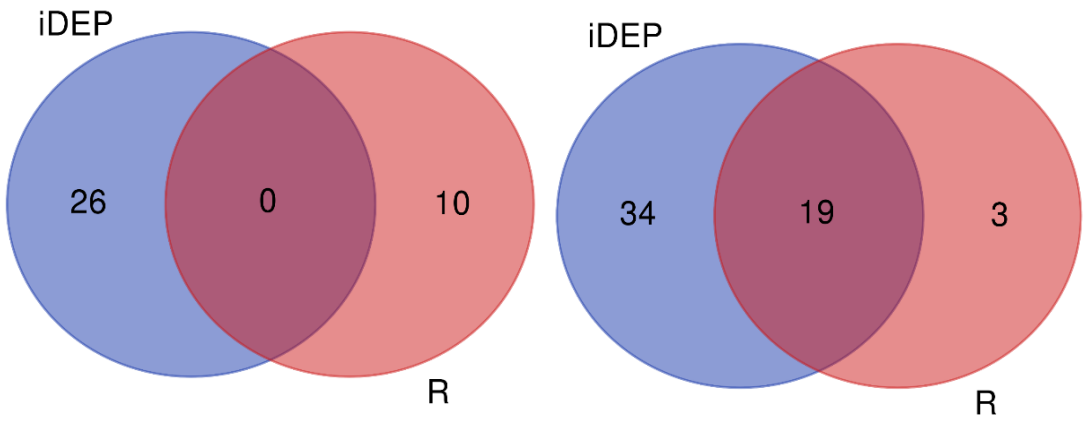
GO:0002181	cytoplasmic translation	0.48	6.40E-08
GO:0042273	ribosomal large subunit biogenesis	0.64	7.40E-08
GO:0030490	maturation of SSU-rRNA	0.61	1.80E-06
GO:0006399	tRNA metabolic process	0.45	3.60E-05
GO:0000462	maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA	0.60	1.20E-04
HSA03008	Ribosome biogenesis in eukaryotes	0.53	1.60E-04
GO:0042255	ribosome assembly	0.56	7.10E-04
GO:0022618	ribonucleoprotein complex assembly	0.38	9.40E-04
GO:0000460	maturation of 5.8S rRNA	0.65	1.30E-03
GO:0000469	cleavage involved in rRNA processing	0.68	1.30E-03
GO:1902570	protein localization to nucleolus	0.71	3.90E-03
GO:0071826	ribonucleoprotein complex subunit organization	0.36	5.00E-03
GO:0000466	maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA	0.73	5.00E-03
HSA00970	Aminoacyl-tRNA biosynthesis	0.57	6.60E-03
GO:0006413	translational initiation	0.45	9.00E-03
HSA04141	Protein processing in endoplasmic reticulum	0.40	9.20E-03
GO:0007005	mitochondrion organization	0.28	1.60E-02
GO:0050821	protein stabilization	0.36	1.60E-02
GO:0043038	amino acid activation	0.55	1.60E-02
GO:0043039	tRNA aminoacylation	0.55	1.60E-02
GO:0006418	tRNA aminoacylation for protein translation	0.56	1.60E-02
GO:0000478	endonucleolytic cleavage involved in rRNA processing	0.70	1.60E-02
GO:0000479	endonucleolytic cleavage of tricistronic rRNA transcript (SSU-rRNA	0.70	1.60E-02
GO:0065002	intracellular protein transmembrane transport	0.55	1.80E-02
GO:0006119	oxidative phosphorylation	0.44	2.10E-02
GO:0000966	RNA 5'-end processing	0.72	2.10E-02
GO:0031647	regulation of protein stability	0.29	2.90E-02
GO:0000027	ribosomal large subunit assembly	0.66	2.90E-02
GO:0071806	protein transmembrane transport	0.53	3.30E-02

GO:0006414	translational elongation	0.47	3.40E-02
GO:0098781	ncRNA transcription	0.54	3.60E-02
GO:1904749	regulation of protein localization to nucleolus	0.76	3.90E-02
GO:0070198	protein localization to chromosome	0.60	4.20E-02
GO:0010608	post-transcriptional regulation of gene expression	0.27	4.40E-02
GO:0071480	cellular response to gamma radiation	0.69	4.40E-02
GO:0042439	ethanolamine-containing compound metabolic process	0.79	4.40E-02
GO:0031123	RNA 3'-end processing	0.42	4.80E-02
GO:0000470	maturation of LSU-rRNA	0.64	4.80E-02
GO:0006589	octopamine biosynthetic process	0.83	4.80E-02
GO:0042415	norepinephrine metabolic process	0.83	4.80E-02
GO:0042421	norepinephrine biosynthetic process	0.83	4.80E-02
GO:0046333	octopamine metabolic process	0.83	4.80E-02
GO:1901522	positive regulation of transcription from RNA polymerase II promoter involved in cellular response to chemical stimulus	0.87	4.80E-02

Table 8: Upregulated Cluster Profile pathways. *Upregulated pathways identified using Cluster profile, ranked by False Discovery Rate (FDR).*

Upreguated patwhay Cluster Profile R			
ID	**Description**	**ES**	**FDR**
GO:0042254	ribosome biogenesis	0.58	7.68E-27
GO:0022613	ribonucleoprotein complex biogenesis	0.49	6.63E-24
GO:0016072	rRNA metabolic process	0.56	1.13E-20
GO:0006364	rRNA processing	0.57	4.65E-19
GO:0034470	ncRNA processing	0.46	9.17E-16
GO:0140053	mitochondrial gene expression	0.54	5.19E-07
GO:0032543	mitochondrial translation	0.55	1.04E-06
GO:0042274	ribosomal small subunit biogenesis	0.57	1.36E-06
GO:0042273	ribosomal large subunit biogenesis	0.61	3.26E-06
GO:0030490	maturation of SSU-rRNA	0.59	4.19E-05
hsa03010	Ribosome	0.48	4.44E-05

hsa03008	Ribosome biogenesis in eukaryotes	0.55	3.51E-04
GO:0006399	tRNA metabolic process	0.43	9.31E-04
GO:0002181	cytoplasmic translation	0.42	1.64E-03
GO:0000462	maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	0.58	4.05E-03
GO:0042255	ribosome assembly	0.55	5.85E-03
GO:0000466	maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	0.68	1.40E-02
GO:0000460	maturation of 5.8S rRNA	0.62	1.61E-02
GO:1902570	protein localization to nucleolus	0.68	2.05E-02
GO:0000469	cleavage involved in rRNA processing	0.63	2.20E-02
hsa05168	Herpes simplex virus 1 infection	0.33	3.54E-02
GO:0098781	ncRNA transcription	0.56	3.89E-02



A *B*

FIGURE 7: Venn plots for up and down regulated pathways between iDEP and Cluster profile: *A and B were Venn plots. The blue area represents unique pathways found with iDEP; in red, unique pathways found with Cluster profile, in the canter, the common pathways are displayed. In A, there are the downregulated pathways; in B, there are the upregulated pathways.*

4. DISCUSSION

4.1 Comparative analysis

Two different approaches were used: iDEP a Graphial User Interface (GUI) tool which internally relies on DESeq2 for differential expression analysis, and RStudio software with the libraries RUVSeq for data normalisation, DESeq2 for differential gene expression and clusterProfile for pathways identification. The DEGs result sowed some differences among the two approaches: in fact, the number of differential expressed genes was higher using RUVSeq+DESeq2 (785) compared to iDEP (649). The higher number of DEGs with RUVSeq+DESeq2 was probably due to the normalization method applied. In the “classical” DESeq2 method, the raw counts are normalised via scaling them by the median of ratios method which is used to calculate the size factor (i.e. a scaling factor that accounts for both differences in sequencing depth and RNA composition) for each sample. This scaling factor is used while performing all the steps of the differential expression analysis. Together with the “classic” DESeq2 normalisation through the median of ratios, the RUVSeq+DESeq2 approach was characterised by the identification, via the RUVs function, of unwanted factors of variations based on replicate samples (e.g., batch, library preparation, and other nuisance effects) which were then added to the experimental design matrix used by DESeq2 during all steps of differential expression analysis. Despite the differences in normalisation approaches, the was a consistent proportion of DEGs shared among the two approaches (430) and the distribution of the logFC was similar between methods. Among the unique DEGs identified with iDEP and RUVSeq, a significant portion of them were involved in extracellular compartment and ion-related process. Although both iDEP and RUVSeq+DESeq2 identified similar functions, most of the genes were different. In the pathways analysis, it was possible to find the main differences, especially in the downregulated pathways in which there were no common pathways. RUVSeq+DESeq2 was stricter since the number of pathways were 32 compared to 79 pathways found with iDEP. 19 upregulated pathways were shared among iDEP and RUVSeq+DESeq2. In conclusion, while the two approaches showed some similarities, they were also characterised by distinct results, offering different perspectives most likely associated to the complexity of RNA-seq data.

4.2 Pathways analysis

4.2.1 Cilium motility and structure downregulation:

Among downregulated pathways found using iDEP, there were ten pathways related to cilium structure and motility. Cilia are microtubule-based organelles, extending from the plasma membrane and coupled to the cytoskeleton [40]. in bivalves, cilia are found on a variety of tissues, such as gill, mantle or stomach, with important functions in, for instance, filtration, respiration, and pseudofeces expulsion [41]. The digestive gland of bivalves is a complex organ composed of digestive and connective tissues. The digestive tissue is comprised of stomach and intestinal epithelial cells, ciliated and non-ciliated cells are present [42]. In their article, Wang et al. (2023) used Hematoxylin-Eosin (HE) staining assays to examine oyster (i.e. *Crassostrea gigas* and *C. angulata*) gill tissues subjected to heat stress at 37°C for 12 hours. Histological analysis on tissue showed a change in cilia, specifically the number of impaired cilia increased under heat stress, particularly in *C. gigas.* Authors of this articles didn't explain directly why cilia in histological exam resulted damaged, they focus more in cytoskeleton remodelling due to differential phosphorylation of key cytoskeletal proteins, including alterations in actin dynamics and microtubule stability [43]. These modifications may indirectly affect cilia integrity, as the cytoskeleton supports cilia. The analysis was conducted on different tissues, and heat exposure was shorter than the condition studied in this thesis, however, it would be interesting to study this process better since 6 pathways related to cilium were found using iDEP: GO:0060271 - Cilium assembly,GO:0044782 - Cilium organization, GO:0060285 - Cilium-dependent cell motility, GO:0003341 - Cilium movement , GO:0060294 - Cilium movement involved in cell motility; and GO:0001539 - Cilium or flagellum-dependent cell motility. Another article revealed that gill tissues in *R. philippinarum* after heat stress at 30° showed damage to the cilia following histological analysis; in particular, researchers highlighted that some of the branchial filaments were uncovered because of desquamation of ciliated epithelial cells and calceiform cells. Exfoliation could be related to a decrease of Na^+, K^+ activated ATPase or a decline in Na^+ and Cl^- concentration [44]. The Na,K-ATPase establishes the electrochemical gradient of cells by driving an active exchange of Na^+ and K^+ ions while consuming ATP, in addition to ion transport, the Na,K-ATPase also acts as a cell adhesion molecule in polarized epithelial cells, indirectly regulating permeability, controlling cellular volume and motility [45].

Cilia damage was also reported in another article, in which a loss of cilia in wild *M. coruscus* was observed under thermal stress [46]. Desquamation of ciliated epithelial cells after heat stress on *R. philippinarum* was also reported by Menike [47].

4.2.2 Sperm motility and structure downregulation:

Using iDEP, pathways related to sperm motility and structure were found, specifically; **GO:0097722** (sperm motility), **GO:0030317** (flagellated sperm motility), **GO:0035082** (axoneme assembly), and **GO:0070286** (axonemal dynein complex assembly). Since the digestive gland is surrendered by reproductive tissues, it is not surprising that significant hits associated with reproductive functions were found analysing the RNA-seq data of the digestive gland of *R. philippinarum*.

Axonemal is the structure of flagella made by microtubules and associated proteins, which provides the motor apparatus for sperm movement [48]. In a published article, Nash et al. [49] exposed *Crassostrea virginica* to control (24°C) and elevated temperatures (28°C and 32°C) for one week. They found at histological analysis that sperm production experienced a significant decrease around 26% when compared with controls. The results of the study demonstrate that short-term exposure to medium and high temperatures (28 and 32°C) attenuates testicular functions by altering body fluid conditions and increasing oxidative stress and apoptosis in testicular tissue of American oysters. More in detail hight temperature increased extrapallial fluid (EPF) pH, that interfere with sperm development [49]. EPF is situate between the inner surface of the shell and the mantle and it is involved in shell-forming system [50].

Heat stress caused an increase of oxidative stress, associated with elevated ROS and ROS level in oyster testicular tissue, resulting in cellular damage. In Nash et al. [53] they measured the levels of nitrotyrosine protein (NTP) as a biomarker for reactive nitrogen species (RNS) and dinitrophenyl protein (DNP) as a biomarker for reactive oxygen species (ROS) using immunohistochemistry [49] [51].

Nash et al. [49] study also showed increased apoptosis rate in spermatogenic cells under hight temperature. For understand that they performed a TUNEL assay (Terminal Deoxynucleotidyl Transferase (TdT) dUTP Nick-End Labeling), this method labels is based on the ligation of dUTP to the 3′-OH phosphate ends including blunt-ended and 5′-recessed DNA fragments, thus measures definite endpoints, which is referred to as DNA damage [57].

Another article, reporter other examples of gamete downregulation; one of these was the decrease of speed in gametogenesis after high temperature exposure in *Crassostrea gigas*: [53].

4.2.3 Mitochondrial gene expression

Mitochondria play a key role in stress responses such as heat by producing ROS and other signalling molecules that induce the cellular response [54]. Mitochondrial gene expression (GO:0140053) pathway is defined as the process by which a mitochondrial gene's sequence is converted into a mature gene product or products [55]. This pathway was up regulated, according to data provided by iDEP. In the study by Georgoulis et al [56], mRNA expression of mitochondria was examined specifically, in ND2 and COX1 in *M. galloprovincialis*. The heat stress was tasted from 26°C to 28°C in heat-hardened and non-hardened mussel. Elevated temperatures caused increased *COX1* and *ND2* mRNA expression in hardened and non-hardened mussels from the 1st day of exposure compared to the baseline levels observed in the control mussels maintained at 18 °C.

Heat-hardened mussels demonstrated a more robust upregulation of these mitochondrial genes, as well as higher electron transport system (ETS) activity, compared to non-hardened mussels, contributed to better mitochondrial integrity and organization under thermal stress. non-hardened mussels showed a lesser increase in mRNA expression and ETS activity, and experienced higher ROS production, reduced ATP synthesis. Enhanced upregulation of mitochondrial pathways in heat-hardened mussels provides a protective mechanism against thermal stress.

In another article, Abele et al. [57] studied the effect on mitochondria under heat exposure. In their work, the bivalve was *Mya arenaria,* control 5-15°C heat stress 20-25°C for 7 days and another group for 2 days, the mitochondria were isolated from mantle tissue. Respiratory control ratio (RCR) was calculated by dividing state 3 by state 4 respiration. State 3 refers to the condition where mitochondria exhibit maximal oxygen consumption and ATP synthesis, while state 4 occur when ATP synthesis has stopped and oxygen consumption derives exclusively from proton leak. [58]

RCR decreased in bivalves exposed to heat condition and this decrease corresponds to a reduced rate of ATP production. ROS formations were measured through the oxidation of DHR (dihydrorhodamine), showing an increase in ROS production under heat stress. State 4 respiration increased compared to state 3 during temperature rising resulting in increased

proton leakage across the inner mitochondrial membrane and an increase in mitochondrial ROS production. This lead to more significant exposure of the cell to hazardous respiratory products under high temperature stress [57]. Another pathway resulted upregulated: Oxidative phosphorylation; mitochondrial electron transport chain utilized electron transfer reaction to generate ATP through oxidative phosphorylation. A consequence of electron transfer is the production of reactive oxygen species [59].Oxidative phosphorylation pathways may be linked to proton leakage across inner mitochondrial membrane, which decrease ATP efficiency and increase ROS production.

In conclusion, enhanced mitochondrial response can provide a protective mechanism against heat stress, but ROS excess and reduced ATP efficiency production contribute to mitochondrial disfunction during stress.

4.2.4 Non-canonical Wnt signalling

Non-canonical Wnt signalling pathways is involved in myogenesis, actin remodelling, cell shape and cytoskeleton regulation. Non-canonical Wnt pathways is also involved in the Wnt/calcium pathway which regulates calcium level in the cells. heat-induced genes affecting the Wnt pathway typically inhibit Wnt signalling. Non-canonical wnt pathways is primarily mediated by the Wnt4, Wnt5a, Wnt5b, Wnt6, Wnt7a, and Wnt11 ligands [60]

Canonical and non-canonical Wnt pathways are not mutually exclusive, they can be activated simultaneously. In the article Bai et al. [61] was studied expression profiles of the Wnt genes at different developmental stage, in adult tissues, during siphon regeneration, in four different shell colour strains. The non canonical Wnt pathway is up-regulated after siphon amputation and play a role in regulating cell fate during regeneration. The pathway identified with iDEP (GO:2000051), negative regulation of the non-canonical Wnt signaling pathway, is downregulated, which suggests that the reduction in signalling leads to increased activity of the pathway. This may play a role in cell regeneration, possibly during heat stress. In particularly Wnt4 genes is reported to be involved in cell proliferation and regeneration [61]

In the article Risha et al. [67] was examined the effect on C2C12 myoblast cells to heat stress including cold stress (35°C), control (37°C), mild heat stress (39°C), and severe heat stress (41°C) using real-rime PCR. They discovered that Wnt5 and Wnt11 involved in Wnt/PCP and Wnt/Calcium signalling, were upregulated under severe heat stress

suggesting that those genes may contribute to cytoskeletal reorganization and calcium homeostasis, thus helping cells adapting to stressful thermal conditions [62].

5. CONCLUSION

iDEP and RUVSeq+DESeq2, use different normalization methods, the first raw counts are normalized via scaling them by the median of rations method, the second uses replicated samples to remove unwanted variations. These two different normalization approaches led to a considerable number of common genes despite differences in normalization.

Among the unique DEGs found by the two methods, there were different genes but with similar functions. The main differences were found in pathways analysis, in which the number of common pathways was low. The RUVseq+DESeq2 method detected fewer pathways compared to iDEP. Both methods used GSEA computation. One difference between the two methods could be that while for clusterProfile, it was possible to set the exponent equal to 0, it is not possible to set this on iDEP, so it is not possible to know the weight applied to gene ranking.

Pathways found significant after *R. philippinarum* exposure to heat were involved in sperm motility and structure, cilium damage, enhanced mitochondria activity and non-canonical Wnt singling.

With iDEP, it was possible to find several pathways related to cilia, which may be directly damaged following heat stress, probably by disrupting ion transport and impairing cytoskeleton remodelling.

Sperm motility and structure were found downregulated with iDEP may be due to changes in body fluid condition, oxidative stress and apoptosis in spermatogenic cells. Oxidative stress was revealed also in mitochondrial enhanced activity in which ATP supply is necessary to survive during stress conditions but due to electron leakage during electron transport chain caused by heat leads to ROS production.

In conclusion, the analysis of heat stress responses in *R. philippinarum* revealed distinct pathways through two different methodologies, emphasizing the complexity of gene regulation under stress conditions.

6. REFERENCES

[1] National Aeronautics and Space Administration. (2024, February). *What is climate change?* NASA, from https://science.nasa.gov/climate-change/what-is-climate-change/

[2] IPCC, 2023: Summary for Policymakers. In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 1-34, doi: 10.59327/IPCC/AR6-9789291691647.001

[3] Gao, K., Zhang, Y. & Häder, DP. Individual and interactive effects of ocean acidification, global warming, and UV radiation on phytoplankton. *J Appl Phycol* 30, 743–759 (2018). https://doi.org/10.1007/s10811-017-1329-6

[4] International Union for Conservation of Nature (IUCN). (2017, November). *Ocean warming*. https://www.iucn.org/resources/issues-brief/ocean-warming

[5] Frölicher, T.L., Laufkötter, C. Emerging risks from marine heat waves. *Nat Commun* 9, 650 (2018). https://doi.org/10.1038/s41467-018-03163-6

[6] Garrabou, J., Gómez-Gras, D., Medrano, A., Cerrano, C., Ponti, M., Schlegel, R., Bensoussan, N., Turicchia, E., Sini, M., Gerovasileiou, V., Teixido, N., ... (2022). Marine heatwaves drive recurrent mass mortalities in the Mediterranean Sea. *Global Change Biology*, *28*(22), 6508–6528. https://doi.org/10.1111/gcb.16301

[7] Galimany, E., Lucas, A., Maynou, F., Solé, M., Pelejero, C., & Ramón, M. (2023). Experimental determination of differential seasonal response in seed of the Manila clam, *Ruditapes philippinarum,* in context of climate change. Aquaculture, 576, 739891. https://doi.org/10.1016/j.aquaculture.2023.739891

[8] Gerdol, M., Gomez-Chiarri, M., Castillo, M. G., Figueras, A., Moreira, R., Novoa, B., Fiorito, G., Pallavicini, A., Ponte, G., Roumbedakis, K., Venier, P., & Vasta, G. R. (2018). Immunity in Molluscs: Recognition and effector mechanisms, with a focus on Bivalvia. In E. L. Cooper (Ed.), *Advances in comparative immunology* (pp. 225-276). Springer International Publishing. https://doi.org/10.1007/978-3-319-76768-0_119]

[9] Banrie. (2012, October 1). Cultured aquatic species: Japanese carpet shell. *The Fish Site*. https://thefishsite.com/articles/cultured-aquatic-species-japanese-carpet-shell

[10] Velez, C., Galvão, P., Longo, R. *et al. Ruditapes philippinarum* and *Ruditapes decussatus* under Hg environmental contamination. *Environ Sci Pollut Res* 22, 11890–11904 (2015). https://doi.org/10.1007/s11356-015-4397-7

[11] Li, D., Nie, H., Jiang, K., Li, N., Huo, Z., & Yan, X. (2020). Molecular characterization and expression analysis of fibrinogen related protein (FREP) genes of Manila clam (*Ruditapes philippinarum*) after lipopolysaccharides challenge. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology, 228, 108672. https://doi.org/10.1016/j.cbpc.2019.108672

[12] Subramoniam, T. (2017). Sexual systems. In *Sexual biology and reproduction in crustaceans* (pp. 57-103). Elsevier. https://doi.org/10.1016/B978-0-12-809337-5.00003-4

[13] Food and Agriculture Organization of the United Nations. (2024). *Global aquaculture production quantity (1950 - 2022)*. FAO. Retrieved September 20, 2024", from https://www.fao.org/fishery/statistics-query/en/aquaculture/aquaculture_quantity

[14] Ponti, M., Castellini, A., Ragazzoni, A., Gamba, E., Ceccherelli, V. U., & Abbiati, M. (2017). Decline of the Manila clam stock in the northern Adriatic lagoons: A survey on ecological and socio-economic aspects. *Acta Adriatica, 58*(1), 89–104.

[15] Biswal, A., Srivastava, P.P., Paul, T. (2021). Effect of Climate Change on Endocrine Regulation of Fish Reproduction. In: Sundaray, J.K., Rather, M.A., Kumar, S., Agarwal, D. (eds) Recent updates in molecular Endocrinology and Reproductive Physiology of Fish. Springer, Singapore. https://doi.org/10.1007/978-981-15-8369-8_21

[16] Pantouw, C.F., Syahidah, D., Hastilestari, B.R. (2024). Ripples of Climate Change: Effects on Reproductive Dynamics of Aquatic Animals and Vegetation. In: Lestari, S., *et al.* Proceedings of the International Conference on Radioscience, Equatorial Atmospheric Science and Environment and Humanosphere Science. INCREASE 2023. Springer Proceedings in Physics, vol 305. Springer, Singapore. https://doi.org/10.1007/978-981-97-0740-9_65

[17] Byers JE (2020) Effects of climate change on parasites and disease in estuarine and nearshore environments. PLoS Biol 18(11): e3000743. https://doi.org/10.1371/journal.pbio.3000743

[18] Wan, Y., Kertesz, M., Spitale, R. C., Segal, E., & Chang, H. Y. (2011). Understanding the transcriptome through RNA structure. *Nature Reviews Genetics, 12*(9), 641-655. https://doi.org/10.1038/nrg3049

[19] Chen, J.-W., Shrestha, L., Green, G., Leier, A., & Marquez-Lago, T. T. (2023). The hitchhikers' guide to RNA sequencing and functional analysis. *Briefings in Bioinformatics, 24*(1), 1-17. https://doi.org/10.1093/bib/bbac529

[20] Van den Berge, K., Hembach, K. M., Soneson, C., Tiberi, S., Clement, L., Love, M. I., Patro, R., & Robinson, M. D. (2019). RNA sequencing data: Hitchhiker's guide to expression analysis. *Annual Review of Biomedical Data Science, 2*, 139-173. https://doi.org/10.1146/annurev-biodatasci-072018-021255

[21] Corchete, L.A., Rojas, E.A., Alonso-López, D. *et al.* Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. *Sci Rep* 10, 19737 (2020). https://doi.org/10.1038/s41598-020-76881-x

[22] National Center for Biotechnology Information (NCBI). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] Available from: https://www.ncbi.nlm.nih.gov/

[23] Peruzza L *et al.,* "Impaired reproduction, energy reserves and dysbiosis: The overlooked consequences of heatwaves in a bivalve mollusc", *Marine Pollution Bulletin*, 2023;193

[24] SRA Toolkit Development Team. (2024). *SRA Toolkit* [Computer software]. National Center for Biotechnology Information (NCBI). Available from: https://github.com/ncbi/sra-tools

[25] Cock, P.J.A., et al. (2010). "The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants." *Nucleic Acids Research*, 38(6), 1767–1771. DOI: https://doi.org/10.1093/nar/gkp1137

[26]https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[27] Shifu Chen, Yanqing Zhou, Yaru Chen, Jia Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics, Volume 34, Issue 17, September 2018, Pages i884–i890, https://doi.org/10.1093/bioinformatics/bty560

[28] Dobin, A., Gingeras, T.R. (2016). Optimizing RNA-Seq Mapping with STAR. In: Carugo, O., Eisenhaber, F. (eds) Data Mining Techniques for the Life Sciences. Methods in Molecular Biology, vol 1415. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-3572-7_13

[29] https://github.com/ncbi/sra-tools?tab=readme-ov-file#the-sra-toolkit

[30] Ge, S.X., Son, E.W. & Yao, R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics* 19, 534 (2018). https://doi.org/10.1186/s12859-018-2486-6

[31] Risso, D. (2022). RUVSeq: Remove Unwanted Variation from RNA-Seq Data. University of Padova. Retrieved from https://bioconductor.org/packages/release/bioc/vignettes/RUVSeq/inst/doc/RUVSeq.html.

[32] Love MI, Huber W, Anders S (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology*, 15, 550. doi:10.1186/s13059-014-0550-8.

[33] Risso, D. (2024). *EDASeq: Exploratory Data Analysis and Normalization for RNA-Seq*. Bioconductor version 3.19. Available at: https://github.com/drisso/EDASeq.

[34] Blighe K, Rana S, Lewis M (2024). *EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling*. R package version 1.22.0, https://github.com/kevinblighe/EnhancedVolcano.

[35] https://bioinformatics.psb.ugent.be/webtools/Venn/

[36] Subramanian, A., Tamayo, P., Mootha, V. K., ... Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences, 102(43), 15545-15550. https://doi.org/10.1073/pnas.0506580102

[37] Xu S, Hu E, Cai Y, Xie Z, Luo X, Zhan L, Tang W, Wang Q, Liu B, Wang R, Xie W, Wu T, Xie L, Yu G (2024). "Using clusterProfiler to characterize multiomics data." *Nature Protocols*. ISSN 1750-2799, doi:10.1038/s41596-024-01020-z, https://www.nature.com/articles/s41596-024-01020-z.

[38] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005).

Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550.

[39] Illumina, Inc. (2024). Quality scores in next-generation sequencing. Illumina. https://emea.illumina.com/science/technology/next-generation-sequencing/plan-experiments/quality-scores.html

[40] May-Simera, H.L., Kelley, M.W. Cilia, Wnt signaling, and the cytoskeleton. *Cilia* 1, 7 (2012). https://doi.org/10.1186/2046-2530-1-7

[41] Ertl, N.G., O'Connor, W.A., Wiegand, A.N. *et al.* Molecular analysis of the Sydney rock oyster (*Saccostrea glomerata*) CO_2 stress response. *Clim Chang Responses* 3, 6 (2016). https://doi.org/10.1186/s40665-016-0019-y

[42] Robledo Y, Cajaraville MP. Isolation and morphofunctional characterization of mussel digestive gland cells in vitro. Eur J Cell Biol. 1997 Apr;72(4):362-9. PMID: 9127736.

[43] Wang, C., Du, M., Jiang, Z., Cong, R., Wang, W., Zhang, G., & Li, L. (2023). Comparative proteomic and phosphoproteomic analysis reveals differential heat response mechanism in two congeneric oyster species. *Ecotoxicology and Environmental Safety*, *258*, 115197. https://doi.org/10.1016/j.ecoenv.2023.115197

[44] Menike, U., Lee, Y., Oh, C. *et al.* Oligo-microarray analysis and identification of stress-immune response genes from manila clam (*Ruditapes philippinarum*) exposure to heat and cold stresses. *Mol Biol Rep* 41, 6457–6473 (2014). https://doi.org/10.1007/s11033-014-3529-3

[45] Kryvenko, V., Vagin, O., Dada, L.A. *et al.* Maturation of the Na,K-ATPase in the Endoplasmic Reticulum in Health and Disease. *J Membrane Biol* 254, 447–457 (2021). https://doi.org/10.1007/s00232-021-00184-z

[46] Dong, Z., Li, H., Wang, Y., Lin, S., Guo, F., Zhao, J., Yao, R., Zhu, L., Wang, W., Buttino, I., Qi, P., & Guo, B. (2023). Transcriptome profiling reveals the strategy of thermal tolerance enhancement caused by heat-hardening in Mytilus coruscus. *Science of the Total Environment*, 903, 165785. https://doi.org/10.1016/j.scitotenv.2023.165785

[47] Menike, U., Lee, Y., Oh, C. *et al.* Oligo-microarray analysis and identification of stress-immune response genes from manila clam (*Ruditapes philippinarum*) exposure to heat

and cold stresses. *Mol Biol Rep* 41, 6457–6473 (2014). https://doi.org/10.1007/s11033-014-3529-3

[48] Mari S Lehti, Anu Sironen, Formation and function of sperm tail structures in association with sperm motility defects, *Biology of Reproduction*, Volume 97, Issue 4, October 2017, Pages 522–536, https://doi.org/10.1093/biolre/iox096

[49] Nash, S., & Rahman, M. S. (2019). Short-term heat stress impairs testicular functions in the American oyster, *Crassostrea virginica*: Molecular mechanisms and induction of oxidative stress and apoptosis in spermatogenic cells. *Molecular Reproduction and Development*, 86(11), 1289–1302. https://doi.org/10.1002/mrd.23268

[50] Wilbur, K. M., & Bernhardt, A. M. (1984). Effects of amino acids, magnesium, and molluscan extrapallial fluid on crystallization of calcium carbonate: In vitro experiments. *The Biological Bulletin, 166*(1), 251-259. University of Chicago Press

[51] Michael Zuykov, Emilien Pelletier, Serge Demers, Method for repeated extrapallial fluid extraction from bivalve molluscs, *Journal of Molluscan Studies*, Volume 76, Issue 4, November 2010, Pages 399–400, https://doi.org/10.1093/mollus/eyq029

[52] Lewis, S. E. M., Aitken, R. J., Conner, S. J., De Iuliis, G., Evenson, D. P., Henkel, R., Giwercman, A., & Gharagozloo, P. (2013). The impact of sperm DNA damage in assisted conception and beyond: Recent advances in diagnosis and treatment. *Reproductive BioMedicine Online, 27*(4), 325-337. https://doi.org/10.1016/j.rbmo.2013.06.014

[53] Chávez-Villalba, J., Pommier, J., Andriamiseza, J., Pouvreau, S., Barret, J., Cochard, J.-C., & Le Pennec, M. (2002). Broodstock conditioning of the oyster *Crassostrea gigas*: Origin and temperature effect. *Aquaculture, 214*(1–4), 115–130. https://doi.org/10.1016/S0044-8486(01)00898-5

[54] Georgoulis, I., Feidantsis, K., Giantsis, I. A., Kakale, A., Bock, C., Pörtner, H. O., Sokolova, I. M., & Michaelidis, B. (2021). Heat hardening enhances mitochondrial potential for respiration and oxidative defense capacity in the mantle of thermally stressed *Mytilus galloprovincialis*. Scientific Reports, 11(1), 17098. https://doi.org/10.1038/s41598-021-96617-9

[55] European Molecular Biology Laboratory - European Bioinformatics Institute. (2024). GO:0140053 - Term details. QuickGO. https://www.ebi.ac.uk/QuickGO/term/GO:0140053->controlalre sai giusto

[56] Georgoulis, I., Feidantsis, K., Giantsis, I.A. *et al.* Heat hardening enhances mitochondrial potential for respiration and oxidative defence capacity in the mantle of thermally stressed *Mytilus galloprovincialis*. *Sci Rep* 11, 17098 (2021). https://doi.org/10.1038/s41598-021-96617-9

[57] Abele, D., Heise, K., Pörtner, H. O., & Puntarulo, S. (2002). Temperature-dependence of mitochondrial function and production of reactive oxygen species in the intertidal mud clam Mya arenaria. Journal of Experimental Biology, 205(13), 1831–1841

[58] Korzeniewski B. 'Idealized' state 4 and state 3 in mitochondria vs. rest and work in skeletal muscle. PLoS One. 2015 Feb 3;10(2):e0117145. doi: 10.1371/journal.pone.0117145. PMID: 25647747; PMCID: PMC4412265.

[59] Nolfi-Donegan, D., Braganza, A., & Shiva, S. (2020). Mitochondrial electron transport chain: Oxidative phosphorylation, oxidant production, and methods of measurement. *Redox Biology*, 37, 101674. https://doi.org/10.1016/j.redox.2020.101674

[60] Qin, K., Yu, M., Fan, J., Wang, H., Zhao, P., Zhao, G., Zeng, W., Chen, C., Wang, Y., Wang, A., Schwartz, Z., Hong, J., Song, L., Wagstaff, W., Haydon, R. C., Luu, H. H., Ho, S. H., Strelzow, J., Reid, R. R., & He, T.-C. (2023). Canonical and noncanonical Wnt signaling: Mult ilayered mediators, signaling mechanisms and major signaling crosstalk.

[61] Bai, Y., Nie, H., Wang, Z., & Yan, X. (2020). Genome-wide identification and transcriptome-based expression profiling of Wnt gene family in *Ruditapes philippinarum.* Comparative Biochemistry and Physiology [1] - Part D, 35(100709).

[62] Risha MA, Ali A, Siengdee P, Trakooljul N, Haack F, Dannenberger D, Wimmers K, Ponsuksili S. 2021. Wnt signaling related transcripts and their relationship to energy metabolism in C2C12 myoblasts under temperature stress. *PeerJ* 9:e11625 https://doi.org/10.7717/peerj.11625

7. ACKNOWLEDGEMENTS

I would like to thank my family for support me during this master, especially my sisters for listen to my drama all days.

I would like to thank my friends in Belluno that never forgot about me, Alba, Giada, Alessia,Giulia, Elisa, Stefania.and Deborah

I thank also to Pierre and him family. Thanks to Alice e Alessio for dealing with crazy goose.

Thanks also to my friend in Padua that help me living the two best years of my life until now during this master, Vale, Elisa, Bea, Gigi, Marty, Nursah and Laura

I’m also grateful to my relator Peruzza for always helping me during this thesis preparation, always kind and precise in his explanation.