universität wien

# MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Antimicrobial peptides as potential mediators of immunity and symbiosis in *Loripes orbiculatus*"

verfasst von / submitted by

Alejandro Llanos-Lizcano

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2024/ Vienna, 2024

| | |
|---|---|
| Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet: | UA 066 830 |
| Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet: | Master's degree programme Molecular Microbiology, Microbial Ecology and Immunobiology |
| Betreut von / Supervisor: | Assoz. Prof. Dr. Jillian Petersen |

# *Table of Contents*

# Acknowledgments

I have to start thanking Ass. Prof. Dr. Jillian Petersen, since I took her practical course and the lecture on symbiosis, I was inspired and intrigued by the principles of how the lucinids symbiotic relationship works at the immune level. Allowing me to take that project with all the challenges ahead was an excellent experience for learning about my limitations and how I could sort them by searching for advice and collaboration from experienced people, all an adventure on how science works.

I also have to thank Prof. Dr. Ron Pinhasi and especially all the group members who gave me the opportunity to work as a lab technician, developing my skills in the wet lab, learning new protocols, and getting an insight into a field that before I never thought that I could work in it.

Both research groups have been successful and productive experiences that will help me on my way to becoming a scientist.

I must mention my family, who were always supporting me. Through them, I was able to reach my goals. For moments when I felt I was not enough for the challenge, they always believed in me. The same I must say about all the friends I made in these three years: Maria, Fer, Paolo, Michael, Isko, Ale, Vojta, Nico, and many, many more; with them, I could enjoy this time, cry, find support and in general experience great moments.

Finally, this would have never been possible without my sponsor, Colfuturo, who allowed me to come to this incredible city and take some steps to fulfill my dreams.

**Abstract**

A diverse range of marine invertebrates have evolved associations with intracellular, chemosynthetic bacteria that provide them with nutrition. The specificity of these interactions is remarkable, but so far, nothing is known about the mechanisms that underpin it. Burrowing clams from the family Lucinidae are a prime example of such a specific chemosynthetic symbiosis. The nutritional relationship develops primarily in the clam's gills, with the rest of the body not known to host symbionts. Antimicrobial peptides (AMPs) are known to play a role in mediating specificity in other intracellular symbioses, but these have not yet been investigated in lucinid clams. To close this knowledge gap, I de-novo assembled a reference transcriptome for the lucinid *Loripes orbiculatus* and developed a custom pipeline to identify AMPs. We used Trinity to assemble the pre-processed raw reads pooled from four different organs: Gills (with symbionts), mantle, foot, and visceral mass (all no symbionts). We used Transdecoder to predict open reading frames. AMPs were predicted based on homology (BLASTp against dbAMP database) or by physicochemical properties (AMPIR). 106,974 transcripts successfully assembled with Trinity were retained. 42,000 potential AMPs were predicted based on physicochemical properties, whereas 709 AMPs were identified by homology. Of these 709, the predominant AMP families based on expression were Buthinin, ovispirin and Lc-NK-lysin. Here I identify for the first time AMPs in the *Loripes orbiculatus* transcriptome. Organ-specific differential expression analysis of the transcripts encoding for AMPs showed a functional difference between organs. The genome mapping using a draft genome showed the need for more efforts to produce a reference genome and study different levels of heterozygosity in populations of *L. orbiculatus*.

**Zusammenfassung**

Eine Vielzahl von wirbellosen Meerestieren hat Verbindungen mit intrazellulären, chemosynthetischen Bakterien entwickelt, die sie mit Nahrung versorgen. Die Spezifität dieser Interaktionen ist bemerkenswert, jedoch sind bisher keine Mechanismen bekannt, die ihnen zugrunde liegen. Ein Paradebeispiel für eine solche spezifische chemosynthetische Symbiose sind die grabenden Muscheln der Familie Lucinidae. Die Nährstoffbeziehung entwickelt sich vor allem in den Kiemen der Muschel, während der Rest des Körpers nicht bekannt ist, dass er Symbionten beherbergt. Antimikrobielle Peptide (AMPs) spielen eine Rolle bei der Vermittlung von Spezifität in anderen intrazellulären Symbiosen, wurden jedoch bei Muscheln bisher nicht untersucht. Um diese Wissenslücke zu schließen, habe ich de novo ein Referenz-Transkriptom für die Muschel Loripes orbiculatus erstellt und eine maßgeschneiderte Pipeline zur Identifizierung von AMPs entwickelt. Wir haben Trinity verwendet, um die vorverarbeiteten Rohdaten aus vier verschiedenen Organen zu assemblieren: Kiemen (mit Symbionten), Mantel, Fuß und Eingeweidemasse (alle ohne Symbionten). Transdecoder wurde verwendet, um offene Leserahmen vorherzusagen. AMPs wurden auf Grundlage von Homologie (BLASTp gegen die dbAMP-Datenbank) oder physikochemischen Eigenschaften (AMPIR) vorhergesagt. Von den 106.974 Transkripten, die erfolgreich mit Trinity assembliert wurden, blieben 42.000 potenzielle AMPs auf Grundlage physikalisch-chemischer Eigenschaften vorhergesagt, während 709 AMPs durch Homologie identifiziert wurden. Von diesen 709 AMP-Familien waren Buthinin, Ovispirin und Lc-NK-Lysin die vorherrschenden AMP-Familien basierend auf der Expression. Hier identifiziere ich zum ersten Mal AMPs im Transkriptom von Loripes orbiculatus. Organspezifische differentielle Expressionsanalysen der für AMPs kodierenden Transkripte zeigten einen funktionellen Unterschied zwischen den

Organen. Die Genomkartierung unter Verwendung eines Genomentwurfs zeigte, dass weitere Anstrengungen erforderlich sind, um ein Referenzgenom zu erstellen und unterschiedliche Heterozygotiegrade in Populationen von L. orbiculatus zu untersuchen.

**List of abbreviations**

AMP: Antimicrobial peptide

ASABF: Ascarussuumantibacterial factor

DE: Differential expression

FDR: False discovery rate

MAG: Metagenome-assembled genome

MGD-1: *Mytilus galloprovicialis* defensin 1

ORF: Open reading frame

PAMP: Pathogen Associated Molecular Pattern

PAV: Prescence/absence variation

PCR: Polymerase chain reaction

PPR: Pattern Recognition Receptor

ROS: Reactive oxygen species

TPM: Transcripts per million

WT: Wild-type

SV: Structural variation

## Introduction

### Symbiosis: *living with others.*

Eukaryotic and prokaryotic organisms have evolved symbiotic relationships that are beneficial to both partners on multiple occasions. Across different branches of the three of life we could observe how these relationships shape the physiology, ecology and, evolution of organisms (Gilbert et al., 2015). A simple way to classify the symbiotic relationship is to divide them into nutritional and defensive types (Wein et al., 2019). Explained as the type of "currency" species exchange between them, which requires molecular adaptations and inheritance to guarantee stable symbiotic relationships.

Permanent, intimate, symbiotic relationships between organisms for the acquisition of a new trait are widespread. Well-studied examples of such beneficial relationships include aphids-*Buchnera,* legumes-*Rhizobium* or the squid-*Vibrio* symbiosis models, among others (Ohbayashi et al., 2020). In the deep of the Hawaiian coastline, a defensive symbiosis develops. The luminous bacterium *Vibrio fischeri* provides to the bobtail squid *Euprymna scolopes* a counter-illumination system that allows the squid to camouflage at night, avoiding predators (Nyholm & McFall-Ngai, 2004). Almost all aphids present in their bacteriocytes a round-shaped bacteria from the genus *Buchnera,* these bacteria are maternally transmitted and encoded in their genome essential biosynthesis pathways of amino acids and vitamins that the host relies on. At the same time, the host provides to *Buchnera* to some nutrients and precursors, reveling a mutually dependent relationship (Shigenobu et al., 2000).

Chemosynthetic symbioses refer to nutritional collaborations between eukaryotic hosts and chemosynthetic bacterial symbionts. Animals form symbiotic partnerships with such bacteria, enabling them to utilize compounds like sulfide that would otherwise be inaccessible. As explained in detail by Sogin et al. In 2021, Winogradsky conducted

a series of experiments that initially revealed how sulfur-oxidizing bacteria derive their growth energy from a reduced inorganic compound, hydrogen sulfide. This phenomenon, termed chemolithoautotrophy and commonly recognized as chemosynthesis, represents the sole alternative form of primary production on Earth apart from photosynthesis. The discovery of chemosynthetic relationships dates back to 1981 when they were first observed at hydrothermal vents in the deep sea, specifically in *Riftia pachyptila* trophosomal tissue, a species of mouthless giant tube worm (Cavanaugh et al., 1981). After so, it became evident that chemosynthetic symbioses had been overlooked for an extended period. These associations are widespread, occurring in diverse habitats ranging from the deep sea to shallow waters worldwide. Notably, at least seven animal phyla have been identified as hosts for chemosynthetic symbionts (Dubilier et al., 2008).

**Lucinids depend on their chemosymbiotic bacteria.**

Burrowing clams from the family Lucinidae are a prime example of specific chemosynthetic symbiosis. As depicted in *Figure 1,* It has been demonstrated that lucinid eggs are aposymbiotic, which means they are free of the symbionts the adults host in their gills. Despite their larvae being in contact with numerous bacteria, only when they migrate into the sediment, and the formation of the gills happens, they acquired a specific bacterium from the genus *Candidatus* Thiodiazotropha in their gills (Zauner et al., 2022). While the nutritional aspects of this symbiosis are widely studied, aspects like recognition, establishment, and maintenance of this relationship are less understood.

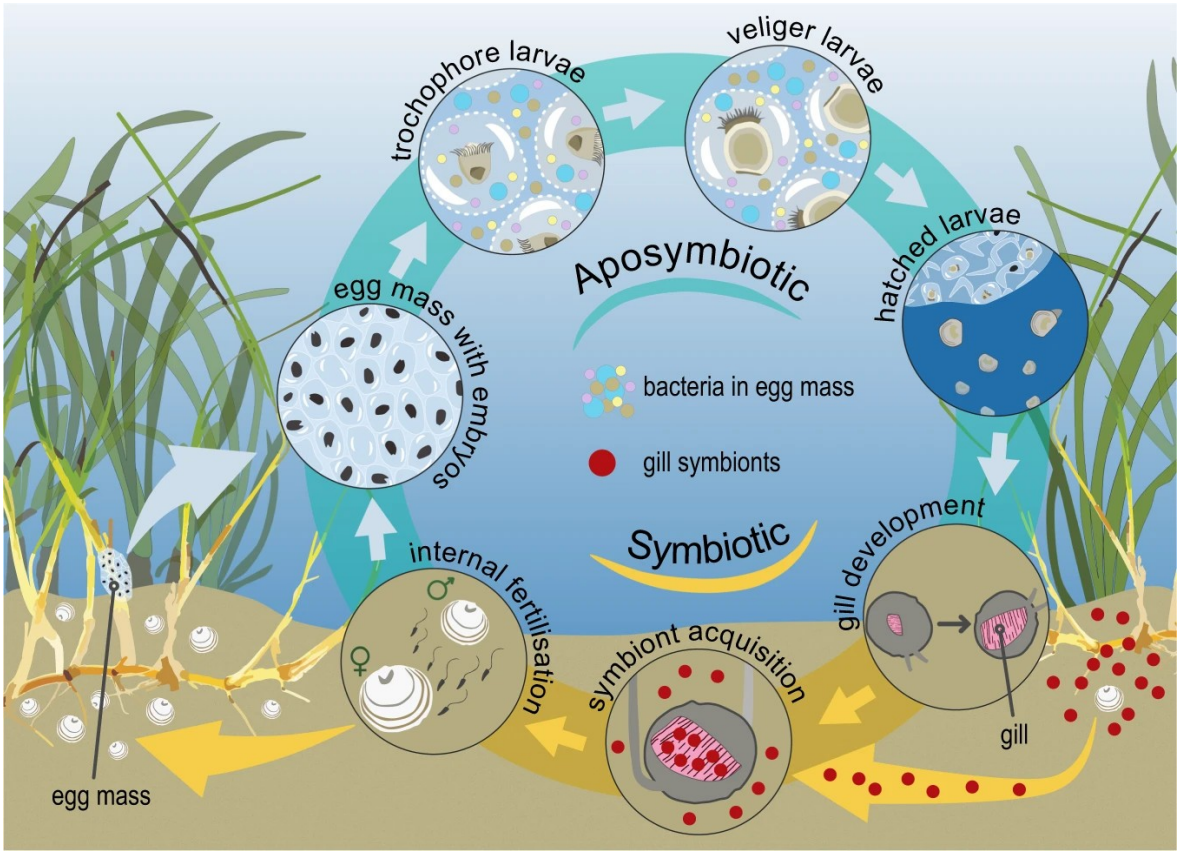



**Figure 1. Lucinds life cycle.** Schematic depicting the life cycle of Lucinds, fixation of the eggs mass to the seagrass blades happens while it still attached to the maternal mucus tube. The larvae within these egg masses exist in an aposymbiotic state, progressing through a trochophore stage (approximately 8 days post-egg mass release) and a veliger larval stage (around 12 days post-egg mass formation) before emerging as fully developed pediveliger larvae. Following hatching, gill development is initiated, and the settled juvenile is colonized by sulfur-oxidizing endosymbionts. Environmental bacteria are visualized in pink, blue, yellow, and brown, while the specific sulfur-oxidizing symbiont is represented in red. *Taken from Zauner et al., 2022.*

All the bacteria described as the main symbiont of lucinds are sulphide-oxidising gammaproteobacteria, they are hosted in a specific cell type in the gills called bacteriocytes and their main role is to provide fixed organic carbon compounds to the host (Taylor & Glover, 2021) As shown by König et al., 2016 and by Petersen et al.,. in 2016 these symbionts also have the molecular machinery to fix nitrogen, which could be an important process for the net nitrogen budget in the seagrass ecosystems (Cardini et al., 2019).

**Overview of defense mechanisms of the bivalve's Immune system.**

Considerable research has been conducted on the immune systems of commercially important bivalves, with a particular focus on families such as Mytilidae, Ostreidae, and Veneridae (Roch, 1999 and Tiscar & Mosca, 2004). In the bivalve's immune system, only the innate immune system is present; this represents a challenge when facing bacterial species when the organism is doing a routine task like filter feeding or digging channels that are used for oxygen transportation. The general components of the innate system are divided into four main mechanisms: Mechanical, physiological, Phagocytic/endocytic, and inflammatory barriers (Marshall et al., 2018).

Specifically in bivalves, we have as mechanical barriers two mechanisms that are constitutively present: First, the hard shell that is employed to support and protect the soft tissues, and the second is a mucosal layer that traps exogenous agents and facilitates their elimination by both ciliary activity and chemical/biological degradation through the action of immune factors associated to it. Between the numerous factors associated with the mucosal membrane, we can find agglutinins and lysozyme (Figure. 2). (Allam & Raftos, 2015).

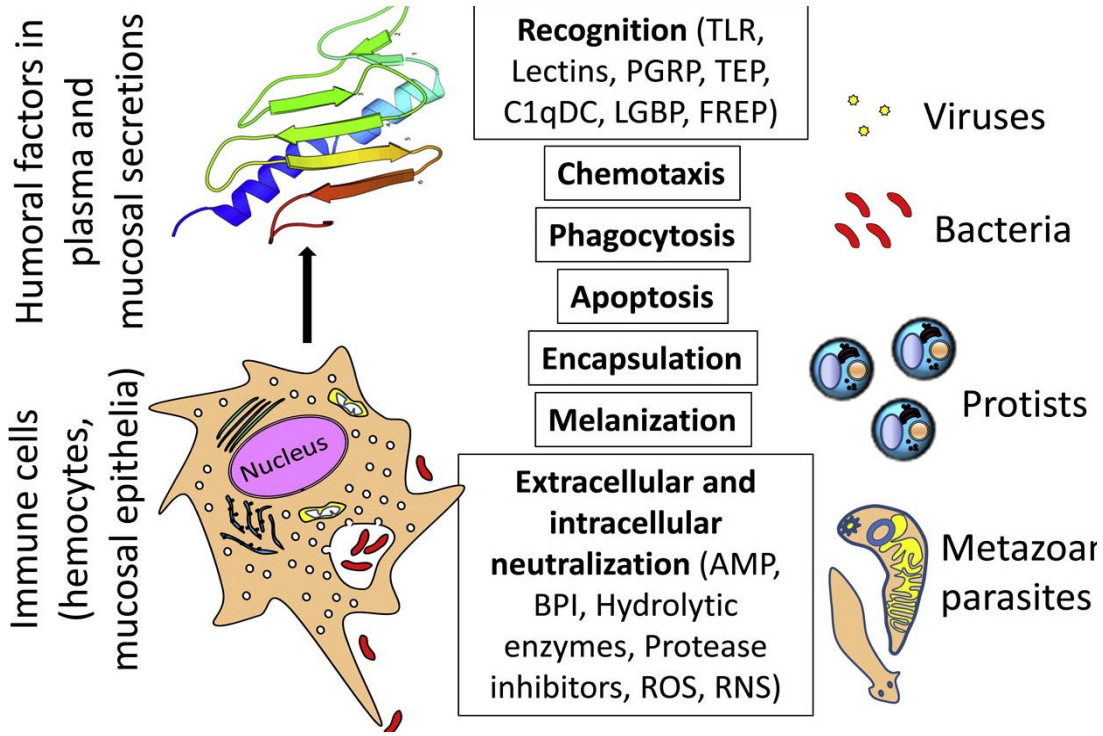



**Figure 2.** Mechanisms for pathogens clearance bivalve's immune system. *Taken from Allam and Raftos. 2015.*

Upon activation of a pattern recognition receptor (PPR) by a pathogen associated molecular pattern (PAMP), generally present on the surfaces of exogenous microorganisms, several biological barriers activate, depending on the nature and location of the immune stimuli (Canesi et al., 2022). Hemocytes in bivalves are crucial cells playing a regulatory role in immunity homeostasis. These cells are present in all the different tissues of bivalves, circulating in the hemolymph; besides being involved in processes like phagocytosis and endocytosis, or even shelf formation and growth between the physiological response after infection, they can exert chemical mediators released, like the complement system which attaches to the target microorganism and promotes lysis and phagocytosis, which produces reactive oxygen species (ROS) pump into the phagosome. (Bouallegui, 2019). One of their leading immune roles is the release of enzymatic effectors, for example, between many substances released by the hemocytes, lysozymes, cytokines, and antimicrobial peptides (AMPs) (Zannella et al., 2017).

Conventionally speaking, antimicrobial peptides[1] are short molecules ranging from 10 to 100 amino acids; they can be classified by either their structural properties as α-helical, β-sheet, loop, and extended peptides (Di Somma et al., 2020) or by their amino acid composition as, for example, proline, histidine or glycine-rich peptides (Huan et al., 2020). The mechanisms of action of AMPs include targeting the cytoplasmic membrane and forming pores; they can also cross the membrane into the cytoplasm and act on intracellular molecules like nucleic acids, tampering with DNA replication or transcription or interfering with protein synthesis via ribosomal obstruction (Zhang et al., 2021).

AMPs were first identified in bivalves in the mussel *Mytilus galloprovincialis*, which shared sequence similarity with the arthropod defensin (Canesi et al., 2022), and it has since considered that AMPs constitute essential molecules generated by hemocytes that help bivalves survive in waters saturated with microorganisms (Yang et al., 2022).

**Antimicrobial peptides support symbiosis as mediator molecules.**

For over thirty years, antimicrobial peptides (AMPs) have been recognized as essential mediators of the innate immune response in plants and animals. Traditionally, the function of AMPs was to aid in destroying pathogenic microorganisms, But it has recently been shown that eukaryotic hosts may also manufacture AMPs when they engage symbiotically with bacteria (Mergaert, 2018). Several symbioses demonstrated the critical role of AMPs in maintaining the homeostasis between the organisms; for

1*Although for the development of the present manuscript, we will continue calling antimicrobial peptides, growing evidence of neglected roles of small peptides is making a redenomination of these molecules to be called small proteins for a review on the topic; I suggest the following bibliography: (Orr et al., 2020 and Yadavalli & Yuan, 2022)*

example, the palm weevil *Rhynchophorus ferrugineus* maintains its endosymbiont in a bacteriocyte by producing ColA. This peptide has bactericidal activity and inhibits the division of the endosymbiont. If ColA production is inhibited, for example, with RNAi, the endosymbiont can divide again, escape from the bacteriocytes, and invade other insect tissues (Login et al., 2011).

Another example of AMPs regulating symbiosis is the previously mentioned Aphids-*Buchnera* symbiosis, in which aphids secreted signal peptides that are cysteine-rich in their bacteriocytes. These genes are first expressed at a developmental point coincident with the strict incorporation of symbionts in the cells that contribute to the bacteriocyte, and this bacteriocyte-specific expression is maintained throughout the aphid's life (Shigenobu & Stern, 2013). Also interesting is the symbiosis between plant roots from the genus *Aeschynomene spp.* nodules and *Bradyrhizobium ,* where it has been demonstrated that AMPs modify bacterial morphology, causing a different shape compared with wild-type (WT) strains when they are exposed to these AMPs in pure cultures (Czernic et al., 2015). Interestingly, bacteria promote pore-forming proteins in their membrane, allowing molecule exchange in the root nodules (Guefrachi et al., 2015) (Figure. 3).

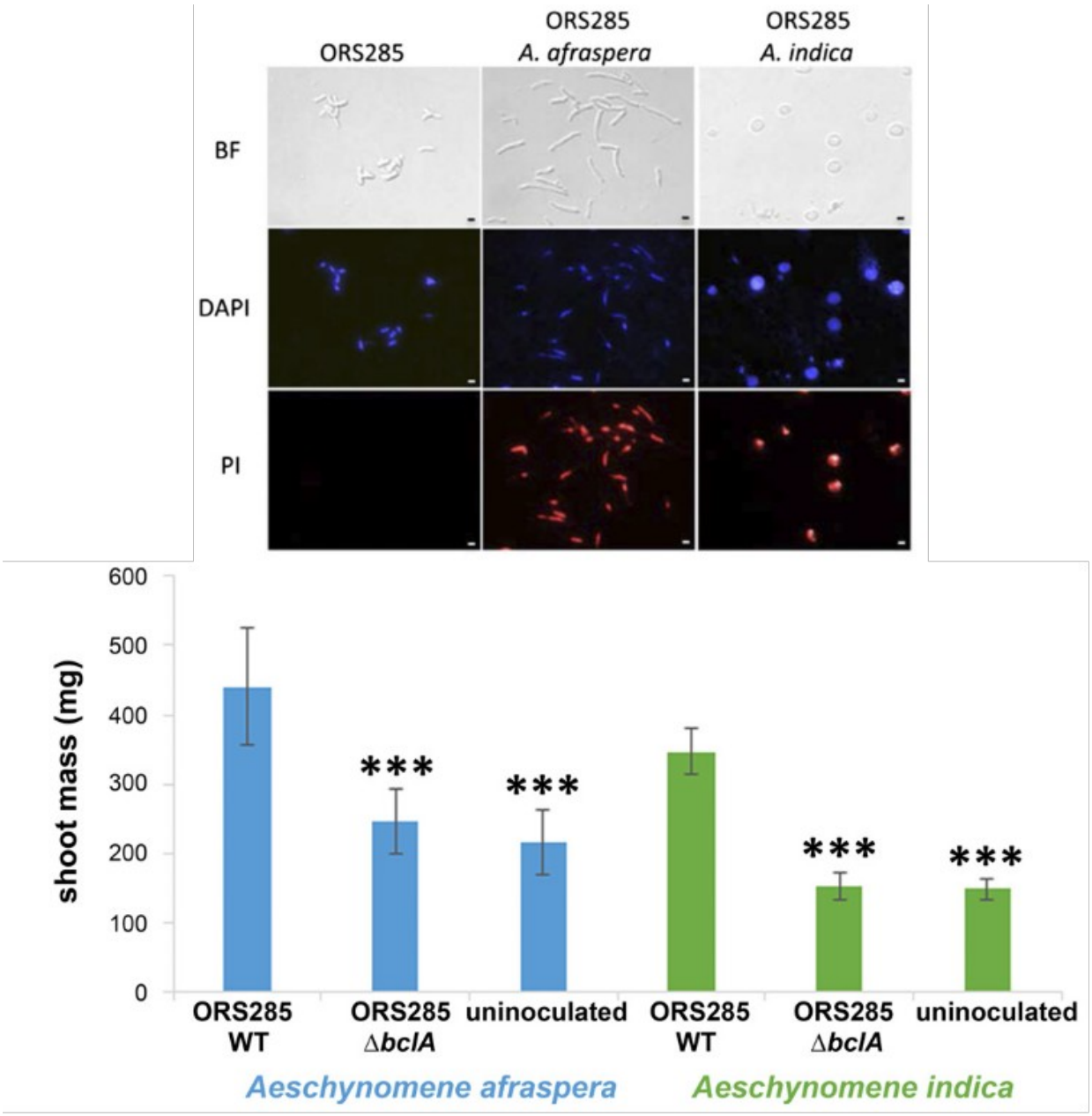



**Figure 3.** Symbiosis in *Aeschynomene spp*. nodules with *Bradyrhizobium*. Top). Terminal differentiation of bacteria exposed to plant AMPs coming from two different *Aeschynomene* species. Bottom). Shoot mass difference between WT bacteria, knockout bclA strain and uninoculated. *Taken and modified from Czernic et al. 2015 and Guefrachi et al. 2015.*

**Aims of this sudy**

There are different strategies for the establishment of symbiosis; while some are vertically transmitted from mothers to offspring via the germ line tissues, others are horizontally transmitted from the environment (Bright & Bulgheresi. 2010) These transmission strategies present different challenges to study, but after the establishment of the symbiosis, the molecular mechanisms involved in the maintenance of the symbiotic relationship and in the homeostasis of the immune system are of great importance to determine for understanding the long-term cross communication process, and lastly how each organism can provide some benefit to its partner.

In this study, I focus on determining if Loripes orbiculatus was producing AMPs and if they were differentially expressed in their tissues. I chose to investigate this molecule due to its importance in the bivalve's immune response and, secondarily, because of the challenge of identification in transcriptome data, the analysis of AMPs was excluded in previous studies.

## Methods

The current exploratory study aims to identify potential AMPs that *L. orbiculatus* produces. Of particular interest are those that are expressed exclusively in the clam gills and may have an impact on symbiosis maintenance. I conducted two bioinformatic approaches to achieve my goals: a de novo transcriptome assembly and genome annotation analysis.

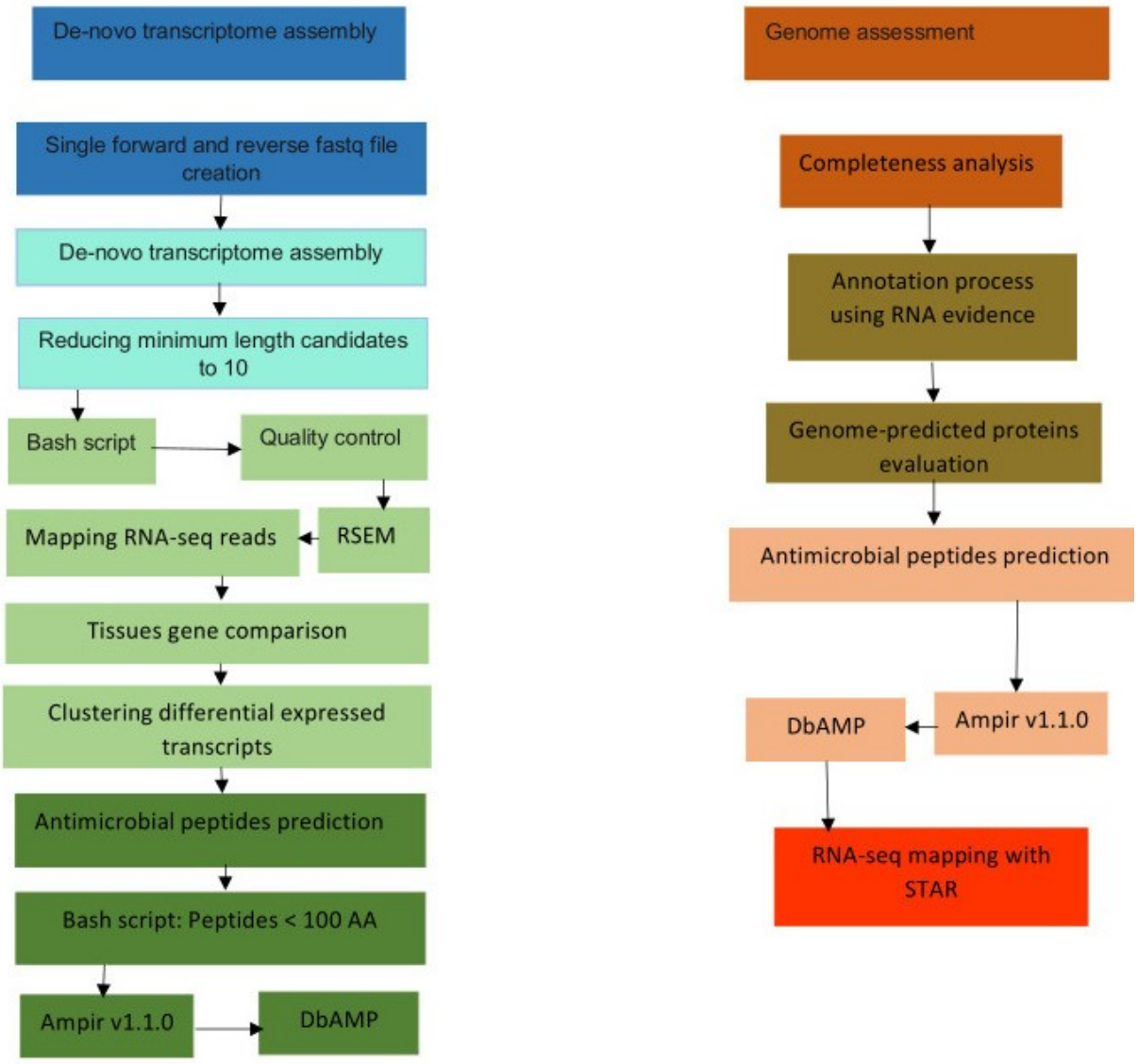



**Figure 4.** Implemented pipelines developed for the present study.

**De-novo transcriptome assembly, differential expression analyses, and AMP prediction**

**Raw reads preprocessing.**

Raw reads were obtained from a published study of *Loripes orbiculatus* transcriptome carried out by Yuen et al. 2019. In short, the raw RNA-seq reads came from the following treatments: they dissected in triplicates different organisms, separating four different tissues (gills, visceral mass, foot, and mantle) and placing them into RNA later until RNA extractions. Non-directional RNA libraries were prepared from Poly A enriched RNA extracts for retaining preferentially mature eukaryotic mRNA, and sequencing was carried out with Illumina HiSeq 2500 technology.

As a guide, I followed the Yuen et al. preprocessing pipeline to ensure certain reproducibility in sample treatment. However, I considered package updates, which could influence some changes in the results. I first used Rcorrector v1.0.4 (Song & Florea, 2015) with default parameters to rectify errors introduced during sequencing; this was particularly helpful for some Illumina reads; read pairs that were marked as uncorrectable were discarded. After that, rRNA contamination from the data was removed with SortMeRNA v4.3.6 (Kopylova et al., 2012) under default settings, comparing our raw reads against 5S, 5.8S, 16S, 23S, 18S, and 28S rRNA sequences, dropping out pair reads that overpass the similarity threshold. This step is crucial for focusing downstream analyses on the biologically relevant mRNA. Following the filtration and correction of reads, trimming was performed using Trimmomatic v0.36 (Bolger et al., 2014), employing the specified settings SLIDINGWINDOW:4:15 MINLEN:36, ILLUMINACLIP option with the adapters/TruSeq3-PE-2.fa:2:30

configuration for Illumina adapter sequences was removed from the paired-end libraries. Last, quality control of the surviving reads was assessed with FastQC v0.12.1 (Andrews, 2017/2024) for guarantee downstream analyses.

**De-novo transcriptome assembly.**

I merged the libraries into a single forward and reverse fastq file for constructing a reference transcriptome with all the possible transcript variants. The De-novo transcriptome assembly of the combined RNA-seq data was performed utilizing Trinity (Grabherr et al., 2011) as the core software package. I then used TransDecoder v5.5.0 (http://transdecoder.sourceforge.net), modifying the minimum length of the candidates' coding regions of the transcripts from the standard 100 amino acids to 10 amino acids long. This modification is crucial for retrieving transcripts encoding antimicrobial peptides. CD-HIT and CD-HIT-Est (W. Li & Godzik, 2006) were employed to reduce transcript complexity and predict proteins based on sequence similarity. I made a simple bash script to select only the longest transcript isoforms, which is fundamental for reducing the database for further analysis.

To evaluate the transcriptome assembly's quality, I conducted several controls. I initially ascertained the percentage of utilization of RNA-seq reads that maps to our assembly as a crucial verification criterion. Second, I assessed the transcriptome's completeness using Busco v5.4.3 (Simão et al., 2015), which examines the number of core genes conserved in a phylogenetic clade that should be represented in the assembled transcriptome. The completeness analysis was made every time that a reduction by similarity or length was carried out to trace the loss of crucial marker genes.

**Differential gene expression analysis.**

First, I used the RSEM program (B. Li & Dewey, 2011)contained in the Trinity pipeline tool set to quantify the transcripts and gene abundances. RNA-seq reads were mapped to the assembled transcriptome using the alignment-based algorithm. By utilizing it, I were able to obtain two practical expression measures: one that estimated the number of RNA-seq fragments derived from each transcript and the other that produced a normalized metric that took into account the length of each transcript, as well as the size of the library and the, reads mapping to it. Subsequently, I generated a matrix containing the normalized expression values in addition to the raw counts. With the two matrices, I removed ultra-low expressed transcripts that would potentially represent assembly artifacts.

To determine the differential gene expression in the four distinct tissues of L. orbiculatus, I made an initial comparison of the tissue triplicates as a way for quality control using the matrices mentioned above. After QC, I employed Deseq2 (Love et al., 2014) for differential analysis of counts transcript per triplicates, and I clustered the differential expressed transcripts, extracting the most significant fold-changed transcripts with a minimum charge of 4-fold.

**Antimicrobial peptide prediction.**

I evaluated the potential of the candidates' encoded proteins in our transcriptome to be antimicrobial. For this purpose, I selected all the peptides that were predicted to be smaller than 100 AA with an in-house bash script. I utilize two methods to determine AMP potential; the first approach employs Ampir v1.1.0 (Fingerhut et al., 2021), which identifies AMPs based on the physicochemical properties of amino acid sequences.

This process is optimal for retrieving new classes of AMP families. The second procedure was using our candidates' proteins as the query in BLASTp (Altschul et al., 1990), which searches by similarity against a curated database of AMPs (Jhong et al., 2022)

I then created a heatmap of only the AMPs that meet the previously stated criteria of 4-fold changes in expression level between the four different tissues of *L. orbiculatus* by cross-comparing the AMP databases of our candidates, as determined by both methods (Physicochemical and similarity) and our differentially expressed transcripts.

**Genome assessment, annotation, and RNA-seq read mapping.**

I obtained a draft genome of *L. orbiculatus* through a collaboration with Associate Professor Dr. Tanguy from the Roscoff Marine Station. To assess the draft genome's condition, I first performed completeness analysis with Busco v5.4.3, n50, and l50 genome metrics with Quast v5.2.0.(Gurevich et al., 2013)

**Annotation using RNA evidence.**

The annotation process was carried out using Braker2 v2.1.6 (Brůna et al., 2021); this program uses both genome information and RNA-seq reads as extrinsic evidence to improve the annotation process of the genome. I evaluated the genome-predicted proteins with Busco v5.4.3 to determine the completeness and accuracy of the annotation. After that, I followed the broad steps of Antimicrobial peptide predictions described above to select the candidates AMPs encoded in the genome.

I then utilize Star v2.7.11a (Dobin et al., 2013), a gap-aware program, to map the RNA-seq reads to the genome. In short, this was for generating a matrix of reads mapping to the genome and then doing the differential expression analysis in the base of the genome-encoded AMPs. Qualimap v.2.3 (García-Alcalde et al., 2012) was utilized as a tool for the evaluation of RNA-seq reads aligned to the genome.

Finally, as an integrative approach, I compared the AMPs inferred with the de-novo transcriptome assembly to the ones encoded in the genome utilizing BLASTp v2.15.0+. I visualized the genome and reads aligned to it with BasePlayer v.1.0.2. (Katainen et al., 2018)

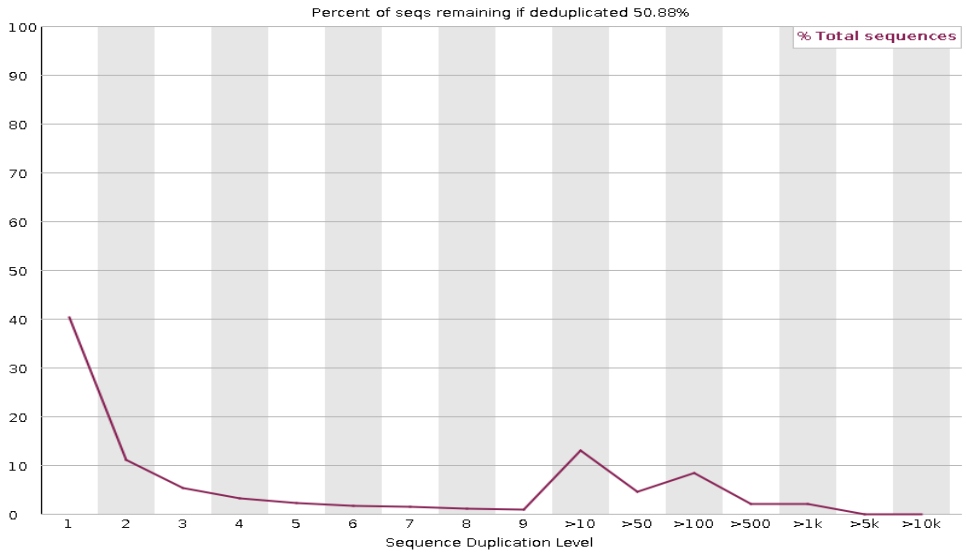
## Results

### Transcriptome assembly.

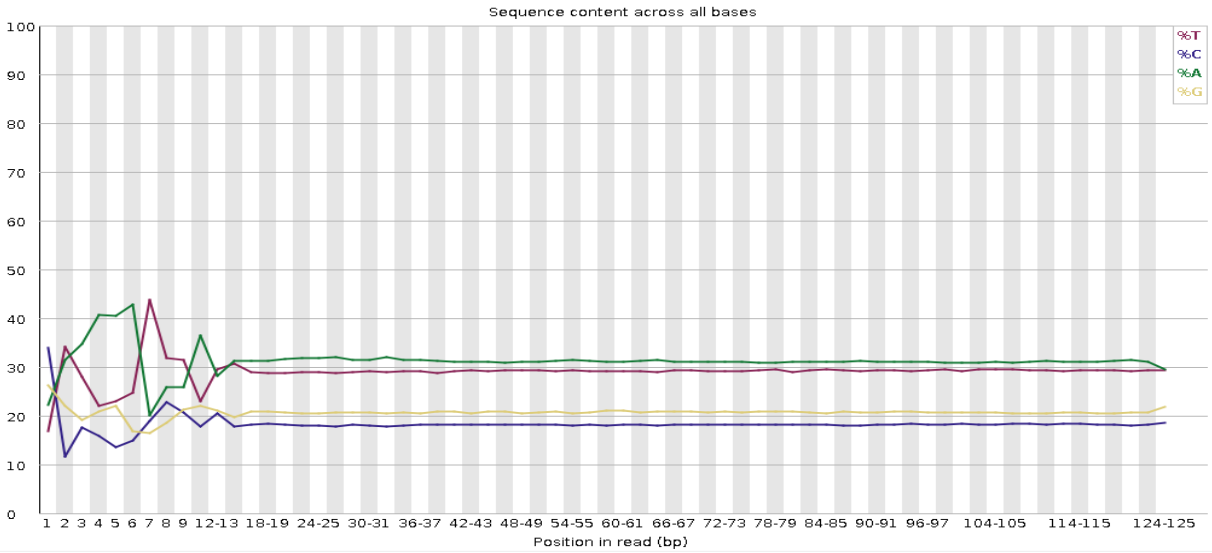
Sequencing the 12 libraries from the four tissues resulted in over 541 million reads, averaging 45 million per library. The preprocessing resulted in a minor reduction of the whole database due to more than 91% of all reads remained. After quality control with FastQC, some of the libraries presented some warnings as I can see in figure 5 such as sequence duplication levels, per base sequence content and per tile sequence quality. Still, those warmings are commonly found in RNA-seq data (Delhomme et al., 2014).

The reference transcriptome assembled with Trinity, resulted in a total of 786,678 transcripts; after transcriptome reduction via transcripts clusterization with CD-HIT-EST I kept 591,518 unique putative transcripts. I employed Transdecoder with a minimum open reading frame of 10 AA to predict the coding regions of the transcripts; I obtained a total of 8 million proteins encoded. Reduction by similarity of the predicted proteins database was not efficiency, because I kept with 7,152,226 predicted proteins, after this post-assembly processing step.

## Sequence Duplication Levels




## Per base sequence content




## Per tile sequence quality

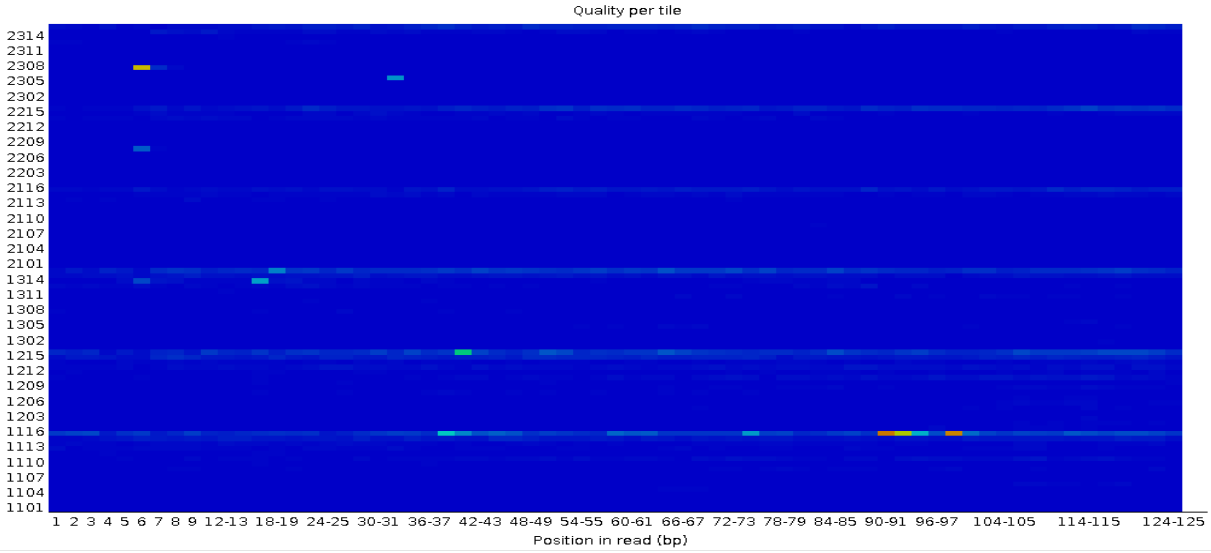


**Figure 5.** Common RNA-seq reads library warnings found with FastQC. Top) The peaks symbolize the amount of repetitive sequences. Middle) At the beginning of the RNA-seq reads, there is not a stable proportion of GC ratio. Bottom) Some tiles are showing possible sequencing errors.

I explore the nature of the predicted proteins (Figure 6.) I identify an algorithmic bias that I refer to as alternative open reading frame bias (ORF), which means that for every single transcript, Transdecoder determines, on a rolling basis, an ORF smaller, annotating it as an isoform of the protein until it reaches the end of the transcript. For this reason, I coded a bash script to select only the longest predicted isoforms, which resulted in a 93% reduction of the predicted proteins.

```
>57784_TRINITY_DN13_c0_g1_i2.p1 type:complete len:101 57784_TRINITY_DN13_c0_g1_i2:231-533(+)
ATGCATAGGCAGGTTATCGTACTGGAGACATCCACTTCTAGAAAAGCATTCACTTTATTTTTCCCCGAAATCAAGTTATCAGTATCGAAAACCAACAAGCTAGTTCACTCACAGGCTAAATTCCATCGAAGCAACACAGTTTTAAGAAGTCCTAGGAACGTCCGAATGCCGTTTCTTTCTGGTGTGGA
TGGAAAGCTCACGACTAGAGTATAGAACTGCCTTCAAGATTAATATGTACCAGAATTCTGGTAAAATGTATAACTGGCAGATGACAATTTGA
>57784_TRINITY_DN13_c0_g1_i2.p2 type:complete len:59 57784_TRINITY_DN13_c0_g1_i2:880-1056(+)
ATGGTAACATGTCCGTATAGAATAGATCCCCTTACATTCCACTTTTATATAGTTGTGTTATGGTTTTACGTCCCACCAACAGCTAAGGTCATACGGAGACGGGACCTCGGTTTAAAGTCTCATCCGAAAGACTGCAAAAATTCTGTTTTCTTATGTTGTGGCTCATATATATTTTGA
>57784_TRINITY_DN13_c0_g1_i2.p3 type:complete len:54 57784_TRINITY_DN13_c0_g1_i2:1129-968(-)
ATGACAAAATATTTCTATATTTTGAGTCCTTTAATCACAGGGAATGTTCATTTCAGCAAATATCAAATTTTGTTCAAAATATATATGAGCCACAACATAAGAAAACAGAATTTTTGCAGTCTTTCGGATGAGACTTTAAACCGAGGTCCCGTCTCCGTATGA
>57784_TRINITY_DN13_c0_g1_i2.p4 type:complete len:45 57784_TRINITY_DN13_c0_g1_i2:236-102(-)
ATGCATTATACTGCGATATTTCACATCTGTATGAATGATAATTTTCAGATGAAAAGTTGTGATATTTTTTCTTATTTTTGCTCAAAACATAAATCTTGGGCACACATTAGAATCGCCACGCTGAGGCGGTTCTAA
>57784_TRINITY_DN13_c0_g1_i2.p5 type:complete len:29 57784_TRINITY_DN13_c0_g1_i2:534-620(+)
ATGCATCCTTTTAGTTGCAGTTTCTTTTTAATGATATCAGGGTCGAAAAACCATCAACAATTGCAAATTGTCGAATTTTATAGTTAA
>57784_TRINITY_DN13_c0_g1_i2.p6 type:complete len:28 57784_TRINITY_DN13_c0_g1_i2:539-456(-)
ATGCATTCAAATTGTCATCTGCCAGTTATACATTTTACCAGAATTCTGGTACATATTAATCTTGAAGGCAGTTCTATACTCTAG
>57784_TRINITY_DN13_c0_g1_i2.p7 type:complete len:27 57784_TRINITY_DN13_c0_g1_i2:102-22(-)
ATGAGTACCAACCATCTCTGTTTTAGAAAAAAAAAATAAGAAAATAAATTATACCCATGTAAACCCCAGTTTAACTATATAA
>57784_TRINITY_DN13_c0_g1_i2.p8 type:5prime_partial len:26 57784_TRINITY_DN13_c0_g1_i2:1251-1174(-)
TTTCAAAACATTTTTTTTTATTTTGTAGATTTGTTCTCTATAACAACATGTAAAATGAAATACTCAATACAATTATGA
>57784_TRINITY_DN13_c0_g1_i2.p9 type:complete len:25 57784_TRINITY_DN13_c0_g1_i2:878-804(-)
ATGGCATTTACAATGCACCCTTGTTCCCAGATTAGTTTGAGTTACTTCCCTTCCTTTGCATGTTATCCAATTTAG
>57784_TRINITY_DN13_c0_g1_i2.p10 type:complete len:20 57784_TRINITY_DN13_c0_g1_i2:46-105(+)
ATGGGTATAATTTATTTTCTTATTTTTTTTTCTAAAACAGAGATGGTTGGTACTCATTAG
>57784_TRINITY_DN13_c0_g1_i2.p11 type:5prime_partial len:18 57784_TRINITY_DN13_c0_g1_i2:2-55(+)
ATACCCCTGCACCACACATTTTATATAGTTAAACTGGGGTTTACATGGGTATAA
>57784_TRINITY_DN13_c0_g1_i2.p12 type:5prime_partial len:18 57784_TRINITY_DN13_c0_g1_i2:1252-1199(-)
TTTTCAAAACATTTTTTTTTATTTTGTAGATTTGTTCTCTATAACAACATGTAA
>57784_TRINITY_DN13_c0_g1_i2.p13 type:complete len:17 57784_TRINITY_DN13_c0_g1_i2:757-807(+)
ATGTCATGTGAAGCTCACTCTTTAAAGATGCAGTTCCATTATTTTACCTAA
>57784_TRINITY_DN13_c0_g1_i2.p14 type:complete len:15 57784_TRINITY_DN13_c0_g1_i2:1086-1042(-)
ATGTTCATTTCAGCAAATATCAAATTTTGTTCAAAATATATATGA
>57784_TRINITY_DN13_c0_g1_i2.p15 type:complete len:14 57784_TRINITY_DN13_c0_g1_i2:1138-1179(+)
```

**Figure 6.** Alternative open reading frame bias in coding sequences when the minimun amino acid parameter is set to 10.

Our final reference transcriptome yielded 518,140 proteins encoded in 591,518 unique putative transcripts, of which 92,556 were longer than 100 AA. Our results showed that the completeness score of 94% (Single: 71%, Duplicate: 23%) remained consistent throughout the post-assembly reduction process. When I analyzed the expression level of the reference transcriptome with the total RNA-seq reads that mapped it, I found, as I can see in Figure 7, that there were 230,501 transcripts whose

general expression falls into only the range of very low expressed, ranging between 0 and 1 transcripts per million (TPM).

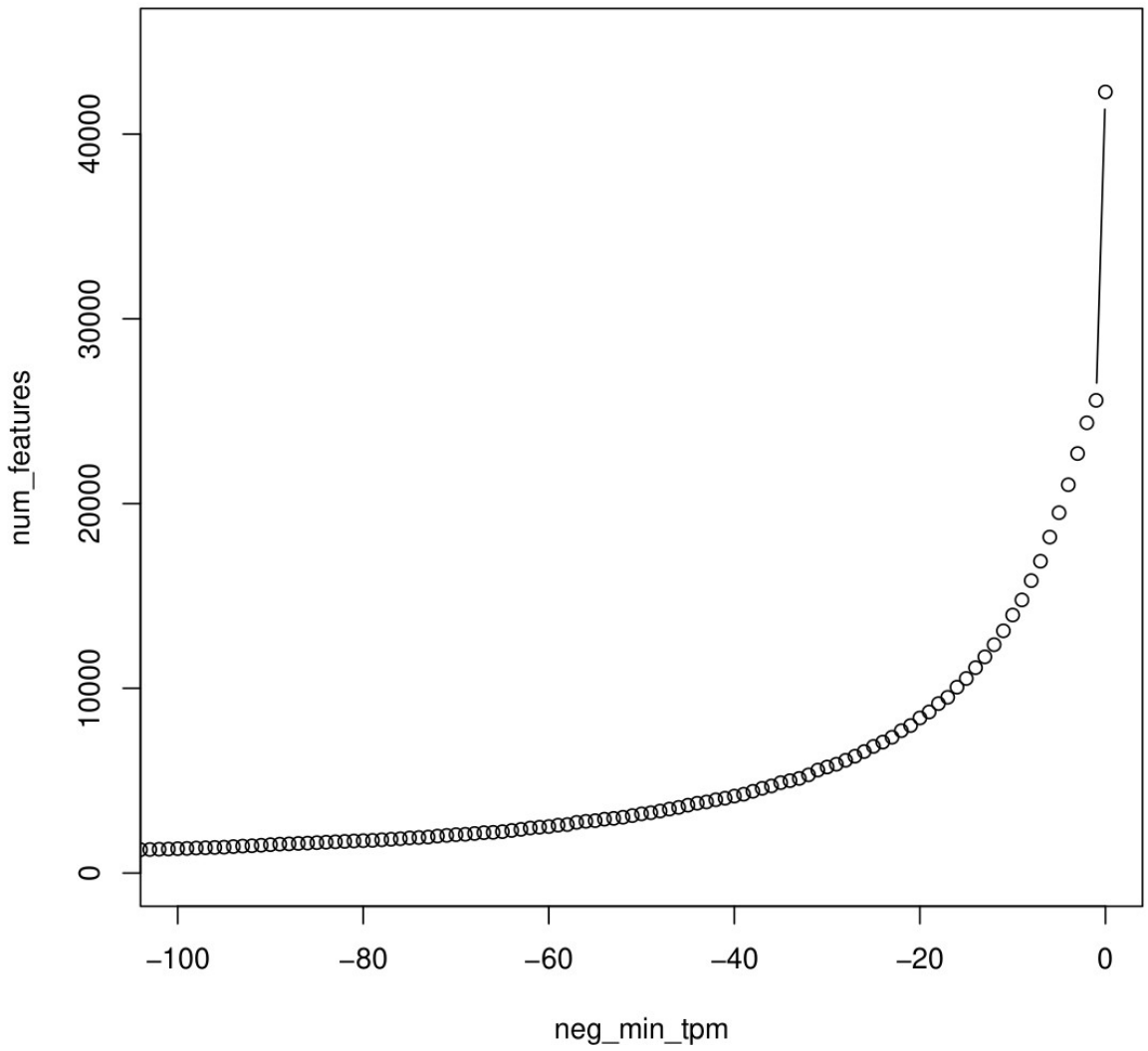



**Figure 7.** Distribution of the TPM per gene in the de-novo transcriptome assembly.

**Determination of Antimicrobial Peptides in *Loripes orbiculatus* transcriptome and differential expression analysis of their predicted AMP.**

First, I proceeded with the AMP prediction based on the physicochemical properties of the small peptides. I used the online tool AMPIR for this analysis, setting the AMP probability cut-off level at 80%. According to this cut-off level, the assembled transcriptome contains 42,284 putative AMPs.

After analyzing the differential expression (DE) pattern of AMPs obtained from the physicochemical prediction assay, I found that only 48 of the AMPs exhibited a DE pattern when grouped according to the DE patterns present throughout the samples. It is worth noting that the gills had the highest number of 22 up-regulated DE peptides, as shown in Figure 8.

In order to validate our predictions, I conducted a thorough search for matches of the predicted peptides in DbAmp, a well-known and curated repository of antimicrobial peptides (AMPs). For this purpose, I used BLASTp, a tool for comparing protein sequences with a database. By comparing our predictions with the DbAmp, I could filter out false-positive results and obtain a more accurate list of potential AMPs, the downside of this procedure, is that potential novel peptides are also removed from the analysis. Our search yielded 709 hits for AMPs in the database, a significant reduction of 98.3% from the initial number of candidates.

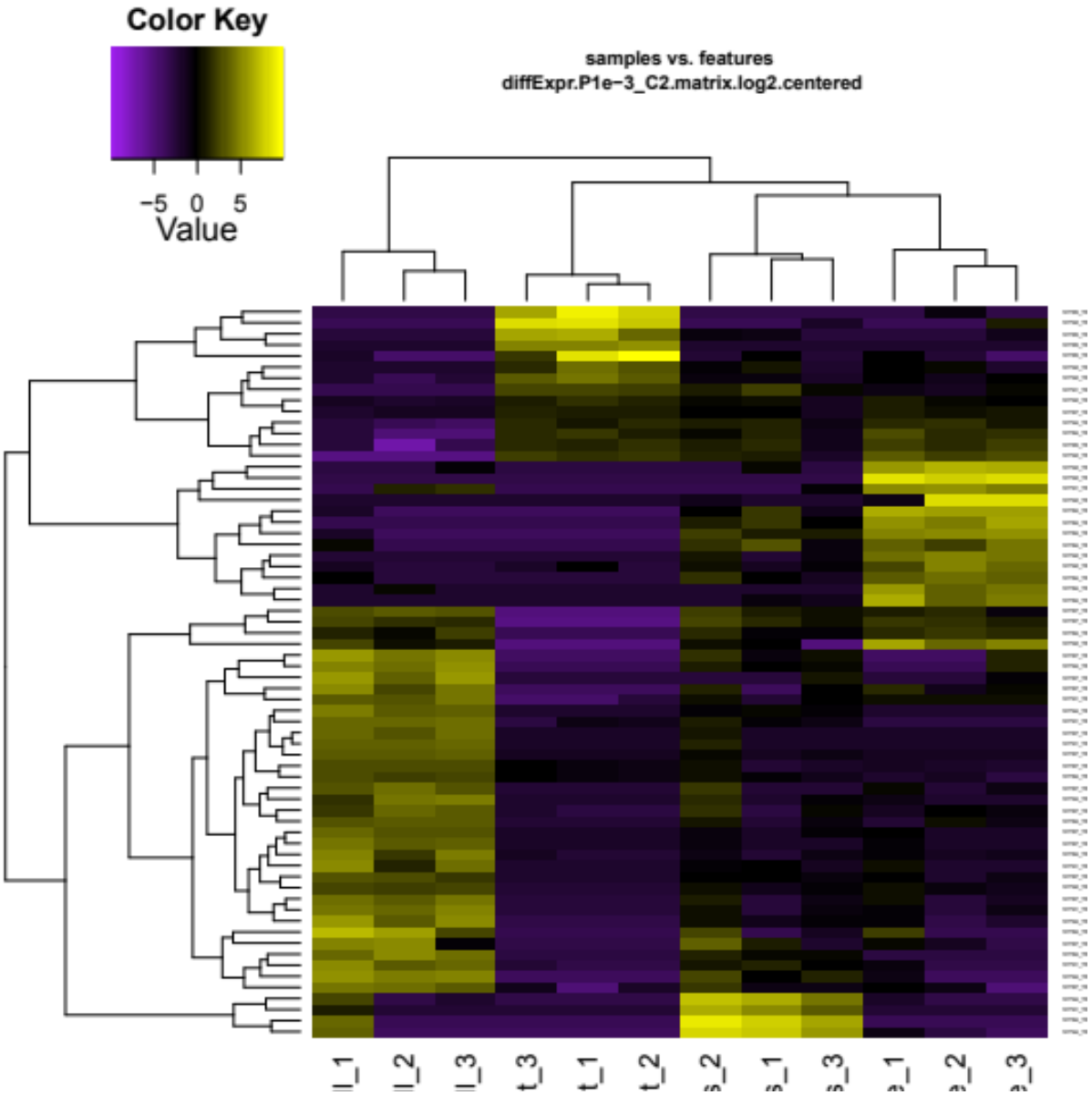



**Figure 8.** Differentlially expressed AMPs identified by physicochemical properties. Yellow = Upregulated peptides. Purple = Downregulate peptides.

I then analyzed the DE pattern of only the annotated AMPs following the same procedure state for the differential AMPs predicted by physicochemical properties (Figure 9). This comparison allowed me to analyze the differences between the predicted physicochemical properties and the annotated AMPs and better understand their expression patterns. I observed that 9 AMPs were upregulated in the gills, representing the tissue with an orchestra of AMPs working to maintain specific immune homeostasis. Strikingly, I did not find that the visceral mass had any specific AMP upregulated, and the foot only had one.

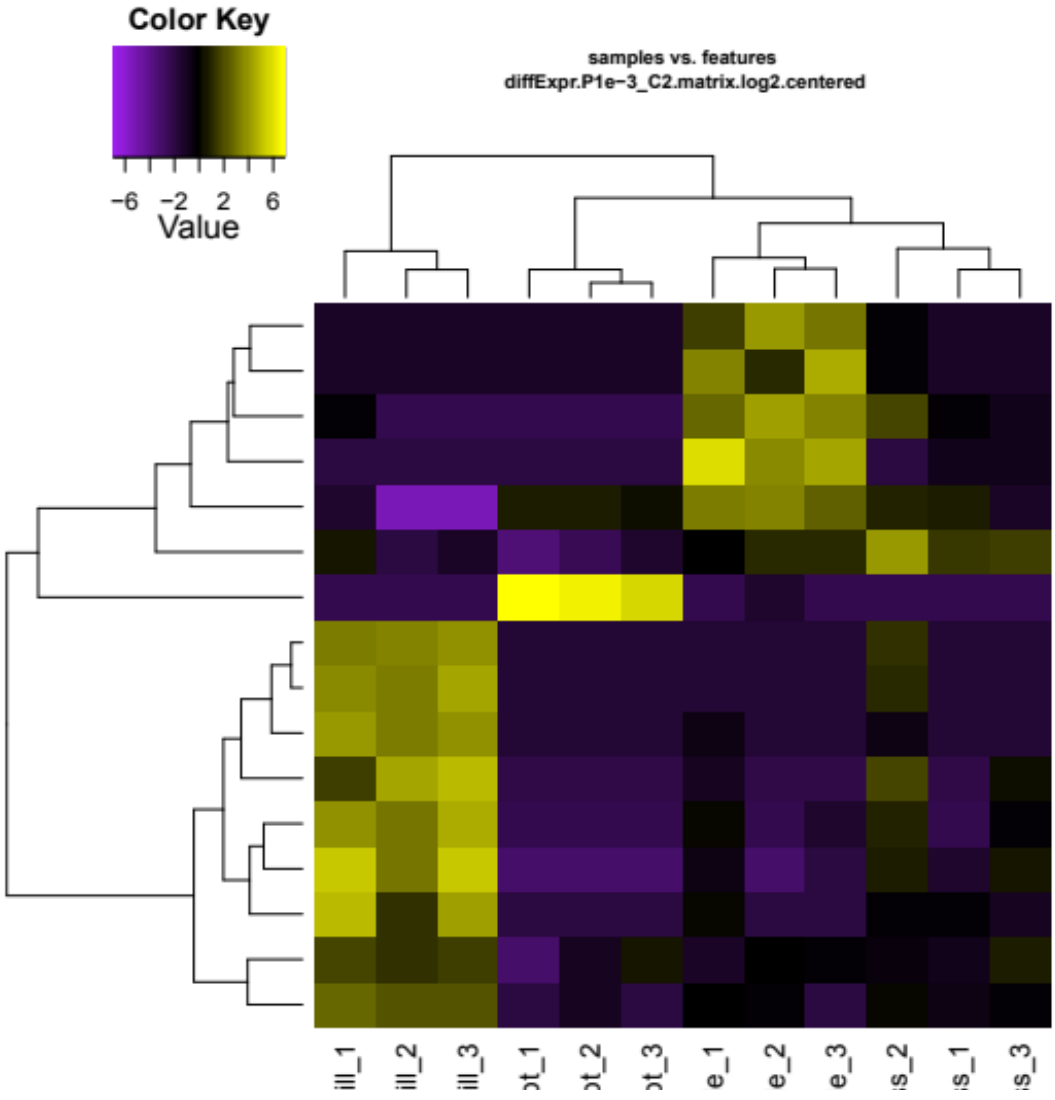



**Figure 9.** Differentially expressed AMPs identified by comparison with DbAMP. Yellow = Upregulated peptides. Purple = Downregulate peptides.

In our research, I closely examined the normalized TPM expression levels of each Differentially Expressed Antimicrobial Peptide (DE AMP). Our findings indicate an inverse relationship with the number of DE AMPs and their expression levels. Specifically, in the foot, where only one AMP is DE, it exhibits high expression when I categorize TPM expression as follows: 0 to 1 (very low), 1 to 10 (low), 10 to 1000 (medium), and above 1000 (high) (Table 1).

**Table 1.** Classes of AMP families identified Differentially Expressed in *L. Orbiculatus.*

| Peptide name | db_AMP_ID | Identity (%) | evalue | AMP Class | Average Expression (TPM) |
|---|---|---|---|---|---|
| **Gills** | | | | | |
| AMP_gill_1 | dbAMP_11074 | 64 | 7.00E-02 | **Buthinin** | **364** |
| AMP_gill_2 | dbAMP_02691 | 46 | 3.91E-04 | rMeuTXK?1 | 166 |
| AMP_gill_3 | dbAMP_02572 | 48 | 4.20E-02 | Defensin MGD-1 | 144 |
| AMP_gill_4 | dbAMP_23507 | 39 | 8.09E-05 | ASABF-related peptide | 132 |
| AMP_gill_5 | dbAMP_11189 | 40 | 7.40E-02 | Psoriasin | 110 |
| AMP_gill_6 | dbAMP_00616 | 42 | 4.00E-03 | Sapecin | 68 |
| AMP_gill_7 | dbAMP_04423 | 39 | 3.96E-04 | Defensin MT3 | 67 |
| AMP_gill_8 | dbAMP_20463 | 40 | 9.10E-02 | Theromacin | 64 |
| **Mantle** | | | | | |
| AMP_mantle_1 | dbAMP_23507 | 40 | 2.70E-02 | **ASABF-related peptide** | **1010** |
| AMP_mantle_2 | dbAMP_15720 | 38 | 4.72E-09 | Kunitz-type serine protease inhibitor PIVL | 229 |
| AMP_mantle_3 | dbAMP_01495 | 36 | 1.00E-01 | Pro-hevein | 106 |
| AMP_mantle_4 | dbAMP_02553 | 41 | 5.10E-02 | Cg-defh2 (Defensin) | 62 |
| AMP_mantle_5 | dbAMP_10274 | 40 | 5.20E-02 | CrustinPm1 | 44 |
| **Foot** | | | | | |
| AMP_foot_1 | dbAMP_08852 | 50 | 2.10E-02 | **Lc-NK-lysin** | **1729** |

In our transcriptome, the second-highest expressed annotated AMP is a putative ASABF-like peptide, which is highly expressed in the mantle. In the gills, I observed medium-expressed AMPs with TPM values ranging from 364 to 64.

In terms of AMP classes, our analysis indicates that they were identified with a low e-value. However, the identity percentage of these AMPs is relatively low, suggesting potential novel properties in our predicted AMPs, this is due to a possible alignment of key domains.

**Draft genome assessment.**

I used QUAST to evaluate the quality of the draft genome assembly and discovered that it was highly complete and contiguous, as indicated by the N50 (1637 contigs) and N50 (379,310 bp) metrics with a GC contant of 36% (Figure 10). I conclude with these findings that Dr. Tanguy's draft genome assembly can be confidently utilized for further research.

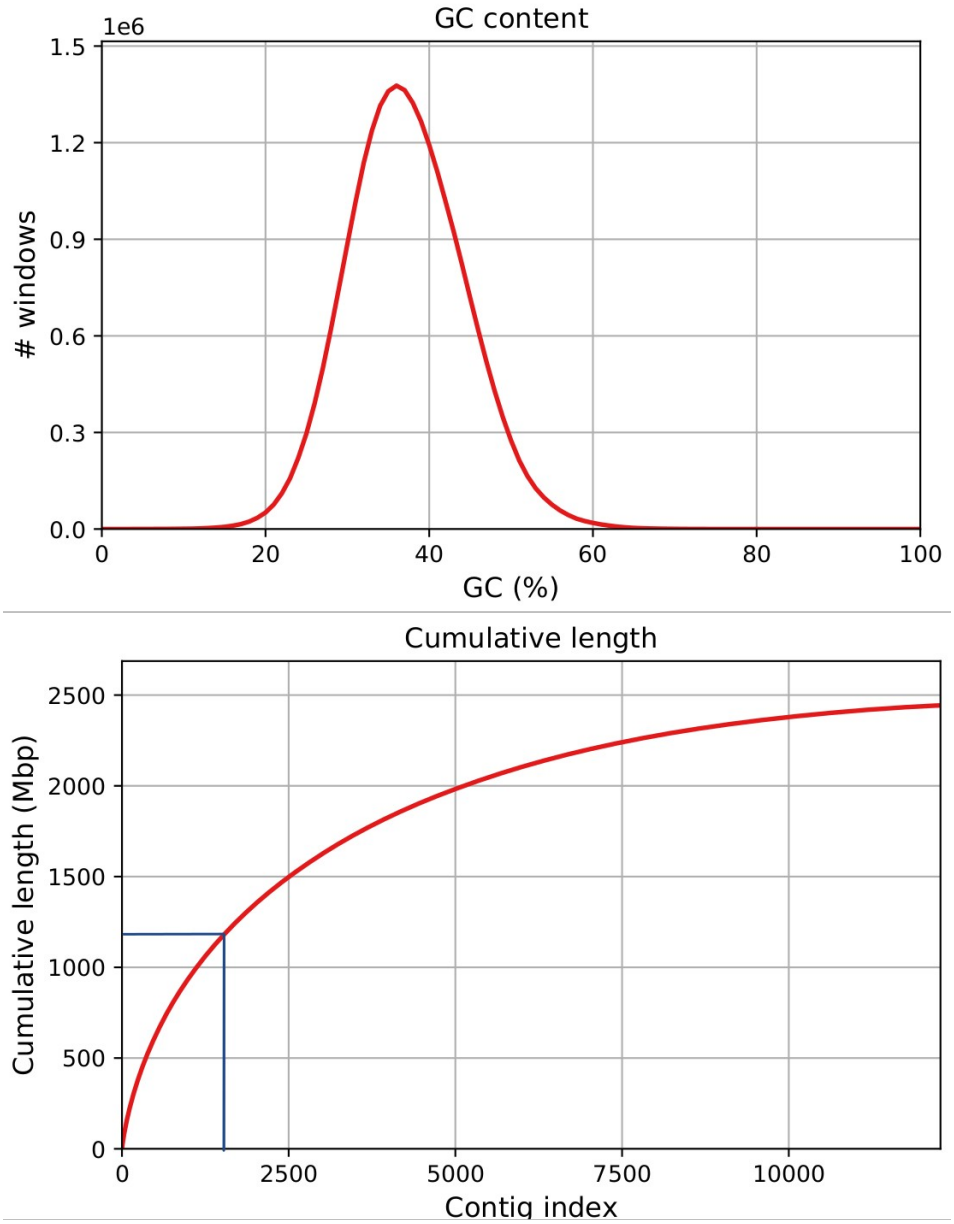



**Figure 10.** Top) GC content of the draft genome of *L. orbiculatus.* Bottom). Cumulative length distribution of contigs. Blue line denotes L50.

During our review of the largest contig, I discovered the complete genome of *Candidatus* Thiodioazotropha lotti, the bacterial symbiont of *Loripes orbiculatus*. Using genome comparison tools implemented in RAST (Aziz et al., 2008), the only differences in gene content between previously assembled MAGs from this species and the complete genome recovered here were mobile elements and hypothetical proteins, found in the complete genome but not in the MAGs (J. Petersen, personal communication).

Using Busco against a metazoan database, I conducted a thorough analysis of the coding completeness of the draft genome. The results demonstrated that 94.8% of the orthologous housekeeping genes were present, with a high duplication ratio of these marker genes. I also discovered that only 3% of marker genes were fragmented, possibly due to the draft genome's remaining fragmentation.

**Genome annotation using RNA evidence.**

Our annotation process allowed us to identify and characterize 252,195 predicted proteins. To assess the efficiency of our annotation, I conducted a completeness assay, which revealed that the annotation reduced 4% the amount of non-orthologous proteins found and increased the proportion of fragmented markers by 3% which overall I considered a successful annotation. The mean coding sequences (CDS) was 1002.35 bp, obtaining a total protein coding-gene density of 10.3 % in our draft genome which is into the upper protein-coding genes in bivalves mollusks (Figure. 11) (Farhat et al., 2022).

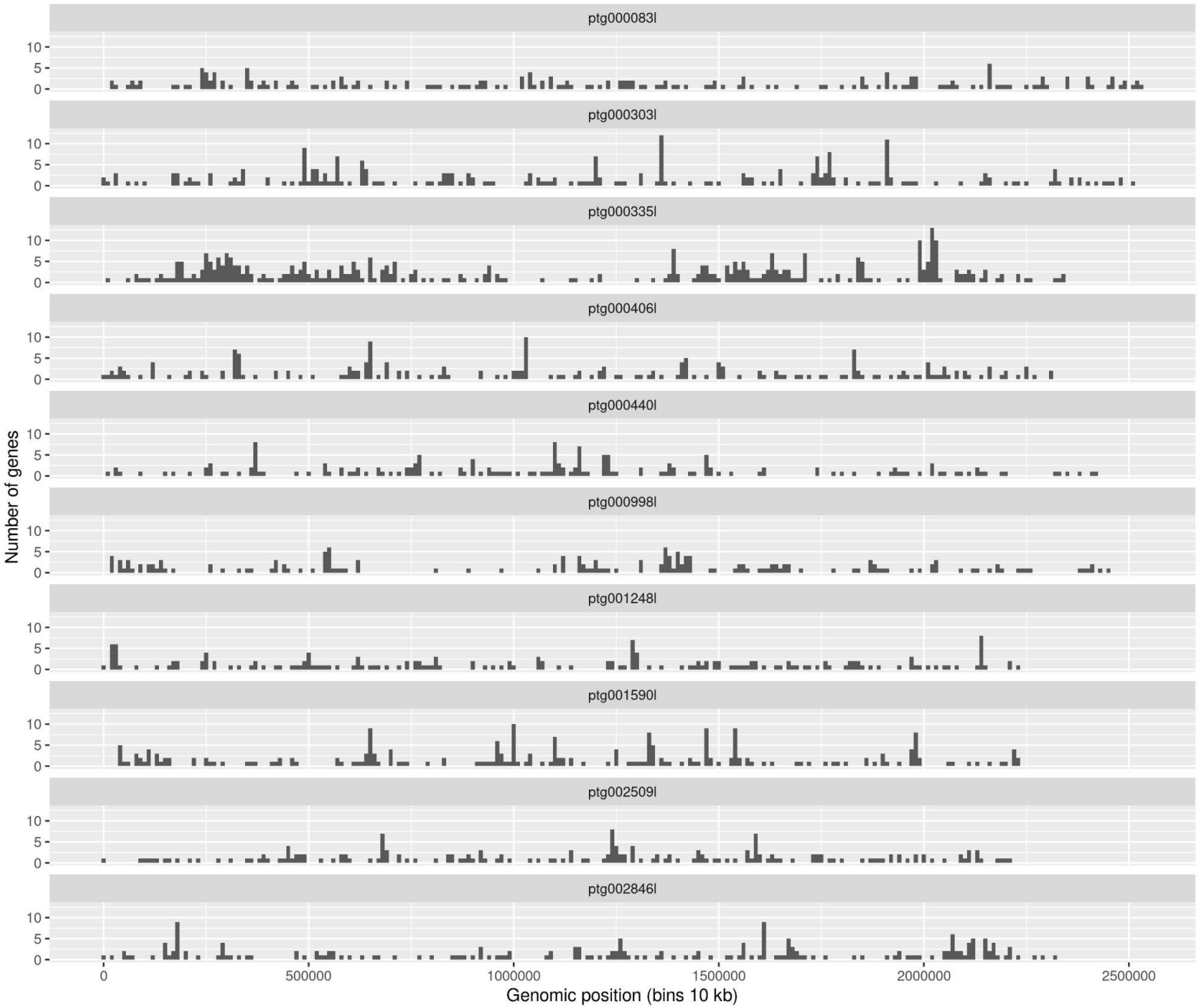



**Figure 11.** Visualization of the 10 largest contigs of the draft genome of *L. orbiculatus*. Gene density is binned by sections of 10 kb

I further explore the amount of predicted small proteins (less than 100 AA) and the number of putative AMPs encoded in the draft genome based on their physicochemical characteristics using AMPIR, as described above for the de novo

transcriptome assembly. I obtained 28,123 small peptides, corresponding to 11% of the genome's predicted proteins. By the physicochemical characterization of these peptides, I concluded that 804 were predicted to have antimicrobial properties if I took a cut-off level of 80% of the score given by AMPIR. From the BLASTp analysis of the total proteins predicted by the genome against DbAmp, I found 162 matches.

**Low efficiency of RNA-seq read mapping to the draft genome of *Loripes orbiculatus.***

To determine the DE expression pattern of these AMPs, I proceeded to map the RNA-seq reads to our annotated draft genome. I found that after using the Star program as our default parameters, I had an average low mapping efficiency of only 56% per library. I tried to improve the mapping efficiency by changing parameters like the minimum mapping length of combined forward and reverse reads (outFilterScoreMinOverLread) and the proportion of mismatches between paired-end reads (peOverlapMMp). However, these changes did not change the amount of unique reads mapping to the genome maintaining a mapping efficiency of 56% per library. Moreover, exploring the reads mapping to the draft genome also revealed that from the ones that mapped uniquely to it, 53% of these reads were aligned to intergenic regions (Figure 12) and configured the reason for which I could not perform a DE expression analysis due to the minimal recommended unique reads per library for performing this kind of analysis is about 10 million reads (Conesa et al., 2016)

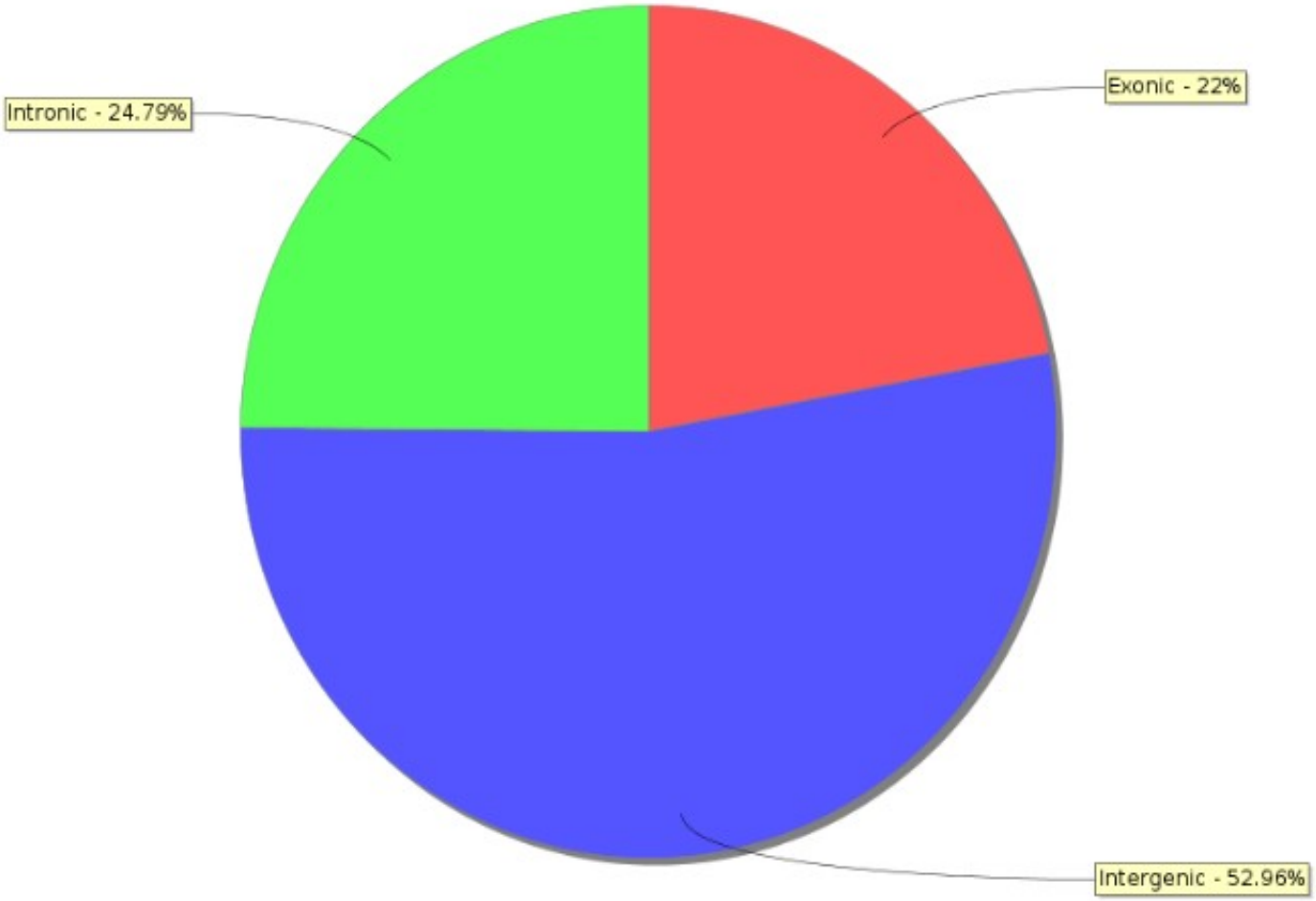



**Figure 12.** Average RNA-seq reads genomic origins per library according to mapping location.

**Comparative analysis between de-novo assembly predicted AMPs and genome annotated AMPs.**

Due to the technical inability to perform a DE analysis based on the genome, I decided to compare the results from the analysis of the putative AMPs that matched DbAMP coming from the de-novo transcriptome assembly and those predicted in the genome. In total, I found 7 AMPs predicted in this study that were shared as the result of the two general pipelines (Table. 2).

**Table 2.** AMPs identified by the comparison of the De-novo transcriptome assembly and the draft genome annotation.

| Genome | Transcriptome peptide name | Identity (%) | evalue | AMP class |
|---|---|---|---|---|
| g239188.t1 | TRINITY_DN597_c13_g1_i2.p1 | 97.778 | 1.7E-61 | rMeuTXK?1 |
| g2065.t1 | TRINITY_DN130_c6_g1_i2.p1 | 97.59 | 1.44E-57 | Sapecin-C |
| g3286.t1 | TRINITY_DN23793_c1_g1_i2.p1 | 97.561 | 9.37E-58 | Defensin MT3 |
| g58514.t1 | TRINITY_DN4873_c1_g1_i1.p1 | 97.5 | 4.49E-54 | ASABF-related peptide |
| g141519.t1 | TRINITY_DN45154_c0_g1_i1.p1 | 96.552 | 5.77E-59 | ASABF |
| g209452.t1 | TRINITY_DN4602_c1_g1_i8.p1 | 95.745 | 2.82E-62 | Theromacin |
| g149187.t1 | TRINITY_DN4524_c0_g1_i2.p1 | 94.444 | 1.88E-59 | Buthinin |
| g113183.t1 | TRINITY_DN597_c13_g1_i2.p1 | 94.444 | 9.09E-58 | rMeuTXK?1 |

**Prospective target validation using molecular biology techniques.**

I utilize the comparative results from the de-novo transcriptome assembly and the genome-annotated peptides to choose the best AMP targets for a prospective study which focus on validation of the bioinformatic results. In detail, I designed with the online tool Primer-blast, under standard parameters, PCR primers for four differentially expressed AMPs (two gills, one foot, and one mantle) and β-Actin as a housekeeping gene to make a positive control (table 3).

**Table 3.** Designed primers for a prospective AMP expression validation assay.

| AMP target (Organ) | lc-nk-lysin primers (foot) | | | | | | |
|---|---|---|---|---|---|---|---|
| Primer name | | **Sequence (5'->3')** | **Tm** | **GC%** | **Self complementarity** | **Self 3' complementarity** | **Product length** |
| nk-lysin_F1 | **Forward primer** | GTGGTGTAGGACTGCAACGA | 59.97 | 55 | 4 | 2 | 71 |
| nk-lysin_R1 | **Reverse primer** | GACAGCCTATAACGCCCCTC | 59.97 | 60 | 4 | 0 | |
| nk-lysin_F2 | **Forward primer** | GGTGAGCGTTTACAGGACCA | 59.97 | 55 | 4 | 2 | 261 |
| nk-lysin_R2 | **Reverse primer** | TCGTTGCAGTCCTACACCAC | 59.97 | 55 | 4 | 1 | |
| AMP target (Organ) | Sapecin (gills) | | | | | | |
| Sapecin_F1 | **Forward primer** | AAAAGATGCGTGCGGAGAGA | 60.04 | 50 | 2 | 0 | 788 |
| Sapecin_R1 | **Reverse primer** | GGTTTGACAAATCGTCGGGC | 60.11 | 55 | 7 | 2 | |
| Sapecin_F2 | **Forward primer** | TGCACTAAAAGATGCGTGCG | 59.83 | 50 | 4 | 2 | 473 |
| Sapecin_R2 | **Reverse primer** | TGTCATATCCGCTGCGTGTT | 60.11 | 50 | 5 | 3 | |
| AMP target (Organ) | Buthinin (Gills) | | | | | | |
| Buthinin_F1 | **Forward primer** | AATTCGGGTCGCAGTGGAAT | 60.04 | 50 | 4 | 2 | 212 |
| Buthinin_R1 | **Reverse primer** | CCAAACTGATGCCGCTTACG | 59.9 | 55 | 3 | 2 | |
| Buthinin_F2 | **Forward primer** | CGTAAGCGGCATCAGTTTGG | 59.9 | 55 | 3 | 0 | 70 |
| Buthinin_R2 | **Reverse primer** | TGGAAGATGTGACTACGGCG | 59.83 | 55 | 3 | 3 | |
| AMP target (Organ) | ASABF-related peptide (mantle) | | | | | | |
| ASABF_F1 | **Forward primer** | AGTGATTGTCCAGCACGGTT | 59.89 | 50 | 5 | 1 | 321 |
| ASABF_R1 | **Reverse primer** | CGACATACTGCTAGAGCGGG | 60.04 | 60 | 4 | 2 | |
| ASABF_F2 | **Forward primer** | ATCGTCGTATTTGGCCTCCC | 59.89 | 55 | 4 | 0 | 152 |
| ASABF_R2 | **Reverse primer** | GCTGTCGCTTTGGGACATTC | 59.83 | 55 | 5 | 2 | |
| AMP target (Organ) | B-actin (Housekeeping gene) | | | | | | |
| b-Actin_F1 | **Forward primer** | AGCCTATTTTGCGTCGGACA | 60.04 | 50 | 5 | 3 | 298 |
| b-Actin_R1 | **Reverse primer** | GCGTCTCTGTCCACTTTCCA | 59.97 | 55 | 2 | 0 | |
| b-Actin_F2 | **Forward primer** | GCGCCTTTGTTTTCGCTACA | 60.04 | 50 | 4 | 2 | 391 |
| b-Actin_R2 | **Reverse primer** | CTTGGCGTCTCTGTCCACTT | 59.97 | 55 | 3 | 0 | |

## Discussion

### De-novo transcriptome assembly reveals a differential AMP expression in *Loripes orbiculatus* tissues.

In the present study, I establish two different in silico methods, the first by comparing the transcriptome data of four different tissues with a de-novo transcriptome assembly approach and the second by annotating a draft genome of *Loripes orbiculatus.* I identified AMPs that might play a role in symbiosis and immune homeostasis. By the first approach, I found 42,282 putative AMPs only by considering the possible physicochemical properties of the predicted peptides. This number of AMPs corresponds probably to an overestimation due to the difficulty of detecting and characterizing them, one of the reasons for their rapid evolution and, thus, change (Hanson et al., 2019). Therefore, I directly compared utilizing a database-assisted system that identifies AMPs and their functional activities. This comparison was achieved through transcriptomic analysis of high-throughput transcriptome data using a curated repository (Jhong et al., 2022) As a result, I could identify 14 distinct peptides with differential expression.

Different studies have previously described a differential usage of AMPs in bivalve tissues. In 2022, Yang et al. used a qPCR analysis targeting specific AMPs described in *Mytilus coruscus* to show that generalistic AMPs, like Arthropod-like defensin-2, Crustin-like or MGD, were constitutively expressed throughout the analyzed tissues. While some were restricted to specific tissues, especially the gonads, they show a marked usage with 5 AMPs upregulated in this tissue. In this study, they could not find a specific AMP related to the gills. More recently, a study determining the response of *Mytilus unguiculatus* to *Vibrio alginolyticus* infection showed an overall

differential expression gene expression in the tissues after pathogen inoculation. However, they only focused on the expression pattern of four well-characterized AMPs (Li et al., 2024).

In *Loripes orbiculatus,* there has only been one study by Yuen et al. in 2019 where they focus on understanding the overall transcriptome profile of four tissues, determining, for example, that despite the host relying on metabolite exchange from the symbiont, in the digestive tract in the visceral mass, I can also identify enzymes related with carbohydrate recognition, digestion, and metabolism. This study focused on two components of the immune system, recognition, and apoptosis, finding differential expression patterns of immune-related genes and upregulating some of these recognition proteins in the gills. Similarly to this study, Yu et al. in 2019, challenged the clam *Meretrix petechialis* to *Vibrio* infection, here as well only characterizing the general immune response of the clam, specifying in the PRRs genes, which could activate downstream AMPs as effector molecules—but overlooking the AMPs as an essential component of the immune systems.

**AMPs discovered in *Loripes orbiculatus* exhibit tissue-specific expression.**

Within the AMPs found in the *L. orbiculatus* gills in this study, I can highlight the antimicrobial capacities of Buthinin and MGD-1 on the In Vitro test that several researchers highly studied; Buthinin initially found in scorpions of the species *Androctonus australis* (Ehret-Sabatier et al., 1996) and MGD-1 *Mytilus galloprovincialis* mussels (Hubert et al., 1996; Mitta et al., 1999) both effective against gram-positive and gram-negative bacteria. Despite ASABF-related peptide being shared in two different tissues, it presented a differential expression; in the gills, the expression averaged 132

TPM, while in the mantle, it was highly expressed at more than one thousand TPM. This class of AMP shows a good response against gram-positive bacteria but has limited effects on gram-negative bacteria, as presented by Zhang et al. in 2000; they tested that seven out of seven gram-positive bacteria were sensitive to ASABF-related-peptide at concentrations between 4 to 130nM in contrast, only 8 out of 14 gram-negative bacteria tested were affected at ASABF concentrations of more than 70nM. Due to the Gram-negative nature of *Candidatus* Thiodiazotropha, the expression of these Gram-specific peptides might indicate an essential role in symbiosis maintenance. In the mantle, ASABF-related peptide is the highest expressed, followed by mediumly expressed AMPs, such as Kunitz-type serine protease inhibitors and Cg-defh2, among others I identify in the same tissue. The two latest mentioned exhibit various antimicrobial activities (Gueguen et al., 2006 and Ranasinghe & McManus, 2013).

Utilizing multiple antimicrobial peptides (AMPs) in tissues may have a synergistic effect, as demonstrated by a study conducted by Yu et al. in 2016 The study analyzed the antimicrobial properties of six different AMPs individually and in combination (up to three AMPs at a time), ultimately concluding that using three AMPs together produced superior results compared to single or two AMP combinations. Comparable to what I see in the gills and mantle of *Loripes orbiculatus,* which could be a way of better controlling the possible pathogens and symbionts. In contrast, our analysis of the foot revealed the presence of only one lc-NK-Lysin-like peptide that exhibited significant upregulated differential expression. This peptide family has been shown to possess potent antimicrobial properties against a wide range of pathogens, characteristic of the members of the saposin-like proteins (Linde et al., 2005). Through our analysis, this

peptide may be crucial in stemming possible pathogens when the foot digs up channels for sulfur acquisition.

**Genomic structural variation drives low RNA-seq reads mapping to the draft genome of *Loripes orbiculatus.***

From hemizygosity to chromosomal inversion, mollusk presents a wide range of genomic structural variantion (SV) that have only recently been explored. In 2020 Gerdol et al., described how *M. galloprovincialis* had approximately 20,000 genes subjected to a presence/absence variation (PAV) controlled mainly by heterozygosity and hemizygosity. Naturally, these PAVs were associated with differential gene expression patterns, especially related to local adaptations, showing, for example, a gene dispensability in mussels caught in the Adriatic Sea compared to the ones in the Atlantic Ocean (Saco et al., 2023). These genomic SV, specifically chromosomal inversion have also been shown to happen in *Pecten maximus* promoted by differential temperature in the settle environment (Hollenbeck et al., 2022). In a preliminary study, Yuen et al unpublished. used 13 protein-coding genes, to understand the phylogenetic relationship between *L. orbiculatus* finding that two distinct clades were formed, one associated with clams (clade “2”) which prefer to lay eggs in the sand without the clear presence of egg masses, whereas the other clade “1” utilized the seagrass bed for leaving visible egg masses. Although there was not a clear physical barrier between these two clades, there was a strong association from the clade “1” coming from sites like Roscoff, and the clade “2” from Elba. To realize the present study, I used RNA-seq reads dissected from clams caught on the coast of Elba (Yuen et al., 2019). The draft genome was kindly shared by a collaborator whose biological station is located in

Roscoff, thus, I hypothesized that a phenomenon like PAV could explain the low-efficiency mapping of the RNA-seq reads to our draft genome. Another feasible explanation related with the SV in mollusks is that it is possible that the low mapping efficiency was due to hemizygosity that could happen even at individual level, as have been demonstrated by Calcino et al., 2021, in which they shown how from molluscs, bivalves are more prone to have this kind of SVs.

**Hypothesized model of cross-communication between *L. orbiculatus* and its symbiont.**

Recently, immunostainings of the bacterial DsrC (Figure. 13), protein being key for the sulfur metabolism in the bacteria, shown a clear dispersal throughout the body of *L. orbiculatus*. Remarkably, a closely related protein RspA (80% similarity), which doesn't have the sulfur binding domain, remains only in the bacteriocytes, suggesting a possible exporting mechanism (Alcaraz et al. *under preparation*). Much research should be done in this field for verifying this hypothesis, here I want to propose a parallel with the *Aeschynomene spp.* and *Bradyrhizobium* system in which is demonstrated the cross-comunication from the bacterial side producing BclA and the plant relesing AMPs, that as an outcome improve the fitness (Guefrachi et al., 2015 and Czernic et al., 2015). *L. orbiculatus,* could be employing AMPs for maintaining the symbiont populations, and *Ca.* Thiodiazotropha could be exchanging its DsrC.

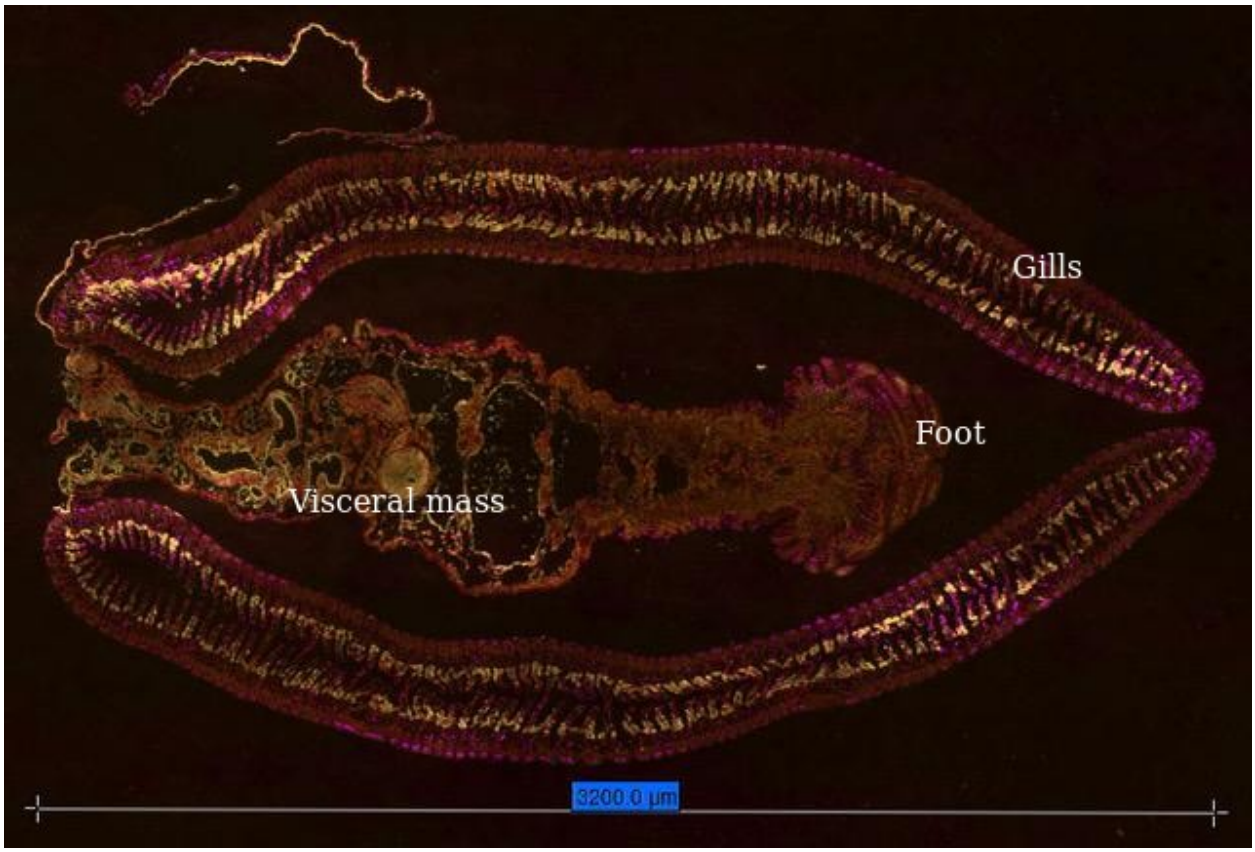



**Figure 13.** Symbiont-encoded protein DsrC throughout the body of *Loripes orbiculatus*. DsrC protein in yellow, DAPI in purple. Courtesy of Cristina Alcaraz.

## Conclusions

To my knowledge, I am presenting the first transcriptomic and genomic analysis of antimicrobial peptides from a symbiotic clam *Loripes orbiculatus*. The study shows how AMPs are differentially expressed in the tissues of *L. orbiculatus* drawing insights into their putative function. Remarkably, I could identify one AMP that might be related with keeping the host from infection when the foot digs up channels for sulfur acquisition. In the gills, I show how an orchestra of AMPs acts possibly mediating the symbiotic relationship. More research should be done to further understand the specific roles of the AMPs in this system.

Despite the low efficiency of mapping, I could annotate the draft genome of *L. orbiculatus* reducing our de-novo transcriptome assembly from 518,140 to 252,195 proteins. Corresponding to the interest of this project, I also narrow the number of predicted AMPs from 42,242 to 804 when only considering the physicochemical properties as the predictor factor. My findings also point to possible genomic structural variants between *L. orbiculatus* populations, that should be addressed in the future. But the first step is to close a reference genome of this species.

## References


Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andrews, S. (2024). *S-andrews/FastQC* [Java]. https://github.com/s-andrews/FastQC (Original work published 2017)

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., … Zagnitko, O. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, *9*(1), 75. https://doi.org/10.1186/1471-2164-9-75

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, *3*(1), lqaa108. https://doi.org/10.1093/nargab/lqaa108

Cardini, U., Bartoli, M., Lücker, S., Mooshammer, M., Polzin, J., Lee, R. W., Micić, V., Hofmann, T., Weber, M., & Petersen, J. M. (2019). Chemosymbiotic bivalves contribute to the nitrogen budget of seagrass ecosystems. *The ISME Journal*, *13*(12), Article 12. https://doi.org/10.1038/s41396-019-0486-9

Cavanaugh, C. M., Gardiner, S. L., Jones, M. L., Jannasch, H. W., & Waterbury, J. B. (1981). Prokaryotic Cells in the Hydrothermal Vent Tube Worm Riftia pachyptila Jones: Possible Chemoautotrophic Symbionts. *Science*, *213*(4505), 340–342. https://doi.org/10.1126/science.213.4505.340

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*(1), 13. https://doi.org/10.1186/s13059-016-0881-8

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Farhat, S., Bonnivard, E., Pales Espinosa, E., Tanguy, A., Boutet, I., Guiglielmoni, N., Flot, J.-F., & Allam, B. (2022). Comparative analysis of the Mercenaria mercenaria genome provides insights into the diversity of transposable elements and immune molecules in bivalve mollusks. *BMC Genomics*, *23*(1), 192. https://doi.org/10.1186/s12864-021-08262-1

Fingerhut, L. C. H. W., Miller, D. J., Strugnell, J. M., Daly, N. L., & Cooke, I. R. (2021). ampir: An R package for fast genome-wide prediction of antimicrobial peptides. *Bioinformatics*, *36*(21), 5262–5263. https://doi.org/10.1093/bioinformatics/btaa653

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T. F., & Conesa, A. (2012). Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics*, *28*(20), 2678–2679. https://doi.org/10.1093/bioinformatics/bts503

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Jhong, J.-H., Yao, L., Pang, Y., Li, Z., Chung, C.-R., Wang, R., Li, S., Li, W., Luo, M., Ma, R., Huang, Y., Zhu, X., Zhang, J., Feng, H., Cheng, Q., Wang, C., Xi, K., Wu, L.-C., Chang, T.-H., … Lee, T.-Y. (2022). dbAMP 2.0: Updated resource for antimicrobial peptides with an enhanced scanning method for genomic and proteomic data. *Nucleic Acids Research*, *50*(D1), D460–D470. https://doi.org/10.1093/nar/gkab1080

Katainen, R., Donner, I., Cajuso, T., Kaasinen, E., Palin, K., Mäkinen, V., Aaltonen, L. A., & Pitkänen, E. (2018). Discovery of potential causative mutations in human coding and noncoding genome with the interactive software BasePlayer. *Nature Protocols*, *13*(11), 2580–2600. https://doi.org/10.1038/s41596-018-0052-3

König, S., Gros, O., Heiden, S. E., Hinzke, T., Thürmer, A., Poehlein, A., Meyer, S., Vatin, M., Mbéguié-A-Mbéguié, D., Tocny, J., Ponnudurai, R., Daniel, R., Becher, D., Schweder, T., & Markert, S. (2016). Nitrogen fixation in a chemoautotrophic lucinid symbiosis. *Nature Microbiology*, *2*(1), Article 1. https://doi.org/10.1038/nmicrobiol.2016.193

Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, *28*(24), 3211–3217. https://doi.org/10.1093/bioinformatics/bts611

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 323. https://doi.org/10.1186/1471-2105-12-323

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Ohbayashi, T., Mergaert, P., & Kikuchi, Y. (2020). Host-symbiont specificity in insects: Underpinning mechanisms and evolution. In *Advances in Insect Physiology* (Vol. 58, pp. 27–62). Elsevier. https://doi.org/10.1016/bs.aiip.2020.03.002

Petersen, J. M., Kemper, A., Gruber-Vodicka, H., Cardini, U., van der Geest, M., Kleiner, M., Bulgheresi, S., Mußmann, M., Herbold, C., Seah, B. K. B., Antony, C. P., Liu, D., Belitz, A., & Weber, M. (2016). Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. *Nature Microbiology*, *2*(1), Article 1. https://doi.org/10.1038/nmicrobiol.2016.195

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Song, L., & Florea, L. (2015). Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience*, *4*(1), 48. https://doi.org/10.1186/s13742-015-0089-y

Taylor, J. D., & Glover, E. (2021). *Biology, evolution and generic review of the chemosymbiotic bivalve family Lucinidae*. Ray Society.

Allam, B., & Raftos, D. (2015). Immune responses to infectious diseases in bivalves. *Journal of Invertebrate Pathology*, *131*, 121–136. https://doi.org/10.1016/j.jip.2015.05.005

Bouallegui, Y. (2019). Immunity in mussels: An overview of molecular components and mechanisms with a focus on the functional defenses. *Fish & Shellfish Immunology*, *89*, 158–169. https://doi.org/10.1016/j.fsi.2019.03.057

Canesi, L., Auguste, M., Balbi, T., & Prochazkova, P. (2022). Soluble mediators of innate immunity in annelids and bivalve mollusks: A mini-review. *Frontiers in Immunology*, *13*. https://www.frontiersin.org/articles/10.3389/fimmu.2022.1051155

Czernic, P., Gully, D., Cartieaux, F., Moulin, L., Guefrachi, I., Patrel, D., Pierre, O., Fardoux, J., Chaintreuil, C., Nguyen, P., Gressent, F., Da Silva, C., Poulain, J., Wincker, P., Rofidal, V., Hem, S., Barrière, Q., Arrighi, J.-F., Mergaert, P., & Giraud, E. (2015). Convergent Evolution of Endosymbiont Differentiation in Dalbergioid and Inverted Repeat-Lacking Clade Legumes Mediated by Nodule-Specific Cysteine-Rich Peptides1. *Plant Physiology*, *169*(2), 1254–1265. https://doi.org/10.1104/pp.15.00584

Di Somma, A., Moretta, A., Canè, C., Cirillo, A., & Duilio, A. (2020). Antimicrobial and Antibiofilm Peptides. *Biomolecules*, *10*(4), Article 4. https://doi.org/10.3390/biom10040652

Dubilier, N., Bergin, C., & Lott, C. (2008). Symbiotic diversity in marine animals: The art of harnessing chemosynthesis. *Nature Reviews Microbiology*, *6*(10), Article 10. https://doi.org/10.1038/nrmicro1992

Ehret-Sabatier, L., Loew, D., Goyffon, M., Fehlbaum, P., Hoffmann, J. A., van Dorsselaer, A., & Bulet, P. (1996). Characterization of novel cysteine-rich antimicrobial peptides from scorpion blood. *The Journal of Biological Chemistry*, *271*(47), 29537–29544. https://doi.org/10.1074/jbc.271.47.29537

Freiberg, C., Fellay, R., Bairoch, A., Broughton, W. J., Rosenthal, A., & Perret, X. (1997). Molecular basis of symbiosis between Rhizobium and legumes. *Nature*, *387*(6631), Article 6631. https://doi.org/10.1038/387394a0

Gerdol, M., Moreira, R., Cruz, F., Gómez-Garrido, J., Vlasova, A., Rosani, U., Venier, P., Naranjo-Ortiz, M. A., Murgarella, M., Greco, S., Balseiro, P., Corvelo, A., Frias, L., Gut, M., Gabaldón, T., Pallavicini, A., Canchaya, C., Novoa, B., Alioto, T. S., … Figueras, A. (2020). Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biology*, *21*(1), 275. https://doi.org/10.1186/s13059-020-02180-3

Gilbert, S. F., Bosch, T. C. G., & Ledón-Rettig, C. (2015). Eco-Evo-Devo: Developmental symbiosis and developmental plasticity as evolutionary agents. *Nature Reviews Genetics*, *16*(10), Article 10. https://doi.org/10.1038/nrg3982

Guefrachi, I., Pierre, O., Timchenko, T., Alunni, B., Barrière, Q., Czernic, P., Villaécija-Aguilar, J.-A., Verly, C., Bourge, M., Fardoux, J., Mars, M., Kondorosi, E., Giraud, E., & Mergaert, P. (2015). Bradyrhizobium BclA Is a Peptide Transporter Required for Bacterial Differentiation in Symbiosis with Aeschynomene Legumes. *Molecular Plant-Microbe Interactions: MPMI*, *28*(11), 1155–1166. https://doi.org/10.1094/MPMI-04-15-0094-R

Gueguen, Y., Herpin, A., Aumelas, A., Garnier, J., Fievet, J., Escoubas, J.-M., Bulet, P., Gonzalez, M., Lelong, C., Favrel, P., & Bachère, E. (2006). Characterization of a defensin from the oyster Crassostrea gigas. Recombinant production, folding, solution structure, antimicrobial activities, and gene expression. *The Journal of Biological Chemistry*, *281*(1), 313–323. https://doi.org/10.1074/jbc.M510850200

Hanson, M. A., Lemaitre, B., & Unckless, R. L. (2019). Dynamic Evolution of Antimicrobial Peptides Underscores Trade-Offs Between Immunity and Ecological Fitness. *Frontiers in Immunology*, *10*. https://www.frontiersin.org/articles/10.3389/fimmu.2019.02620

Huan, Y., Kong, Q., Mou, H., & Yi, H. (2020). Antimicrobial Peptides: Classification, Design, Application and Research Progress in Multiple Fields. *Frontiers in Microbiology*, *11*. https://www.frontiersin.org/articles/10.3389/fmicb.2020.582779

Hubert, F., Noel, T., & Roch, P. (1996). A member of the arthropod defensin family from edible Mediterranean mussels (Mytilus galloprovincialis). *European Journal of Biochemistry*, *240*(1), 302–306. https://doi.org/10.1111/j.1432-1033.1996.0302h.x

Jhong, J.-H., Yao, L., Pang, Y., Li, Z., Chung, C.-R., Wang, R., Li, S., Li, W., Luo, M., Ma, R., Huang, Y., Zhu, X., Zhang, J., Feng, H., Cheng, Q., Wang, C., Xi, K., Wu, L.-C., Chang, T.-H., … Lee, T.-Y. (2022). dbAMP 2.0: Updated resource for antimicrobial peptides with an enhanced scanning method for genomic and proteomic data. *Nucleic Acids Research*, *50*(D1), D460–D470. https://doi.org/10.1093/nar/gkab1080

Li, H., Zhao, J., Li, Y., Dong, Z., Lin, S., Guo, B., & Qi, P. (2024). Transcriptome analysis reveals tissue-specific responses of Mytilus unguiculatus to Vibrio alginolyticus infection. *Fish & Shellfish Immunology*, *144*, 109301. https://doi.org/10.1016/j.fsi.2023.109301

Linde, C. M. A., Grundström, S., Nordling, E., Refai, E., Brennan, P. J., & Andersson, M. (2005). Conserved Structure and Function in the Granulysin and NK-Lysin Peptide Family. *Infection and Immunity*, *73*(10), 6332–6339. https://doi.org/10.1128/IAI.73.10.6332-6339.2005

Login, F. H., Balmand, S., Vallier, A., Vincent-Monégat, C., Vigneron, A., Weiss-Gayet, M., Rochat, D., & Heddi, A. (2011). Antimicrobial Peptides Keep Insect Endosymbionts Under Control. *Science*, *334*(6054), 362–365. https://doi.org/10.1126/science.1209728

Marshall, J. S., Warrington, R., Watson, W., & Kim, H. L. (2018). An introduction to immunology and immunopathology. *Allergy, Asthma & Clinical Immunology*, *14*(2), 49. https://doi.org/10.1186/s13223-018-0278-1

Mergaert, P. (2018). Role of antimicrobial peptides in controlling symbiotic bacterial populations. *Natural Product Reports*, *35*(4), 336–356. https://doi.org/10.1039/C7NP00056A

Mitta, G., Vandenbulcke, F., Hubert, F., & Roch, P. (1999). Mussel defensins are synthesised and processed in granulocytes then released into the plasma after bacterial challenge. *Journal of Cell Science*, *112 ( Pt 23)*, 4233–4242. https://doi.org/10.1242/jcs.112.23.4233

Nyholm, S. V., & McFall-Ngai, M. (2004). The winnowing: Establishing the squid–vibrio symbiosis. *Nature Reviews Microbiology*, *2*(8), Article 8. https://doi.org/10.1038/nrmicro957

Orr, M. W., Mao, Y., Storz, G., & Qian, S.-B. (2020). Alternative ORFs and small ORFs: Shedding light on the dark proteome. *Nucleic Acids Research*, *48*(3), 1029–1042. https://doi.org/10.1093/nar/gkz734

Ranasinghe, S., & McManus, D. P. (2013). Structure and function of invertebrate Kunitz serine protease inhibitors. *Developmental and Comparative Immunology*, *39*(3), 219–227. https://doi.org/10.1016/j.dci.2012.10.005

Roch, P. (1999). Defense mechanisms and disease prevention in farmed marine invertebrates. *Aquaculture*, *172*(1), 125–145. https://doi.org/10.1016/S0044-8486(98)00439-6

Saco, A., Rey-Campos, M., Gallardo-Escárate, C., Gerdol, M., Novoa, B., & Figueras, A. (2023). Gene presence/absence variation in Mytilus galloprovincialis and its implications in gene expression and adaptation. *iScience*, *26*(10), 107827. https://doi.org/10.1016/j.isci.2023.107827

Shigenobu, S., & Stern, D. L. (2013). Aphids evolved novel secreted proteins for symbiosis with bacterial endosymbiont. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1750), 20121952. https://doi.org/10.1098/rspb.2012.1952

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., & Ishikawa, H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. *Nature*, *407*(6800), Article 6800. https://doi.org/10.1038/35024074

Sogin, E. M., Kleiner, M., Borowski, C., Gruber-Vodicka, H. R., & Dubilier, N. (2021). Life in the Dark: Phylogenetic and Physiological Diversity of Chemosynthetic Symbioses. *Annual Review of Microbiology*, *75*, 695–718. https://doi.org/10.1146/annurev-micro-051021-123130

Tiscar, P. G., & Mosca, F. (2004). Defense mechanisms in farmed marine molluscs. *Veterinary Research Communications*, *28 Suppl 1*, 57–62. https://doi.org/10.1023/b:verc.0000045379.78547.23

Wein, T., Picazo, D. R., Blow, F., Woehle, C., Jami, E., Reusch, T. B. H., Martin, W. F., & Dagan, T. (2019). Currency, Exchange, and Inheritance in the Evolution of Symbiosis. *Trends in Microbiology*, *27*(10), 836–849. https://doi.org/10.1016/j.tim.2019.05.010

Yadavalli, S. S., & Yuan, J. (2022). Bacterial Small Membrane Proteins: The Swiss Army Knife of Regulators at the Lipid Bilayer. *Journal of Bacteriology*, *204*(1), e0034421. https://doi.org/10.1128/JB.00344-21

Yang, J., He, J., Liu, L., He, M., Zhang, X., Buttino, I., Guo, B., Yan, X., & Liao, Z. (2022). Expression profiles of antimicrobial peptides in Mytilus coruscus. *Aquaculture*, *548*, 737709. https://doi.org/10.1016/j.aquaculture.2021.737709

Yu, G., Baeder, D. Y., Regoes, R. R., & Rolff, J. (2016). Combination Effects of Antimicrobial Peptides. *Antimicrobial Agents and Chemotherapy*, *60*(3), 1717–1724. https://doi.org/10.1128/AAC.02434-15

Yu, J., Wang, H., Yue, X., & Liu, B. (2019). Dynamic immune and metabolism response of clam Meretrix petechialis to Vibrio challenge revealed by a time series of transcriptome analysis. *Fish & Shellfish Immunology*, *94*, 17–26. https://doi.org/10.1016/j.fsi.2019.08.057

Yuen, B., Polzin, J., & Petersen, J. M. (2019). Organ transcriptomes of the lucinid clam Loripes orbiculatus (Poli, 1791) provide insights into their specialised roles in the biology of a chemosymbiotic bivalve. *BMC Genomics*, *20*(1), 820. https://doi.org/10.1186/s12864-019-6177-0

Zannella, C., Mosca, F., Mariani, F., Franci, G., Folliero, V., Galdiero, M., Tiscar, P. G., & Galdiero, M. (2017). Microbial Diseases of Bivalve Mollusks: Infections, Immunology and Antimicrobial Defense. *Marine Drugs*, *15*(6), Article 6. https://doi.org/10.3390/md15060182

Zauner, S., Vogel, M., Polzin, J., Yuen, B., Mußmann, M., El-Hacen, E.-H. M., & Petersen, J. M. (2022). Microbial communities in developmental stages of lucinid bivalves. *ISME Communications*, *2*(1), Article 1. https://doi.org/10.1038/s43705-022-00133-4

Zhang, H., Yoshida, S., Aizawa, T., Murakami, R., Suzuki, M., Koganezawa, N., Matsuura, A., Miyazawa, M., Kawano, K., Nitta, K., & Kato, Y. (2000). In vitro antimicrobial properties of recombinant ASABF, an antimicrobial peptide isolated from the nematode Ascaris suum. *Antimicrobial Agents and Chemotherapy*, *44*(10), 2701–2705. https://doi.org/10.1128/AAC.44.10.2701-2705.2000