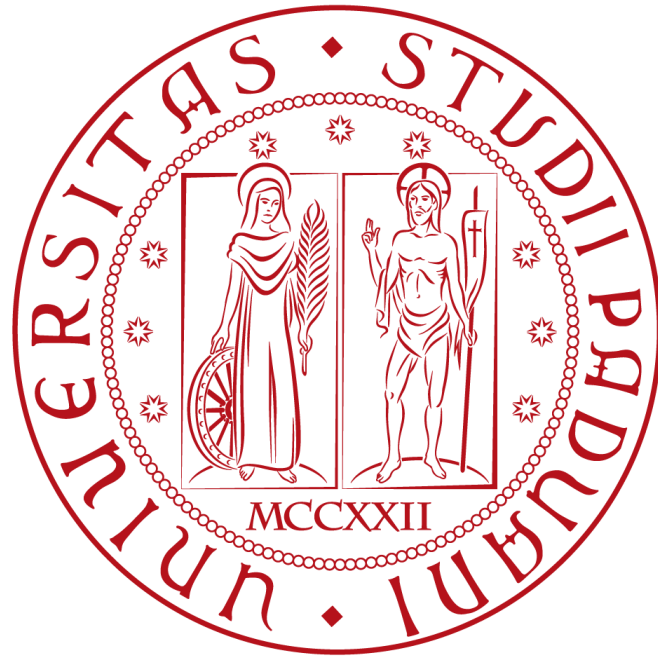


UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA
CORSO DI LAUREA IN BIOTECNOLOGIE



ELABORATO DI LAUREA
IDENTIFICAZIONE DI PEPTIDI ANTIMICROBICI
DI MITILO TRAMITE SEQUENZIAMENTO
NANOPORE

Tutor: Professore Rosani Umberto

Dipartimento: Biologia

Università degli studi di Padova

Laureanda: Elisa Geronazzo

Matricola 1216592

Anno Accademico 2022/2023

ANNO ACCADEMICO 2022/2023

Sommario

ABSTRACT.....	6
1 INTRODUZIONE	7
2 APPROCCIO SPERIMENTALE.....	9
2.1 OTTENIMENTO DEI CAMPIONI.....	9
2.2 COSTRUZIONE DELLA LIBRERIA DI RNA E SEQUENZIAMENTO NANOPORE	9
2.3 IDENTIFICAZIONE DELLE SEQUENZE	10
2.4 ANALISI BIOINFORMATICA DELLE SEQUENZE	10
3 RISULTATI E DISCUSSIONE.....	11
3.1 ESTRAZIONE DELL'RNA E SEQUENZIAMENTO NANOPORE	11
3.2 IDENTIFICAZIONE E PROCESSAMENTO DELLE SEQUENZE	11
3.3 ANALISI BIOINFORMATICA.....	14
3.4 OSSERVAZIONI CONCLUSIVE	18
4 BIBLIOGRAFIA.....	20

ABSTRACT

In questo studio è stato effettuato il sequenziamento del trascrittoma di un campione di emociti ottenuto da *Mytilus galloprovincialis* e sono state analizzate le sequenze ottenute con vari tool bioinformatici.

Tramite *BLAST* è stato possibile identificare 5.230 trascritti noti, e con un confronto tra questi ed il genoma di *M. galloprovincialis* il 75% dei trascritti è stato mappato sul genoma. Sono stati predetti 3.482 geni, di cui 3.328 già annotati e 154 possibili nuovi geni.

Utilizzando *prodigal*, sono state identificate 11.874 sequenze codificanti che sono state annotate identificando i domini proteici conservati.

Sono stati visualizzati i vari domini identificati e la loro frequenza in vari dataset: mappati sul genoma, presenti nelle sequenze proteiche predette e presenti nelle sequenze predette che presentano anche il peptide segnale. In particolare, confrontando le sequenze predette destinate alla secrezione extracellulare con database proteici tramite *BLAST* risultano 214 hit, 78 delle quali con funzione nota. Di queste 78, solo 10 sono state effettivamente riconosciute come AMP. Lo studio presenta anche una panoramica sui dati ottenuti dal sequenziamento di RNA diretto applicato ad un organismo non modello quale il mitilo mediterraneo.

1 INTRODUZIONE

I peptidi antimicrobici (AMP) sono delle corte sequenze peptidiche coinvolte nella difesa immunitaria innata [1]. Nonostante un'ampia variabilità di dimensione, sequenza, modalità di azione e struttura, presentano delle caratteristiche comuni come una carica cationica, un'importante presenza di cisteine e ponti disolfurici e la capacità di promuovere le interazioni peptide-membrana sfruttando le differenze fondamentali delle membrane cellulari procariotiche dalle eucariotiche, in modo da generare un'azione specifica contro i patogeni senza danneggiare l'ospite [1,2].

Gli AMP di specie marine presentano forti proprietà idrofobiche e struttura anfipatica specifica che ne rafforzano l'azione antimicrobica. Tra questi, gli AMP di mitilo sono stati i tra i primi ad essere studiati [1]. Ad oggi sono stati divisi in quattro categorie principali: defensine, mitiline, miticine e mitimicine [1,3], nonostante altre famiglie di AMP siano state identificate.

Tra gli AMP più studiati, le defensine sono molto diffuse in tutti gli organismi viventi, in mitilo sono presenti due isoforme che codificano per peptidi con un'alta omologia di sequenza tra loro: *Mytilus galloprovincialis* defensin 1 (MGD1) e MGD2. Entrambe presentano quattro ponti disolfuro, invece che tre come nella maggior parte delle altre molecole analoghe, e quello che è stato chiamato γ core motif, che invece è altamente conservato in tutti gli organismi viventi [2]. Come azione, le MGD dimostrano una certa efficacia contro i batteri gram-positivi e in più MGD2 ha anche effetto sui batteri gram-negativi.

la seconda categoria, le mitiline, sono presenti in cinque isoforme: A, B, C, D e G1. Le mitiline sono hanno effetto sia sui batteri gram-positivi che gram-negativi, a parte l'isoforma G1 che invece agisce solo sui batteri gram negativi. Inoltre, tramite vari esperimenti, si è potuto accertare che ad alte concentrazioni le isoforme D e G1 necessitavano di varie ore per agire, mentre alle isoforme A e B sono bastati pochi minuti, con B a dimostrare l'efficacia maggiore.

Le miticine presentano tre isoforme A, B e C di 40 aa con effetto antibatterico contro i gram-positivi e alcune volte i gram-negativi, e l'isoforma C è anche un potente antivirale. Questa classe di AMP è ricca di cisteine e presenta quattro ponti disolfuro.

La mitimicina (MytM) è un peptide strettamente antifungino che presenta 12 cisteine nella sequenza, rilevata inizialmente da *M. edulis*. Una molecola analoga chiamata miticousin-1 di 104 aa, di cui 10 cisteine, è stata scoperta invece nella specie *M. coruscus* e invece dimostra efficacia antibatterica contro i gram-positivi piuttosto che gram-negativi e funghi [1,3].

Le varie caratteristiche fondamentali degli AMP ne fanno delle molecole con un alto potenziale biotecnologico. In particolare, l'efficienza che dimostrano a basse

concentrazioni e il fatto che il meccanismo di azione dipenda dalle caratteristiche base della membrana cellulare procariotica (rendendo la possibilità di sviluppare resistenza molto bassa in quanto i batteri dovrebbero modificare fattori base della loro morfologia) le rende ottimi candidati per sostituire l'uso di antibiotici [4]. Le conoscenze sugli AMP sono ancora limitate, e sono quindi necessari studi più approfonditi dal punto di vista genetico, trascrittomico e del meccanismo di azione effettivo di queste molecole.

Comunemente lo studio del trascrittoma si basa sul sequenziamento dell'RNA dopo la sua conversione in cDNA e la sua frammentazione per costruire librerie di sequenziamento adatte alla tecnologia illumina. Questo, inevitabilmente, comporta una perdita di informazioni importanti, quali la complessità della sequenza completa e le modificazioni di sequenza apportate a livello di RNA [5]. Nel 2017 è stata pubblicata una nuova tecnologia in grado di sequenziare sequenze complete senza dover ricorrere a PCR o frammentazione. La Oxford Nanopore technologies (ONT, Oxford, UK) ha introdotto un nuovo protocollo: Direct RNA sequencing (dRNA-seq) per il sequenziamento di RNA nativo [5]. Se effettivamente l'utilizzo del metodo dRNA-seq per identificare e analizzare le diverse AMP e le loro isoforme dimostrasse una buona efficacia ed accuratezza questo potrebbe portare ad un importante avanzamento nelle conoscenze dell'organizzazione genica e potrebbe permettere di capire meglio il meccanismo di evoluzione di questi peptidi [5,6]. Inoltre, una migliore conoscenza di base permetterebbe di aprire la via ad utilizzi biotecnologici, industriali e farmaceutici di queste molecole.

2 APPROCCIO SPERIMENTALE

2.1 OTTENIMENTO DEI CAMPIONI

I campioni analizzati in questa tesi consistono in emolinfa collezionata dal muscolo adduttore di 3 esemplari di *Mytilus galloprovincialis* adulti in condizioni *naïve*, prelevati in ambiente di laguna a Chioggia (Italia) nel luglio 2022.

I campioni, rispettivamente 1H, 2H e 3H, sono stati centrifugati a 800 rpm e dopo aver rimosso il surnatante gli è stato aggiunto 1 ml di Trizol (Thermo Fisher Scientific, Waltham, Massachusetts, USA) e sono stati conservati a -80 °C.

L'RNA è stato estratto utilizzando le colonnine RNeasy MinElute Cleanup kit (Qiagen, Hiden, Germany) e quantificato tramite QuBit (Thermo Fisher Scientific). Per le seguenti analisi è stato utilizzato il campione 3H.

2.2 COSTRUZIONE DELLA LIBRERIA DI RNA E SEQUENZIAMENTO NANOPORE

La preparazione della libreria di RNA è stata effettuata con il kit Nanopore dRNA (SQK-RNA002, ONT, Oxford, UK).

Partendo dal campione di RNA purificato 3H, l'adattatore poly(T) è stato legato all'RNA poliadenilato incubandolo a temperatura ambiente per 15 minuti con T4 DNA ligase (New England Biolabs) e il buffer Quick Ligase reaction (New England Biolabs).

In seguito, si sono aggiunti alla provetta i dNTP e la trascrittasi inversa SuperScript III (Thermo Fisher Scientific, Waltham, Massachusetts, USA) e incubato il campione nel termociclatore (Eppendorf, Amburgo, Germania), a 50°C per 50 minuti e 70°C per 10 minuti, così da sintetizzare il filamento cDNA complementare all'RNA.

Al termine del ciclo con le biglie magnetiche Agentcount RNACleans XP (Beckman Coulter, Brea, California, USA) si sono rimossi i reagenti del passaggio precedente e si è recuperato l'ibrido RNA-cDNA in acqua nuclease-free per passare allo step successivo.

Aggiungendo l'RNA Adapter RNAX e la ligase T4 DNA il campione è stato incubato nell'HULA Mixer per 5 minuti a temperatura ambiente lasciando agire gli enzimi. Utilizzando nuovamente le biglie Agentcount RNACleans XP si sono eliminati i reagenti in eccesso e si è recuperata la libreria di RNA pronta da sequenziare in 21µl di Elution Buffer EB, di cui 1 µl è stato utilizzato per la quantificazione con Qubit.

Caricando i restanti 20 µl rimanenti nella flow cell del sequenziatore MinION (ONT, Oxford, UK) secondo le istruzioni fornite dal produttore, si è sequenziata la libreria generando circa 1 milione di reads in formato FAST5.

Processando i file FAST5 ottenuti da MinION Nanopore con il software *guppy* v.1.2.0 ([Github](#)) basecaller si sono ottenuti file FASTQ che sono stati in seguito analizzati con il software *Rattle* v.1.0 ([Github](#)).

2.3 IDENTIFICAZIONE DELLE SEQUENZE

Una volta ottenute, le sequenze del trascrittoma sono state elaborate con il software *prodigal* v.2.6.0 ([Github](#)) per la predizione delle sequenze codificanti. Quest'ultime sono state analizzate con *pfam* v.35.0 per l'identificazione di vari domini e regioni conosciuti e la frequenza con cui si presentano nelle reads. Essendo il target del nostro studio i AMP, sono state isolate tutte le sequenze che presentano la regione del peptide segnale, identificato utilizzando *SignalP* v.6 ([DTU Health Tech](#)), che identifica le proteine destinate alla secrezione.

2.4 ANALISI BIOINFORMATICA DELLE SEQUENZE ED ELABORAZIONE DEI DATI

Tramite *BLAST*, le sequenze sono state mappate sul genoma di *M. galloprovincialis* e i domini trovati tramite *pfam*. Si sono poi messi a confronto i domini presenti e la loro frequenza con quelli mappati sul genoma, identificati nel trascrittoma e nelle proteine col peptide segnale.

I dati sono stati elaborati e visualizzati utilizzando i pacchetti *dplyr*, *ggplot2*, *magrittr* e *ggpubr* di R v. 4.3.0.

3 RISULTATI E DISCUSSIONE

3.1 ESTRAZIONE DELL'RNA E SEQUENZIAMENTO NANOPORE

In questa tesi verranno esposti i risultati del sequenziamento e l'analisi delle reads ottenute dal campione 3H, riassunti nella tabella 1.

Tabella 1. Quantificazioni dei campioni di *M. galloprovincialis* raccolti. Per ogni campione è indicato l'identificativo rispettivamente 1H, 2H e 3H.

ID campione	ng/ μ l	A ₂₆₀	$\frac{A_{260}}{A_{230}}$	$\frac{A_{260}}{A_{280}}$	QUANTITÀ DI RNA (μ g)	NUMERO DI READS
1H	121,97	3,049	1,66	1,96	2.439,4	238.517
2H	236,28	5,907	1,78	1,98	4.725,6	512.345
3H	172,84	4,321	1,84	1,96	3.456,8	1.008.018

Il campione 2H, con A₂₆₀ a 5,907, mostra la maggior quantità di RNA tra i campioni. Il rapporto di assorbanza 260/230, che dimostra il grado di purificazione degli acidi nucleici da contaminanti e carboidrati, è buono per i campioni 2H e 3H (solo leggermente più basso dei valori ottimali 2-2,2) e meno per il campione 1H. Il rapporto 260/280, che quantifica il grado di purificazione rispetto alle proteine, è ottimo per tutti i campioni (il valore ottimale per l'RNA è circa 2). Dal sequenziamento dei tre campioni si ottiene un discreto numero di reads, con il campione 3H a dare il maggior numero di sequenze.

3.2 IDENTIFICAZIONE E PROCESSAMENTO DELLE SEQUENZE

Dopo il sequenziamento Nanopore e il basecalling effettuato con *guppy* si sono ottenute in totale 1.008.018 reads.

Sapendo che durante il sequenziamento è facile incorrere in scambi di basi ed errori, le reads ottenute da *guppy* sono state ulteriormente processate con *rattle cluster*.

In seguito, le sequenze corrette sono state analizzate per identificare trascritti noti ed elaborate col tool *prodigal* per la predizione di sequenze codificanti.

Le sequenze predette sono state analizzate col tool *pfam* per rilevare la presenza di domini proteici.

I risultati sono sintetizzati nella tabella 2.

Tabella 2. Risultati ottenuti con *rattle cluster*, *prodigal* e *pfam*

	Numero effettivo delle sequenze
Reads	1.008.018
Read corrette	944.346
No. di trascritti identificati	5.287
Sequenze codificanti predette	1.530
Sequenze predette con peptide segnale	660

Da questa tabella si può vedere che *rattle cluster* permette di correggere e conservare circa il 94% delle reads ottenute dal basecalling.

Le sequenze corrette sono state confrontate con database di sequenze di RNA, permettendo l'identificazione di 5.287 trascritti noti nel trascrittoma del campione 3H.

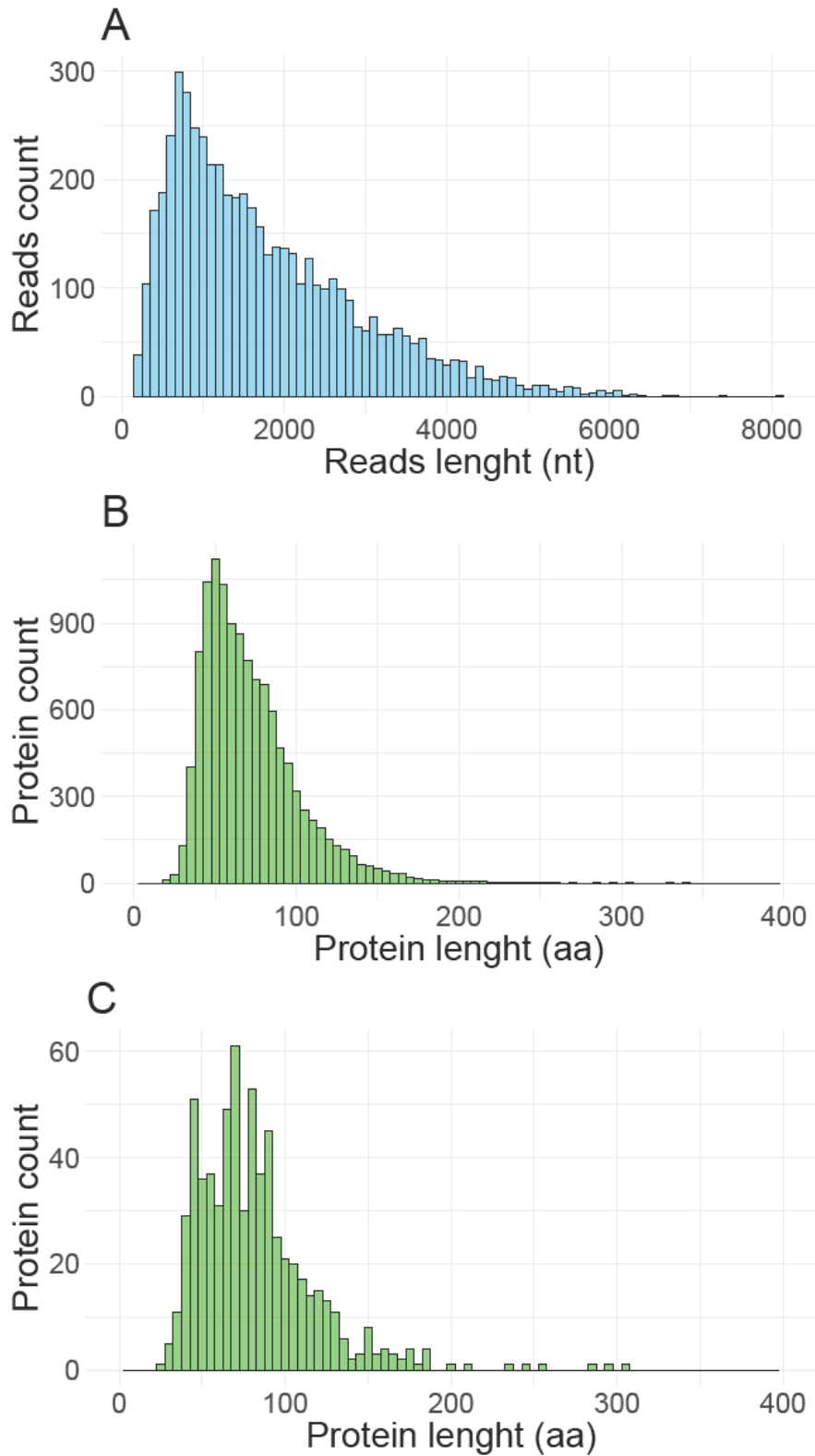
Inoltre, le 944.346 sequenze corrette sottoposte al software *prodigal* per la predizione di sequenze codificanti danno 11.874 sequenze peptidiche, che potrebbero effettivamente corrispondere a proteine o precursori di proteine esistenti.

Tramite un'analisi delle proteine predette, effettuata con il database *pfam*, si sono rilevati 1.530 domini proteici; tra questi è presente anche il dominio del peptide segnale, una sequenza amminoacidica che determina la localizzazione extracellulare delle proteine.

Sapendo che gli AMP sono dei peptidi destinati alla secrezione extracellulare, sono state quindi selezionate le 660 sequenze che presentano il peptide segnale per ulteriori analisi.

Nel processo del sequenziamento è facile che le reads ottenute siano incomplete, specialmente nel caso di sequenziamento dell'RNA, data la relativa instabilità di questa molecola, ed è quindi necessario tenere conto della lunghezza delle reads ottenute per valutare la completezza e la qualità delle informazioni ricavate. Nella figura 1 sono messe a confronto le distribuzioni delle reads, delle sequenze codificanti predette e delle sequenze predette che presentano il peptide segnale; la figura è stata ottenuta utilizzando il software *R studio*.

Figura 1. Distribuzione della lunghezza delle reads (nt) e delle proteine (aa), rispettivamente in azzurro e in verde.



Dalla figura 1 si possono ricavare alcune osservazioni:

- A. Il grafico A mostra la distribuzione delle lunghezze in nucleotidi delle reads ottenute dal sequenziamento Nanopore dRNA-seq. La lunghezza delle molecole di mRNA va da alcune centinaia a migliaia di nucleotidi, mentre la distribuzione delle reads ottenute dal sequenziamento dà un picco alla lunghezza di circa 700-800 basi e la maggior parte delle reads è di lunghezza minore di 3000 basi. Tenendo in considerazione la difficoltà di ottenere sequenze full-length, questo si può considerare un buon risultato.
- B. Il grafico B mostra la distribuzione delle lunghezze in amminoacidi delle proteine tradotte. Le proteine arrivano fino alla lunghezza di migliaia di amminoacidi e la lunghezza media è di circa 300 amminoacidi, mentre le proteine predette dalle reads sono per la maggior parte al di sotto di 150 amminoacidi. Questo può suggerire che nella purificazione del trascrittoma del campione siano stati conservati per la maggior parte i trascritti più corti, quindi per la codifica di piccole proteine, o che le reads siano incomplete o più probabilmente che siano successe entrambe le ipotesi, in quanto è difficile ottenere reads full-length e le sequenze più corte hanno una minor probabilità di degradazione.
- C. Il grafico C mostra la distribuzione delle lunghezze in amminoacidi del sottogruppo di proteine tradotte che presentano il peptide segnale. Come per il grafico B, le proteine hanno una lunghezza inferiore alla media, e inoltre la distribuzione è meno ordinata. D'altra parte, la lunghezza di queste proteine è più in linea con le aspettative, in quanto le proteine destinate alla secrezione extracellulare hanno mediamente una dimensione minore di quelle localizzate intracellularmente.

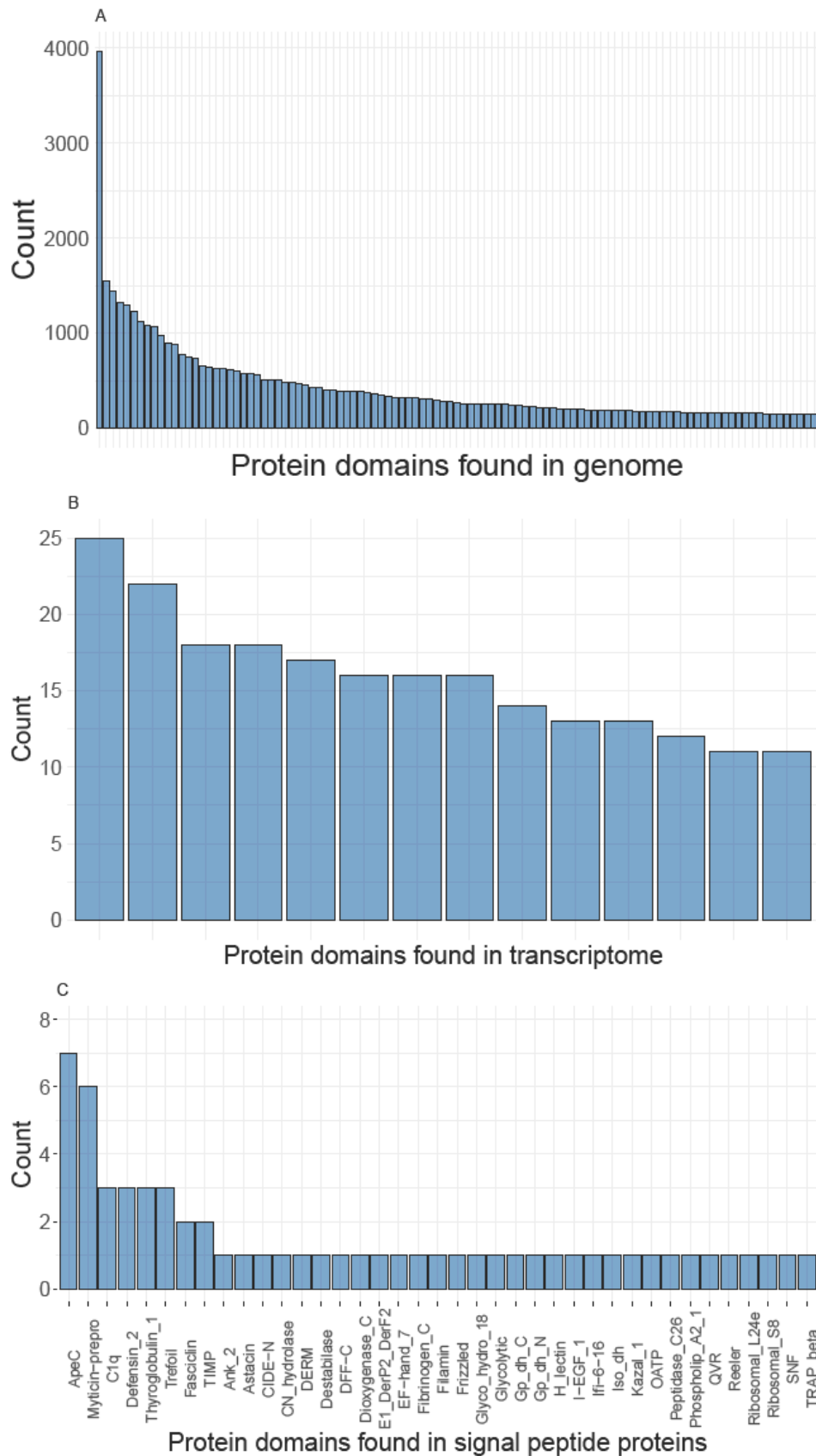
3.3 ANALISI BIOINFORMATICA

Tramite il tool *BLAST* si sono mappati il 75% dei trascritti sul genoma di *M. galloprovincialis*.

I risultati indicano che la predizione dei geni ha supportato 3482 geni; di questi 3328 risultano già annotati e 154 sono nuovi, non conosciuti.

I trascritti sono stati utilizzati per mappare domini *pfam* sul genoma, per osservare la presenza e la frequenza dei domini proteici mappati sul genoma e confrontarli con quelli presenti nel trascrittoma e con quelli presenti nel gruppo di proteine che presentano il peptide segnale; si veda a questo proposito la figura 2.

Figura 2. Lista e frequenza assoluta dei domini proteici presenti nei rispettivi dataset.



Dal momento che il numero di domini mappati sul genoma e individuati nei trascritti è molto alto e pertanto di difficile rappresentazione, nei primi due grafici A e B della figura 2 sono stati rappresentati il 98% dei domini, in ordine di frequenza decrescente (individuati tramite la funzione quantile di *R studio*) e per maggiore leggibilità sono stati omessi i nomi dei domini proteici.

Dalla figura 2 si possono evidenziare i 10 domini proteici che si presentano con maggiore frequenza:

- A. Nel grafico A è rappresentata la frequenza dei domini proteici mappati sul genoma, fino al 98° percentile. I 10 domini proteici più frequenti in ordine decrescente sono:
 1. Ank_2
 2. Zf-C2H2
 3. EGF
 4. Ig_3
 5. RTV_1
 6. I-set
 7. TSP_1
 8. LRR_8
 9. WD40
 10. Lectin_C
- B. Nel grafico B è rappresentata la frequenza dei domini proteici nel trascrittoma, fino al 98° percentile. I 10 domini proteici più frequenti in ordine decrescente sono:
 1. ApeC
 2. Fibrogen_C
 3. Aldedh
 4. ubiquitin
 5. C1q
 6. Proteasome
 7. RRM_1
 8. Ras
 9. Actin
 10. Kazal_1
- C. Nel grafico C è rappresentata la frequenza assoluta di tutti i domini presenti nelle proteine con il peptide segnale. I 10 domini più frequenti in ordine decrescente sono:
 1. ApeC
 2. Myticin_prepro
 3. C1q

4. Defensin_2
5. Thyroglobulin_1
6. Trefoil
7. Fascicolin
8. TIMP
9. Ank_2
10. Astacin

Osservando i domini proteici presenti nel gruppo di proteine col peptide segnale, ovvero le proteine destinate alla secrezione di cui fanno parte anche la classe di AMP, si tratta per la maggior parte di domini con funzione di o correlati alla difesa immunitaria. In particolare:

- Il dominio ApeC (Apextrin C-terminal) media il riconoscimento batterico legandosi ai MDP batterici.
- Il dominio Myticin-prepro è il precursore degli AMP miticine.
- Il dominio C1q è un dominio C-terminale dell'enzima C1 che ha la funzione di attivare il complemento.
- Il dominio Defensin_2 è stato isolato dalle defensine degli artropodi e consiste in sei cisteine conservate a formare ponti disolfuro.
- Al dominio Thyroglobulin_1 invece non è ancora stata attribuita una funzione perché è molto diffuso in varie proteine diverse tra loro.
- Il dominio Trefoil si sa che è molto diffuso nelle proteine eucariotiche extracellulari.
- Il dominio Fascicolin sembra essere un dominio di adesione cellulare antico.
- Il dominio TIMP ha la funzione di complessarsi con le metalloproteinasi inattivandole.
- Molto interessanti sono i domini E1_DerP2_DerF2 e H_lectin, che hanno le funzioni rispettivamente di riconoscimento dei lipidi dei patogeni e di riconoscimento self/non-self, ed è prevedibile che siano quindi correlate alla difesa immunitaria.

Eseguendo un *BLAST* (*E-value* *minimun* 10^{-5}) delle 660 proteine che presentano il peptide segnale, i risultati mostrano 446 sequenze senza hit e 214 con una hit. I risultati sono sintetizzati nella tabella 3.

Tabella 3. Risultati dell'analisi di *BLAST* sui database proteici.

Tipo di hit	Numero di hit
Ipotetici (hypotetical)	88
Non caratterizzati (uncharacterized)	5
Non noti (unnamed)	43
Con funzione nota	78
- di cui AMP	10
TOTALE	214

Più di due terzi delle proteine predette che presentano il peptide segnale, pertanto, non ha trovato riscontro nei database e buona parte delle hit trovate risultano come proteine ipotetiche (ipotizzate da sequenze geniche o trascrittomiche), non caratterizzate (senza famiglia proteica e/o funzione nota) o non note. Tra le 78 proteine risultanti con funzione nota, 10 sono state identificate come AMP:

- 3 miticine
- 2 CRP-I
- 1 defensina
- 1 miticalina C6
- 1 miticofensina A2
- 1 precursore di mitilina D
- 1 precursore di mitimicina

Si ricorda che dall'analisi pfam risultavano 6 domini Myticin-prepro (figura 2); pertanto ci sono 3 miticine in meno dell'atteso. Analizzando i risultati di *BLAST*, risulta che le tre sequenze mancanti, che non sono state riconosciute come miticine, sono state invece identificate come: una no hit, una hypotetical predicted protein e una precursore mitilina D.

La quantità di AMP risultanti dal campione è molto minore dell'atteso, in quanto si tratta di proteine solitamente con un alto grado di espressione.

3.4 OSSERVAZIONI CONCLUSIVE

Il sequenziamento del trascrittoma di un organismo presenta difficoltà, a partire da una buona purificazione del RNA, modo di ridurre la degradazione (l'RNA è una

molecola relativamente instabile), fino a ottenere le sequenze complete e sequenziare anche i trascritti meno rappresentati.

In questo studio si sono eseguiti il sequenziamento e l'analisi del trascrittoma di un campione di *M. galloprovincialis* mediante il protocollo Nanopore dRNA-seq, ottenendo un discreto numero di reads.

Durante l'analisi dei dati ottenuti, e più specificatamente nella visualizzazione delle distribuzioni di lunghezza di reads e delle proteine predette (Figura 2), è risultata evidente la minor lunghezza delle sequenze nucleotidiche e peptidiche ottenute rispetto alla lunghezza media delle molecole di mRNA e a quella delle proteine effettive.

Tenendo in considerazione il fatto che l'RNA è una molecola che si degrada facilmente e che anche i protocolli di sequenziamento più avanzati hanno difficoltà a sequenziare molecole molto lunghe per intero, era previsto e atteso che una parte delle sequenze risultassero incomplete. Come conseguenza, in tal caso, anche le sequenze peptidiche predette hanno lunghezza inferiore a quella effettiva e risultano incomplete, oppure si tratta di piccole proteine derivanti da trascritti di lunghezza ridotta (più facilmente sequenziati).

Dall'elaborazione dei trascritti identificati tramite un *BLAST* sul genoma di *M. galloprovincialis* risulta che sono predetti 3482 geni; di questi 154 non sono noti e possono quindi costituire una nuova scoperta.

Con una analisi incrociata con *pfam* è stato possibile identificare i possibili domini proteici mappati sul genoma, quelli presenti nelle sequenze codificanti predette e quelli presenti nel sottogruppo che contengono il peptide segnale. Tra i domini proteici delle proteine destinate alla secrezione, si osserva che molti hanno funzione correlata alla difesa immunitaria e al riconoscimento di patogeni.

Confrontando tramite *BLAST* le sequenze proteiche predette contenenti peptide segnale con i database proteici sono state identificate 78 proteine di funzione nota, di cui 10 AMP. Il numero e la frequenza degli AMP in generale, e in particolare quelli delle miticine, trovati nel trascrittoma è molto minore dell'atteso.

Si è visto che solo una frazione ridotta delle sequenze dei trascritti e delle sequenze predette è stata identificata. Ciò può significare che molte delle reads ottenute dal sequenziamento sono ancora ignote e si tratta quindi di nuove informazioni.

Il vantaggio del sequenziamento diretto dell'RNA è che è possibile confermare le sequenze ottenute come parte effettiva del trascrittoma, e non far ricorso a quelle predette da DNA genomico e cDNA. Questo fornisce una maggiore attendibilità ai dati e consente di ottenere una maggiore quantità di informazioni sulla espressione e sulla traduzione del trascrittoma dell'organismo.

La tecnologia dRNA-seq è recente e necessita di ulteriore sviluppo, ma mostra di un notevole potenziale che potrebbe superare gli ostacoli che si osservano nel sequenziamento del trascrittoma e far avanzare la ricerca scientifica.

4 BIBLIOGRAFIA

1. [Mitta, G., Vandenbulcke, F., & Roch, P. \(2000\). Original involvement of antimicrobial peptides in mussel innate immunity. FEBS letters, 486\(3\), 185–190. https://doi.org/10.1016/s0014-5793\(00\)02192-x.](https://doi.org/10.1016/s0014-5793(00)02192-x)
2. [Schmitt, P., Rosa, R. D., & Destoumieux-Garzón, D. \(2016\). An intimate link between antimicrobial peptide sequence diversity and binding to essential components of bacterial membranes. Biochimica et biophysica acta, 1858\(5\), 958–970. https://doi.org/10.1016/j.bbamem.2015.10.011.](https://doi.org/10.1016/j.bbamem.2015.10.011)
3. [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\(6\), 182. https://doi.org/10.3390/md15060182.](https://doi.org/10.3390/md15060182)
4. [Bachère, E., Rosa, R. D., Schmitt, P., Poirier, A. C., Merou, N., Charrière, G. M., & Destoumieux-Garzón, D. \(2015\). The new insights into the oyster antimicrobial defense: Cellular, molecular and genetic view. Fish & shellfish immunology, 46\(1\), 50–64. https://doi.org/10.1016/j.fsi.2015.02.040.](https://doi.org/10.1016/j.fsi.2015.02.040)
5. [Gleeson, J., Leger, A., Praver, Y. D. J., Lane, T. A., Harrison, P. J., Haerty, W., & Clark, M. B. \(2022\). Accurate expression quantification from nanopore direct RNA sequencing with NanoCount. Nucleic acids research, 50\(4\), e19. https://doi.org/10.1093/nar/gkab1129.](https://doi.org/10.1093/nar/gkab1129)
6. [Furlan, M., Delgado-Tejedor, A., Mulroney, L., Pelizzola, M., Novoa, E. M., & Leonardi, T. \(2021\). Computational methods for RNA modification detection from nanopore direct RNA sequencing data. RNA biology, 18\(sup1\), 31–40. https://doi.org/10.1080/15476286.2021.1978215.](https://doi.org/10.1080/15476286.2021.1978215)