Передовая
Инженерная
Школа
ИТМО

Training Program: Biology (06.04.01)
Educational Program: Applied Genomics

Research Project Report

Identification of Differentially Expressed Genes in *Metridia longa* Based on RNA-Seq Data

Presenter:
Student, Group A4150: A.I. Kozyreva

Relevance

Bioluminescence is a widespread phenomenon in marine ecosystems and plays a critical role in the survival strategies of marine organisms. It enables various functions such as communication, camouflage, and predator deterrence.

In particular, Metridia longa, a bioluminescent copepod, utilizes this light-emitting ability through a biochemical reaction involving luciferin. Understanding the metabolic pathways responsible for the synthesis and regulation of luciferin in *Metridia longa* holds great potential for application in biotechnology and medicine.

Figure 1. Bioluminescence phenomenon in (a) marine organisms, (b) Metridia longa.

Источник:https://wsbs-msu.ru/2021/12/22/новости-науки-89/

Objective and Research Questions

The main objective of this study is to investigate the hypothesis that luciferin in Metridia longa is synthesized from peptides containing an FYY motif. Additionally, the research aims to identify and annotate differentially expressed genes across various tissues of the copepod.

1. Search for and identify sequences containing the FYY motif in the Metridia longa transcriptome.

2. Perform differential expression analysis to assess the presence and variation of FYY motif-containing target sequences across tissues exhibiting bioluminescence and those without.

3. Compare the identified sequences with metagenomic data to explore the hypothesis that luciferin may be synthesized by symbiotic bacteria.

Theoretical Framework

This study evaluates two potential pathways for luciferin synthesis from one molecule of L-phenylalanine (F) and two molecules of L-tyrosine (Y):

Ribosomal Synthesis: In this scenario, luciferin is formed from the short tripeptide FYY located at the C-terminus of a longer peptide precursor, hypothesized to be H2N-F-Y-Y-COOH.

Non-ribosomal Synthesis: This pathway involves the formation of the tripeptide FYY from free amino acids via non-ribosomal peptide synthetase or through multiple tripeptide intermediates.

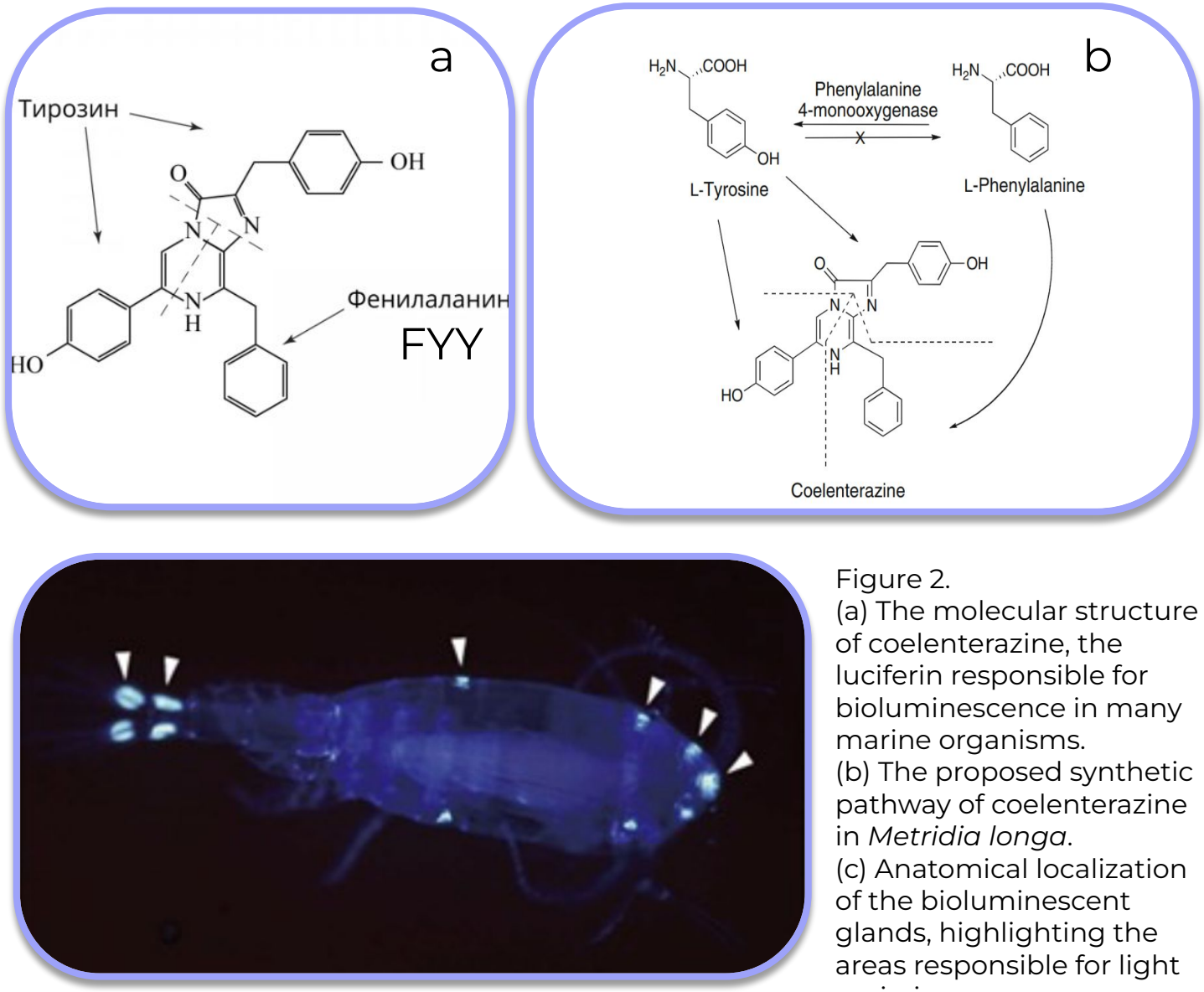

Figure 2.
(a) The molecular structure of coelenterazine, the luciferin responsible for bioluminescence in many marine organisms.
(b) The proposed synthetic pathway of coelenterazine in *Metridia longa*.
(c) Anatomical localization of the bioluminescent glands, highlighting the areas responsible for light emission.

Methods

Sample collection took place during an expedition to the Barents Seas.

- 1 sample of whole copepods
- 3 samples of *Metridia* with glands (Samples 2, 4, 6)
- 3 samples of *Metridia* without glands (Samples 3, 5, 7)

№	Название образца*	Тип образца	Концентрация (нг/мкл)	Объём (мкл)	Растворитель
1	M. longa whole 13.11.23	Тотальная РНК эукариот	162,3	40	Вода
2	M. longa glands + #1 13.11.23	Тотальная РНК эукариот	41,7	40	Вода
3	M. longa glands - #1 13.11.23	Тотальная РНК эукариот	51,1	40	Вода
4	M. longa glands + #2 13.11.23	Тотальная РНК эукариот	73,3	40	Вода
5	M. longa glands - #2 13.11.23	Тотальная РНК эукариот	44,6	40	Вода
6	M. longa glands + #3 13.11.23	Тотальная РНК эукариот	55,6	40	Вода
7	M. longa glands - #3 13.11.23	Тотальная РНК эукариот	50,9	40	Вода

Table 1. Samples metadata

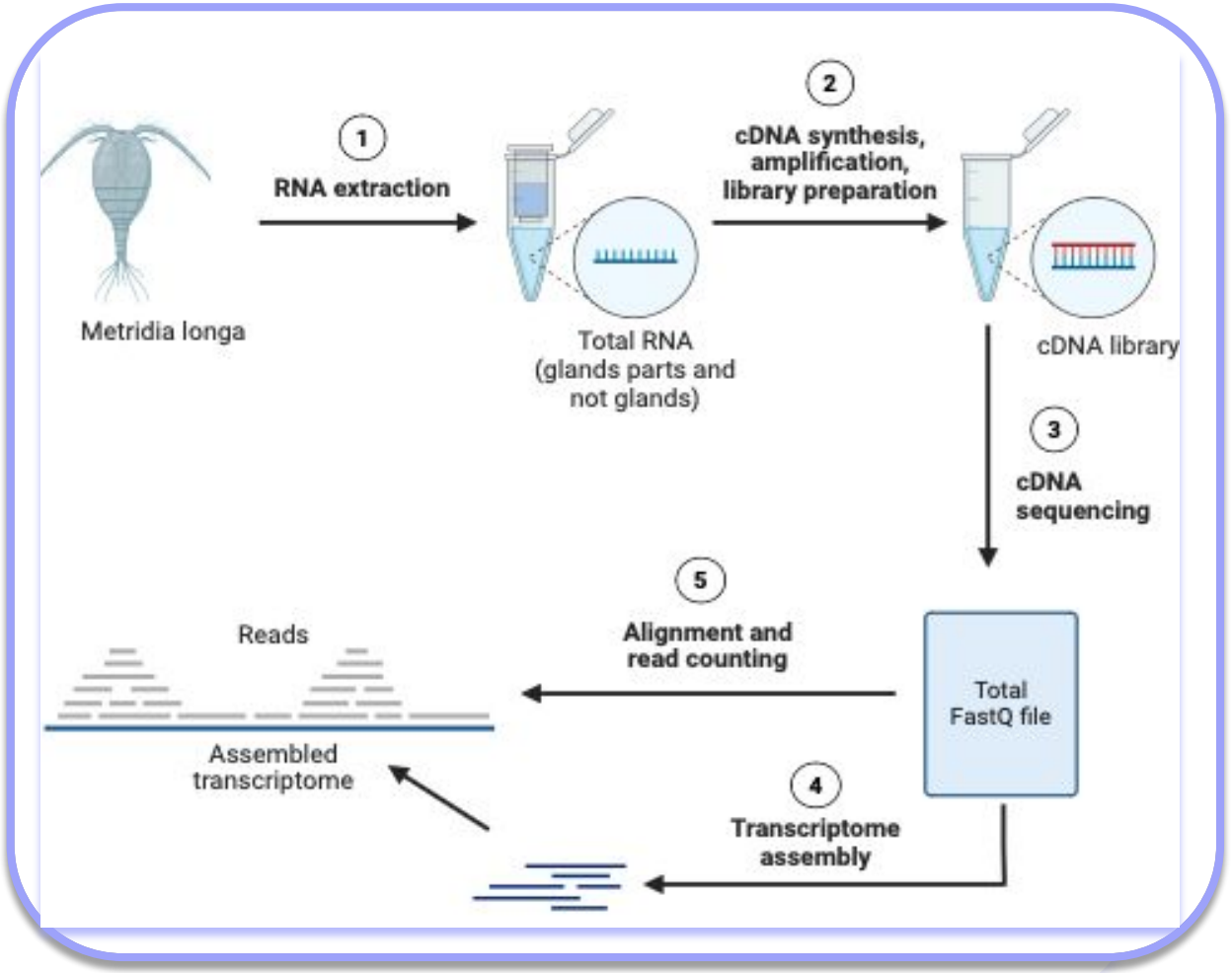

Figure 3. RNA extraction, cDNA sequencing, and data processing workflow: transcriptome assembly and alignment of reads.

Методы исследования

1 FYY Motif Sequence Identification

- RNA-seq Quality Assessment: The quality of raw RNA-seq data was evaluated using FastQC and rRNASort.
- Transcriptome Assembly: De novo transcriptome assembly was performed using Trinity. The quality of the assembled transcriptome was assessed with BUSCO and QUAST to confirm completeness and accuracy.
- Coding Sequence Identification: Open reading frames (ORFs) were predicted using TransDecoder.

2 Differential Gene Expression (DE) Analysis

- Read Mapping: RNA-seq reads were aligned to the assembled transcriptome using HISAT2.
- Count Matrix: Gene expression levels were quantified by generating a count table using FeatureCounts.
- DE Analysis: Differential expression of genes across samples (with and without bioluminescent glands) was analyzed using the DESeq2 library in R.

! Target Sequence Alignment: Metagenomic sequences containing the FYY motif were compared with the identified FYY-sequences using BLAST to explore potential symbiotic bacterial involvement.

Results

The de novo transcriptome assembly of Metridia longa was successfully performed.

Open Reading Frames (ORFs) and potential coding sequences were predicted through TransDecoder, leveraging homology-based methods.

Among the predicted amino acid sequences, we filtered those that ended with the **FYY motif -** resulting in **85 candidate sequences**. These sequences were further analyzed using the RefSeq Protein database as the reference.

Comparison of the obtained sequences with metagenomic data using the **blastp tool** showed an identity of over 95% for all samples, further supporting the potential link between these sequences and their functional relevance to bioluminescence mechanisms.

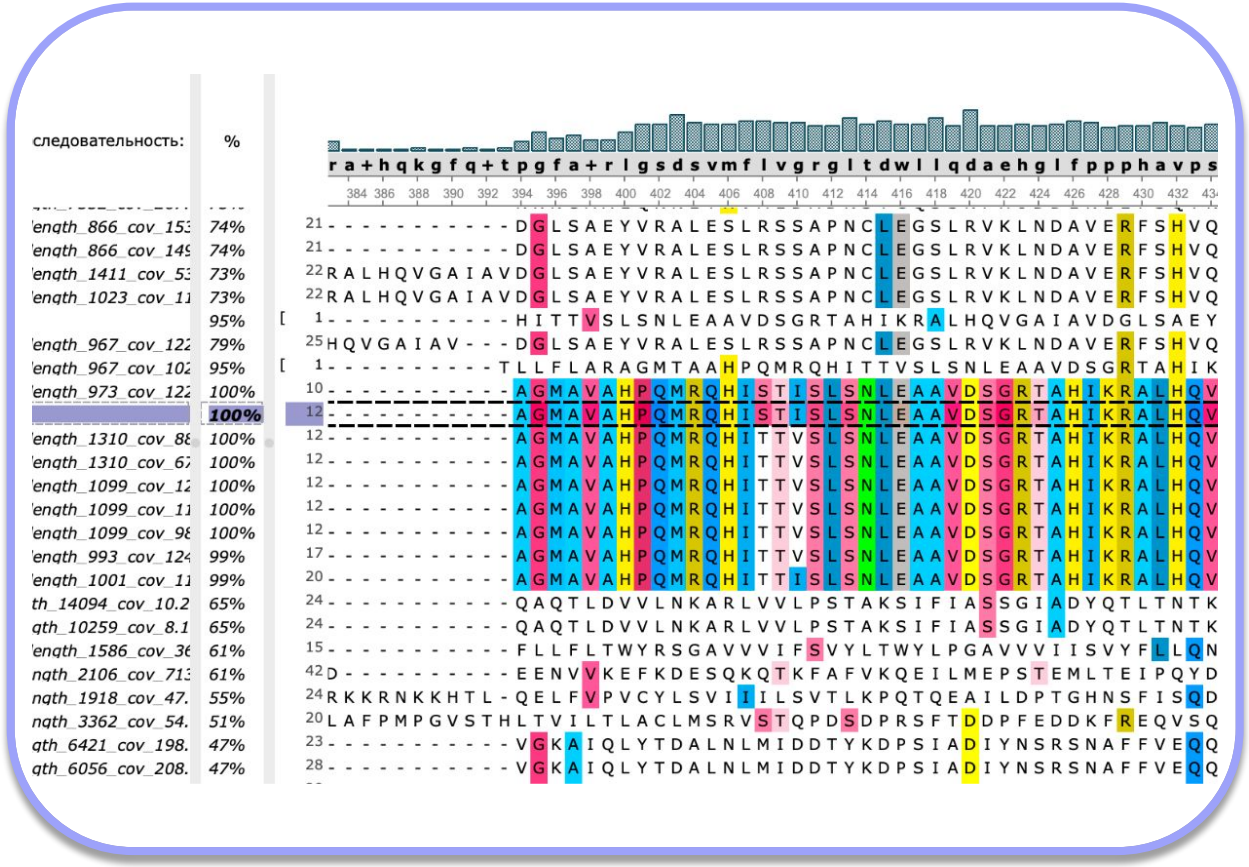

Figure 4. Comparison of FYY-motif peptides from de-novo assembly with literature data from connected metagenomic research.

Results

In the analysis of **Differentially Expressed Genes (DE)** using DESeq2, we identified **88 differentially expressed sequences**.

From these, we filtered sequences based on the standard error of the log2FoldChange (lfCSE < 1.0), leaving **9 significant sequences** with all adjusted p-values (p-adj) < 0.01. These results indicate robust statistical support for differential expression across tissues.

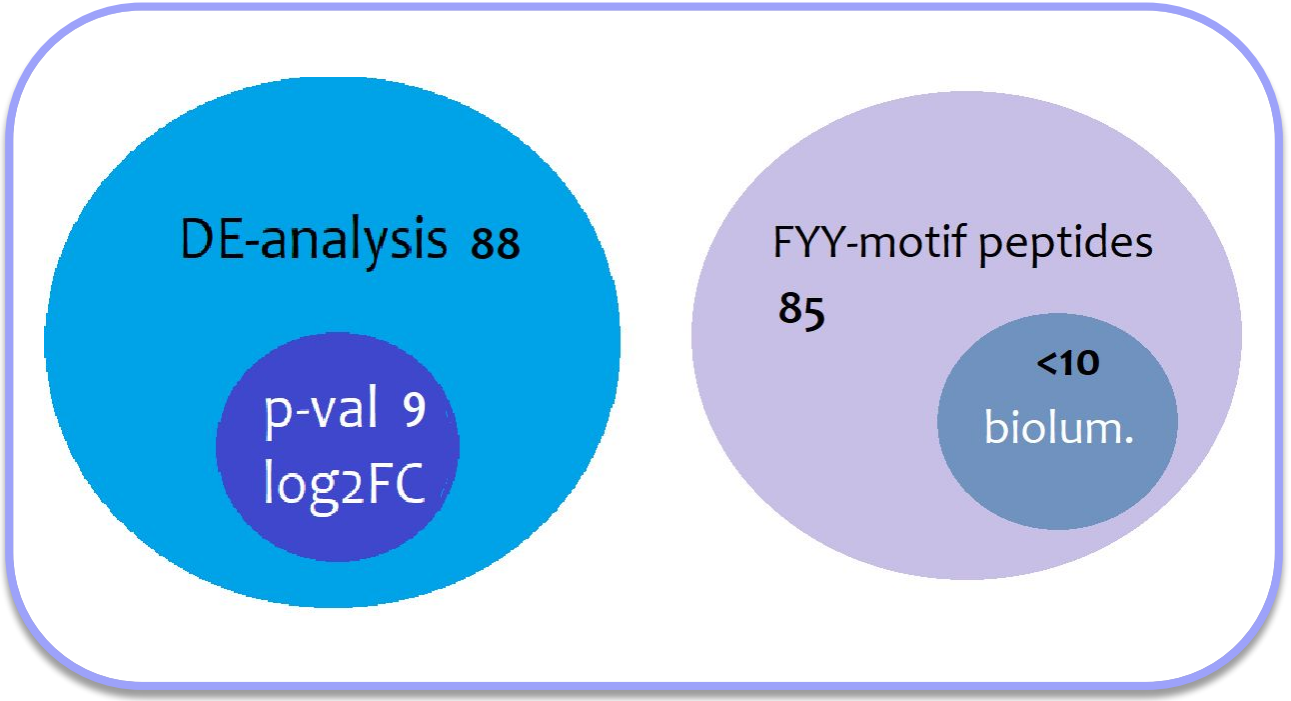

Figure 5. Visualization of research results demonstrates connection absence between sequence detected by different methods.

Conclusions

1. In this study, we successfully performed a de novo transcriptome assembly for Metridia longa and identified 85 coding sequences containing the FYY motif.

2. Despite extensive analysis, the DE analysis revealed no statistically significant differences in the presence of FYY-containing target sequences between bioluminescent and non-bioluminescent tissues of the copepod.

3. Furthermore, when comparing these sequences with metagenomic data, the hypothesis that luciferin may be synthesized by symbiotic bacteria was supported, as we observed an Identity score greater than 95% in multiple sequence alignments.

Literature

- Tessler, Michael et al. “Luciferin production and luciferase transcription in the bioluminescent copepod Metridia lucens.” PeerJ vol. 6 e5506. 14 Sep. 2018, doi:10.7717/peerj.5506

- Haas, Brian J et al. “De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.” Nature protocols vol. 8,8 (2013): 1494-512. doi:10.1038/nprot.2013.084

- Маркова, С. В., and Е. С. Высоцкий. "Целентеразин-зависимые люциферазы (обзор)." Биохимия 80.6 (2015): 845-866.

- Larionova, Marina D et al. “Bioluminescent and structural features of native folded Gaussia luciferase.” Journal of photochemistry and photobiology. B, Biology vol. 183 (2018): 309-317. doi:10.1016/j.jphotobiol.2018.04.050

- Markova, Svetlana V., et al. "Cloning and expression of cDNA for a luciferase from the marine copepod Metridia longa: a novel secreted bioluminescent reporter enzyme." Journal of Biological Chemistry 279.5 (2004): 3212-3217.