

Msc Bioinformatics thesis

Study of Division of Labor in Pseudomonas through single-cell RNA-seq

Valentin Goupille

Master 2 in Bioinformatics

Academic Year: 2024-2025

Internship conducted at Ecobio UMR 6553 CNRS-University of Rennes



Ecobio UMR 6553 CNRS-University of Rennes

Campus de Beaulieu, 35042 Rennes Cedex, France

Under the supervision of:

Solène Mauger-Franklin, Postdoctoral Researcher

Philippe Vandenkoornhuyse, Professor

Presented on 2025-07-01

Table of contents

Copyright notice	v
Declaration	vi
Abstract	vii
Acknowledgements	viii
List of Abbreviations	x
List of Symbols	xi
List of Figures	xii
List of Tables	xiii
1 Introduction	1
1.1 Litterature review	1
2 Materials and Methods	3
2.1 Materials	3
3 Results	4
3.1 Overview	4
3.2 Single-cell RNA-seq Analysis	4
3.3 Functional Analysis	4
3.4 Integration with Previous Studies	4
3.5 Summary of Key Findings	4
4 Discussion	5
4.1 Interpretation of Key Findings	6
4.2 Comparison with Existing Literature	6
4.3 Methodological Strengths and Limitations	6
4.4 Biological Implications	6
4.5 Future Research Directions	6
4.6 Conclusion	6
5 Conclusion and Future Work	7
5.1 Summary of Main Findings	7
5.2 Impact on the Field	7
5.3 Future Research Directions	7

5.4	Final Remarks	7
5.5	References	7
Bibliography		8
Appendices		9
A		9
B	Annexe B: erferfrefref	10
C	Annexe C: codcefe	11

Copyright notice

Produced on 4 avril 2025.

© Valentin Goupille (2025).

Declaration

Statement of originality



I, the undersigned, **Valentin Goupille**, a student in the **Master's program in Bioinformatics**, hereby declare that I am fully aware that plagiarism of documents or parts of documents published on any type of medium, including the internet, constitutes a violation of copyright laws as well as an act of fraud.

As a result, I commit to citing all the sources I have used in the writing of this document.

Date : **01/04/2025**

Signature :

A handwritten signature in black ink, enclosed within an oval shape.

Reproducibility statement

This thesis is written using Quarto. All materials (including the data sets and source files) required to reproduce this document can be found at the Github repository github.com/vgoupille/Internship_2025.

This work is licensed under a [Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).



Abstract

Study of *Pseudomonas brassicacearum* gene expression variation in environmental constraints, towards the validation of Division Of Labor.

Phasellus enim erat, vestibulum vel, adedededliquam a, posuere eu, velit. Nullam sapien sem, ornare ac, nonummy non, lobortis a, enim. Nunc tincidunt ante vitae massa. Duis ante orci, molestie vitae, vehicula venenatis, tincidunt ac, pede. Nulla accumsan, elit sit amet varius semper, nulla mauris mollis quam, tempor suscipit diam nulla vel leo. Etiam commodo dui eget wisi. Donec iaculis gravida nulla. Donec quis nibh at felis congue commodo. Etiam bibendum elit eget erat. Maecenas ipsum velit, consectetur eu, lobortis ut, dictum at, dui. In rutrum. Sed ac dolor sit amet purus malesuada congue. In laoreet, magna id viverra tincidunt, sem odio bibendum justo, vel imperdiet sapien wisi sed libero. Suspendisse sagittis ultrices augue. Mauris metus. Nunc dapibus tortor vel mi dapibus sollicitudin. Etiam posuere lacus quis dolor. Praesent id justo in neque elementum ultrices. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. In convallis. Fusce suscipit libero eget elit. Praesent vitae arcu tempor neque lacinia pretium. Morbi imperdiet, mauris ac auctor dictum, nisl ligula egestas nulla, et sollicitudin sem purus in lacus.

Etiam posuere quam ac quam. Maecenas aliquet accumsan leo. Nullam dapibus fermentum ipsum. Etiam quis quam. Integer lacinia. Nulla est. Nulla turpis magna, cursus sit amet, suscipit a, interdum id, felis. Integer vulputate sem a nibh rutrum consequat. Maecenas lorem. Pellentesque pretium lectus id turpis. Etiam sapien elit, consequat eget, tristique non, venenatis quis, ante. Fusce wisi. Phasellus faucibus molestie nisl. Fusce eget urna. Curabitur vitae diam non

Keywords :

Single-cell RNA-seq, *Pseudomonas brassicacearum*, Division Of Labor, (4-5 keywords) bacterial population, metabolism, specialization, root colonization

Acknowledgements

I would like to thank ... Ecobio ANR Divide

In accordance with Chapter 7.1.4 of the research degrees handbook, if you have engaged the services of a professional editor, you must provide their name and a brief description of the service rendered. If the professional editor's current or former area of academic specialisation is similar your own, this too should be stated as it may suggest to examiners that the editor's advice to the student has extended beyond guidance on English expression to affect the substance and structure of the thesis.

If you have used generative artificial intelligence (AI) technologies, you must include a written acknowledgment of the use and its extent. Your acknowledgement should at a minimum specify which technology was used, include explicit description on how the information was generated, and explain how the output was used in your work. Below is a suggested format:

"I acknowledge the use of [insert AI system(s) and link] to [specific use of generative artificial intelligence]. The output from these was used to [explain use]."

Free text section for you to record your acknowledgment and gratitude for the more general academic input and support such as financial support from grants and scholarships and the non-academic support you have received during the course of your enrolment. If you are a recipient of the "Australian Government Research Training Program Scholarship", you are required to include the following statement:

"This research was supported by an Australian Government Research Training Program (RTP) Scholarship."

You may also wish to acknowledge significant and substantial contribution made by others to the research, work and writing represented and/or reported in the thesis. These could include significant contributions to: the conception and design of the project; non-routine

technical work; analysis and interpretation of research data; drafting significant parts of the work or critically revising it to contribute to the interpretation.

List of Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
ANR	Agence Nationale de la Recherche
DNA	Deoxyribonucleic Acid
DOL	Division Of Labor
NGS	Next Generation Sequencing
RNA	Ribonucleic Acid
RNA-seq	RNA sequencing
scRNA-seq	single-cell RNA sequencing

List of Symbols

Symbol	Definition
α	Alpha
β	Beta
γ	Gamma

List of Figures

1.1	image d'un elephant	1
1.2	image d'un elephant	1

List of Tables

Chapter 1

Introduction

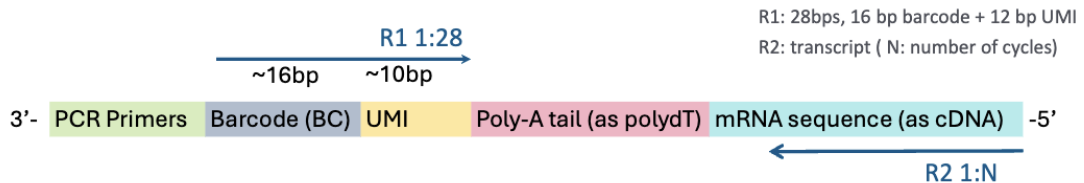


Figure 1.1: image d'un elephant

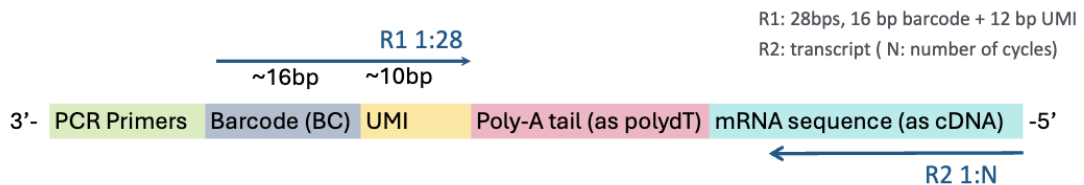


Figure 1.2: image d'un elephant

test de citation de figure Table C.1

je cite la figure Figure 1.1

test ajout

1.1 Litterature review

deddfefde^{1,2} Internship description

The survival of organisms in evolving environments is driven by their fitness. The cost-benefit ratio of traits is constantly balanced and gives rise to different populational evolutionary strategies. To succeed, organisms will have to compete, cooperate and/or specialize as a result of how fit their

traits are considering their biotic and abiotic environment. Bacteria are unicellular organisms with therefore little option to specialize and give up certain traits production to limit their metabolic costs, unlike multicellular organisms that present many different forms of specialized cells in one single organism. However, [[auxotrophic bacteria]] (i.e bacteria lacking genes coding for a molecule essential for their survival) have been studied (Morris et al., 2012).³

Auxotroph bacteria can take advantage of leaky functions of helper's organisms to fulfill their needs in specific compounds (Morris et al., 2014, Estrela et al., 2016).^{4,5} With a reduced genetic material, the beneficiary organism fitness is improved, at the risk of being dependent on the helpers presence in their environment. The conditions in which patterns of such [[division of labor (DOL)]] arise are still obscure, but its advantages for bacterial population are clear: DOL allows to diminish the cost associated to certain functions and the possibility of cohabitation of various mutants/specialized cells within the population to respond as a whole to environmental constraints, and thrive. New technologies allow us to access within-species diversity and study the possible metabolic specialization between cells. Single-cell -omics have been developed for this purpose in human health and are now applied to microbial systems. However, analyzing such datasets still requires custom pipelines to respond to the specificity of bacterial biology and technical challenges.

The goal of this internship is to explore [[scRNA-seq]] (single-cell RNA-seq) datasets of [[*Pseudomonas brassicacearum*]], a root colonizer. The student will analyse samples datasets from various nutritional conditions to determine if DOL can be detected within this species as a strategy for efficient root colonization. The intern will have to implement transcriptomic data analyses from ultra-high throughput sequence run(s). Thus the main aim of the intern will be to set up bioinformatic workflow(s) from existing tools to produce interpretable results.

Chapter 2

Materials and Methods

2.1 Materials

Chapter 3

Results

3.1 Overview

This chapter presents the findings of our single-cell RNA-seq analysis of *Pseudomonas*, focusing on the division of labor within bacterial populations.

3.2 Single-cell RNA-seq Analysis

3.2.1 Data Quality and Preprocessing

3.2.2 Cell Type Identification

3.2.3 Differential Expression Analysis

3.2.4 Division of Labor Patterns

3.3 Functional Analysis

3.3.1 Pathway Enrichment

3.3.2 Gene Set Analysis

3.3.3 Regulatory Network Analysis

3.4 Integration with Previous Studies

3.5 Summary of Key Findings

Chapter 4

Discussion

4.1 Interpretation of Key Findings

4.1.1 Division of Labor Mechanisms

4.1.2 Biological Significance

4.1.3 Technical Considerations

4.2 Comparison with Existing Literature

4.2.1 Similarities with Previous Studies

4.2.2 Novel Insights

4.2.3 Discrepancies and Their Implications

4.3 Methodological Strengths and Limitations

4.3.1 Technical Advantages

4.3.2 Potential Limitations

4.3.3 Future Methodological Improvements

4.4 Biological Implications

4.4.1 Ecological Significance

4.4.2 Evolutionary Perspectives

4.4.3 Potential Applications

4.5 Future Research Directions

4.5.1 Open Questions

4.5.2 Suggested Follow-up Studies

4.5.3 Technical Improvements

4.6 Conclusion

Chapter 5

Conclusion and Future Work

5.1 Summary of Main Findings

5.1.1 Key Discoveries

5.1.2 Methodological Contributions

5.1.3 Biological Insights

5.2 Impact on the Field

5.2.1 Contribution to Single-cell RNA-seq Methodology

5.2.2 Contribution to Pseudomonas Research

5.2.3 Broader Implications for Microbial Ecology

5.3 Future Research Directions

5.3.1 Technical Improvements

5.3.2 Biological Questions to Address

5.3.3 Potential Applications

5.4 Final Remarks

5.5 References

Bibliography

1. Kuchina, A. *et al.* [Microbial single-cell RNA sequencing by split-pool barcoding](#). *Science* **371**, eaba5257 (2021).
2. Gaiser, K. D. *et al.* [High-throughput single-cell transcriptomics of bacteria using combinatorial barcoding](#). *Nature Protocols* **19**, 3048–3084 (2024).
3. Morris, J. J., Lenski, R. E. & Zinser, E. R. [The black queen hypothesis: Evolution of dependencies through adaptive gene loss](#). *mBio* **3**, e00036–12 (2012).
4. Morris, E. K. *et al.* [Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories](#). *Ecology and Evolution* **4**, 3514–3524 (2014).
5. Estrela, S., Kerr, B. & Morris, J. J. [Transitions in individuality through symbiosis](#). *Current Opinion in Microbiology* **31**, 191–198 (2016).

Appendix A

Vous pouvez insérer ici des tableaux supplémentaires ou des sorties d’analyses.

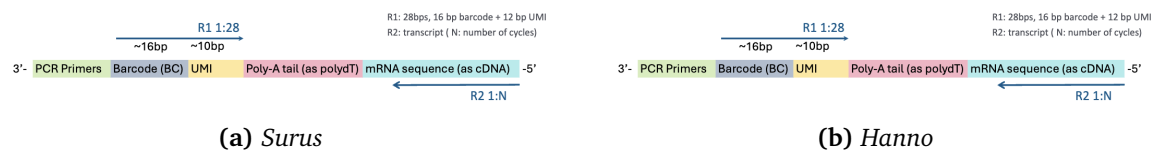


Figure A.1: Famous Elephants

Table A.1: Titre du tableau

Column 1	Column 2	Column 3
Cell 1	Cell 2	Cell 3
Cell 4	Cell 5	Cell 6
Column 1	Column 2	Column 3
Cell 1	Cell 2	Cell 3
Cell 4	Cell 5	Cell 6
Column 1	Column 2	Column 3
Cell 1	Cell 2	Cell 3
Cell 4	Cell 5	Cell 6
Column 1	Column 2	Column 3
Cell 1	Cell 2	Cell 3
Cell 4	Cell 5	Cell 6
Column 1	Column 2	Column 3

Appendix B

Annexe B: erferfrefref



Figure B.1: *Figure A*

Appendix C

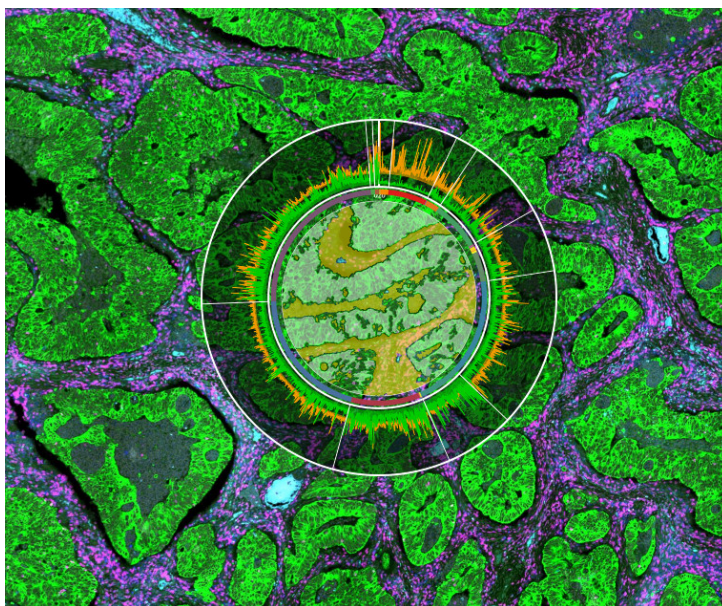
Annexe C: codcefe

Table C.1: *Tableau représentant les résultats des expériences*



Master's Thesis in Bioinformatics

University of Rennes



This thesis was conducted in the framework of the Master's program in Bioinformatics at the University of Rennes. The research presented here contributes to the field of computational biology and bioinformatics.

© Valentin Goupille - ?meta:year

All rights reserved