

Evolution and adaptation in *Pseudomonas* driven by plant volatiles

Filipa Rafaela Lourenço Soares

UMinho |2024



Universidade do Minho
Escola de Ciências

Filipa Rafaela Lourenço Soares

Evolution and adaptation in *Pseudomonas* driven by plant volatiles

June 2024

Universidade do Minho
Escola de Ciências

Filipa Rafaela Lourenço Soares

Evolution and adaptation in *Pseudomonas* driven by plant volatiles

Ph.D. Thesis
Doctoral Program in Molecular and Environmental Biology Biotechnology

Work supervised by
Professor Doctor Pedro Santos
Doctor Pedro Castro

June 2024

COPYRIGHT AND TERMS OF USE OF THE WORK BY THIRD PARTIES

This is an academic work that can be used by third parties as long as they respect internationally accepted rules and good practices, with regard to copyright and related rights. Thus, the work may be used under the terms of the license indicated below. If the user needs permission to use the work under conditions not provided for in the licensing, should contact the author through the RepositóriUM of the University of Minho.

License granted to users of this work



Attribution-NonCommercial-NoDerivatives CC BY-NC-ND
https://creativecommons.org/licenses/by-nc-nd/4.0

Acknowledgments

The submission of this thesis marks the conclusion of my years as a student and the start of my professional career as a scientific researcher. Over the past five years, I've worked hard to learn about different scientific topics and gain the skills needed to make important discoveries in my field. This thesis represents all the effort I've put in, which wouldn't have been possible without the help of many people and organizations.

I am deeply grateful to my advisors Pedro Santos and Pedro Castro for their steady guidance, support, and encouragement throughout this journey. Their expertise, patience (lots of patience), and mentorship have been invaluable in shaping both my research and personal growth. We worked together through tough times, including the Covid-19 saga, delivering scientific knowledge despite the hardships.

A special thank you is addressed to Prof. Dirk Bendorff and his Ph.D. student Patrick Hellwig, from Otto Von Guerick university (Germany), who hosted me in Magdeburg giving me the conditions, time, and guidance to perform most of the proteomics work presented in this thesis.

I am also thankful to Prof. Victor de Lorenzo, from CNB- CSIC (Madrid) who kindly provided the pSEVA Crispr-Cas9 vectors to create the mutant constructed in this thesis. I also acknowledge the members of my "Comissão de Acompanhamento de Tese", namely Prof. Leonilde Moreira from Instituto Superior Técnico (Lisboa), for the helpful insights regarding the work of this thesis.

This work was mainly funded by Fundação para a Ciência e Tecnologia (FCT), Portugal through the Ph.D. grant PD/BD/146184/2019 and the project POCI-01-0145-FEDER-030488. The work was also supported by the Centre of Molecular and Environmental Biology (CBMA, University of Minho), in which the majority of the experimental activities were carried out.

On a more personal note, I thank my friends, colleagues, peers, technicians of CBMA and all the students that accompanied me through this journey (especially Vitor, Andreia and Rafaela who many times endured my bad mood and tiredness particularly by the end of this thesis) for their camaraderie, collaboration, and stimulating discussions. The diverse perspectives and shared experiences have enriched my understanding and improved my passion for research.

As last, but not least, I want to express my gratitude to my family and boyfriend Dinis for their huge support, encouragement, understanding and unwavering love, throughout this very challenging yet rewarding endeavour. Their constant support has been my source of strength and motivation, and they know how these last few years were not the easiest for us. I hope you are all very proud of me for finally finish this journey.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Evolução e adaptação de *Pseudomonas* guiada por voláteis de plantas

A bactéria *Pseudomonas* sp. M1 possui a capacidade de mineralizar o monoterpeno β-mirceno devido a um conjunto especializado de genes presentes numa ilha genómica de 28-kb (IG) integrada no seu genoma. No âmbito dos trabalhos de investigação, foi confirmado que *Pseudomonas* sp. M1 pertence à espécie *P. citronellolis*, uma espécie caracterizada pela sua capacidade de mineralização de diferentes terpenos. Verificou-se que a estirpe M1, assim como vários isolados obtidos por bioprospeção, constituem uma clade própria dentro dessa espécie, associada ao catabolismo de β-mirceno. O principal objetivo deste projeto foi melhorar a capacidade de *Pseudomonas* sp. M1 para biotransformar β-mirceno, através de Evolução Adaptativa e Laboratorial. Para tal, a estirpe M1 foi cultivada em β-mirceno ou lactato como única fonte de carbono, durante 600 gerações. A sequenciação e análise do gDNA das culturas evoluídas revelaram a ausência de mutações dentro da IG associada ao catabolismo de mirceno, em culturas adaptadas a β-mirceno. Alguns isolados derivados destas culturas mostraram eficiência aumentada no seu catabolismo. Entre estes isolados evoluídos, verificou-se que mutações no gene que codifica para regulador FleQ resultaram em mobilidade diminuída, aumento de agregação celular e aumento de atividade transcricional de genes associados ao catabolismo de β-mirceno. Uma análise comparativa do perfil proteómico da estirpe ancestral de *Pseudomonas* sp. M1 e linhagens derivadas evidenciou um aumento na quantidade de proteínas codificadas pela IG associada ao catabolismo de β-mirceno nas linhagens evoluídas/mutante em comparação com a estirpe ancestral, corroborando os resultados observados a nível transcricional. Além disso, 1 das metiltransferases identificadas no perfil proteómico da estirpe M1 ancestral apresentava uma abundância relativa cerca de 1.5x maior em β-mirceno, sugerindo que os genes necessários ao catabolismo de β-mirceno poderiam ser hipermetilados durante o crescimento neste composto, justificando a ausência de mutações dentro deste cluster durante a experiência de adaptação. A previsão dos níveis de metilação do gDNA na estirpe ancestral após o crescimento em β-mirceno, por sequenciação com Oxford-Nanopore, revelou um aumento dos níveis de metilação em gDNAs obtidos de culturas de M1 em β-mirceno, incluindo hipermetilação da IG associada ao catabolismo de β-mirceno. Esta tese constitui o primeiro relato de aumento de capacidade catabólica de uma *Pseudomonas* a um monoterpeno, através de evolução laboratorial, contribuindo para um melhor conhecimento dos mecanismos moleculares subjacentes ao catabolismo de terpenos e à sua adaptação. Os resultados poderão ter implicações práticas na otimização de processos de biotransformação relacionados com monoterpenos.

Palavras-chave: Biotransformação de terpenos, Evolução, Adaptação, Genómica, Proteómica

Evolution and Adaptation in *Pseudomonas* driven by plant volatiles

The bacterium *Pseudomonas* sp. M1 has the ability to mineralize the monoterpene β-myrcene due to a specialized set of genes present in a 28-kb genomic island (GI) integrated into its genome. The work developed in the scope of this thesis confirmed that the strain *Pseudomonas* sp. M1 belongs to the species *P. citronellolis*, known for its ability to mineralize different terpenes. Apparently, *Pseudomonas* sp. M1, along with other isolates obtained through bioprospection, form a distinct clade within this species, all associated with β-myrcene catabolism. The main purpose of this project was to enhance the ability of *Pseudomonas* sp. M1 to metabolize β-myrcene via Adaptive Laboratory Evolution (ALE). To elucidate the adaptation mechanisms, independent cultures were grown with β-myrcene or lactate as the sole carbon sources, for 600 generations. Genome-wide analysis of the gDNAs extracted from the evolved cultures revealed the absence of mutations within the GI associated with β-myrcene catabolism, under β-myrcene cultivation. Isolates derived from these cultures showed increased efficiency in β-myrcene catabolism. Some of these evolved isolates had mutations in the gene encoding the regulator FleQ, resulting in decreased motility, increased cellular aggregation, and increased transcriptional activity of genes associated with β-myrcene catabolism. Comparative proteomic analysis of the *Pseudomonas* sp. M1 ancestor strain and derived lineages evidenced an increased relative abundance of proteins encoded by the GI in the evolved/mutant lineages, corroborating the results observed at the transcriptional level. Furthermore, the relative abundance of one of the methyltransferases identified in the proteomic profile of the M1 ancestor strain, was approximately 1.5-fold increased under β-myrcene cultivation conditions. This observation suggested that β-myrcene catabolism-associated genes could be hypermethylated during growth in this compound, and, thus, justifying (at least partially) the absence of mutations within the 28-kb GI during ALE experiments. Indeed, the prediction of gDNA methylation levels in the ancestor strain grown in β-myrcene, using Oxford-Nanopore sequencing, revealed increased methylation in gDNAs obtained from M1 cultures grown in β-myrcene, including hypermethylation of the GI associated with β-myrcene catabolism. As a whole, this thesis constitutes the first report of increased catabolic ability of a *Pseudomonas* to a monoterpene through ALE, contributing to a better understanding of the molecular mechanisms underlying terpene catabolism and its adaptation. The results obtained and discussed in this thesis may have practical implications for the optimizing of biotransformation processes related to monoterpenes.

Keywords: Biotransformation of terpenes, Evolution, Adaptation, Genomics, Proteomics

Index

Chapter I – General Introduction 24
1.2. Contribution of high-throughput technologies for biotechnology 26
1.3. Engineering microorganisms for enhanced production of desired compounds 30
1.3.1. Laboratorial adaptation and evolution 30
1.3.2. Genetic engineering techniques 32
1.4. Importance of Taxonomy in the biotech context 36
1.5. *Pseudomonas* genus overview and its relevance to the biotech context 38
1.6. *Pseudomonas* sp. M1 as a biotechnologically relevant strain 42
1.6. Outline of this thesis 46
Chapter II - Revisiting the taxonomy context of *Pseudomonas* sp. M1 as an isolate with great biotech potential 47
2.1. Introduction 48
2.2. Material and Methods 49
2.2.1. Phylogenetic analysis – 16S rRNA 49
2.2.2. Phylogenetic analysis – Multi locus sequencing typing (MLST) 49
2.2.3. Phylogenetic analysis – Maximum likelihood tree 50
2.2.4. Genome-based analysis – Average Nucleotide Identity (ANI) and In-silico DNA-DNA hybridization (DDH) 50
2.2.5. Genome-based analysis – Pan-genome analysis of *Pseudomonas citronellolis* 50
2.2.6. Metabolic Fingerprint – GENIII MicroPlate™ test panel Biolog 51
2.2.7. Growth evaluation of *Pseudomonas* sp. M1 and *Pseudomonas citronellolis* in benzoate 51
2.2.8. Verification of phenol metabolism in *Pseudomonas citronellolis* using the β-galactosidase assays 52
2.2.9. Transmission Electron Microscopy (TEM) 52
2.3. Results and Discussion 53
2.3.1. *Pseudomona*s sp. M1 morphology 53
2.3.2. M1 strain is closely related to *P. humi* and *P. citronellolis* through phylogenetic analysis. 53
2.3.3. Pan-genome of *Pseudomonas citronellolis* 57
2.3.4. Comparison of metabolic profiles between *Pseudomonas* sp. M1 and *P. citronellolis* 60
2.3.5. The genomic machinery to metabolize phenol/benzoate is shared between M1 and *P. citronellolis* 61
2.3.6. *Pseudomonas* sp. M1 is spread across the soil 62

Chapter III - Improvement of *Pseudomonas* sp. M1 efficiency towards the catabolism of β-myrcene – Adaptive Laboratorial Evolution (ALE) 66
3.1. Introduction 67
3.2. Material and methods 68
3.2.1. Bacterial strains and culture conditions 68
3.2.2. Adaptive Laboratorial Evolution Experiment 69
3.2.3. Total DNA extraction of evolved cultures, sequencing and datasets treatment 70
3.2.4. Isolation and conjugation of evolved M1 strains to β-myrcene 71
3.2.5. Monitoring the fluorescence of the isolated strains 71
3.2.6. Total DNA extraction of isolated strains, sequencing, and datasets treatment 72
3.2.7. *fleQ* gene deletion in the ancestor strain 73
3.2.8. Motility assay 76
3.3. Results and Discussion 76
3.3.1. Comparative genomics 78
3.3.2. Strains isolated from β-myrcene adapted cultures showed increased performance to catabolize this monoterpene 90
Chapter IV – Proteome assessment of *Pseudomonas* sp. M1 *wt* and M1-derived strains – M2C19, M3C22 and M1ΔFleQ 101
4.1. Introduction 102
4.2. Material and Methods 103
4.2.1. Protein Extraction 103
4.2.2. Protein Quantification 103
4.2.3. SDS-PAGE 104
4.2.4. LC-MS/MS measurements 104
4.2.5. Results analysis 105
4.3. Results 106
4.3.1. Overall analysis 106
4.3.2. Overall fold change of protein abundance between β-myrcene and lactate cultured cells 107
4.3.3. Overall fold change of protein content between evolved isolates and ancestor strain 115
4.3.4. The isolates M2C19 and M3C22, and the mutant M1ΔFleQ have major differences in abundance of proteins regulated by the FleQ regulator 118
4.3.5. Proteome assessment of M2C19, M3C22 isolates and M1ΔFleQ showed increased expression of the 28-kb genomic *locus* proteins in comparison to the ancestor 119
Chapter V – First insights on the methylome dynamics of *Pseudomonas* sp. M1 123

5.1. Introduction 124
5.1.1. DNA adenine methylation (Dam) 124
5.1.2. DNA cytosine methylation (Dcm) 125
5.1.4. Role of methylation in transposition 126
5.1.5. Methylation detection with Oxford-Nanopore technology 127
5.2. Material and Methods 128
5.2.1. High molecular weight genomic DNA extraction 128
5.2.2. gDNA sequencing with MinION from Oxford-Nanopore technology 128
5.2.3. Onward analysis of MinION data 129
5.3. Results and discussion 129
Chapter VI – Final remarks and Future Perspectives 137
6.1. Confirmation that *Pseudomonas* sp. M1 belongs to *Pseudomonas citronellolis* species 138
6.2. ALE increased the efficiency of *Pseudomonas* sp. M1 to mineralize β-myrcene 139
6.3. Evaluation of the impact of *fleQ*-deletion in the biotransformation of other terpenes 140
6.4. Isolate and sequence β-myrcene adapted individuals with decreased susceptibility to antibiotics 140
6.5. Genome methylation levels may impact the evolution course as well as the efficiency of β-myrcene catabolism in *Pseudomonas* sp. M1 141
6.6. Further adaptive laboratorial evolution of *Pseudomonas* sp. M1 mutants towards terpenes 141
6.7. Long-term perspectives 141
Bibliography 143
Supplementary material 164

Abbreviations index

3MB - 3-methylbenzoate

4mC – N4-methylcytosine

5mC – C5-methylcytosine

6mA – N6-methyladenine

ALE – Adaptive Laboratory Evolution

ANI – Average Nucleotide Identity

CFU – Colony Formation Unit

CID – Collision-Induced Dissociation

COG – Cluster of Orthologous Genes

CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats

crRNA – mature CRISPR RNAs

Dam – DNA adenine methylation

Dcm – DNA cytosine methylation

DDA – Data Dependent Acquisition

DDH – DNA-DNA hybridization

DIA – Data Independent Acquisition

dNTP – Deoxynucleotide triphosphate

EMS – Ethyl Methane Sulfonate

FAME – Fatty Acid Methyl Esters

FC – Fold-change

FLB – Flush Buffer

FLT – Flush Tether

GC-MS – Gas Chromatography coupled with Mass Spectrometry

gDNA – genomic DNA

GFP – Green Fluorescent Protein

GGDC – Genome-To-Genome Distance Calculator

GI – Genomic Island

Gm – Gentamycin

GRAS – Generally Recognized As Safe

gRNA – Guide RNA

GST – Glutathione S-Transferases

HDR – Homology-Directed Repair

HPLC – High-Performance Liquid Chromatography

HSP – High-scoring Segment Pairs

INDELs – insertion/deletion mutations

KDO - 2-keto-3-deoxy-manno-octonic acid

Km – Kanamycin

LB – Lysogenic Broth Lennox

LC-MS/MS – Liquid Chromatography-tandem Mass Spectrometry

m/z – mass-to-charge ratio

MLST – Multilocus Sequence Typing

MM – Mineral Medium

MS – Mass spectrometry

MUM – Maximally Unique Matches

NCBI – National Centre for Biotechnology Information

NHEJ – Non-Homologous End Joining

ORF – Open Reading Frame

PAM – Protospacer Adjacent Motif

PASEF – Parallel Accumulation Serial Fragmentation

PCA – Principal Component Analysis

pDNA – plasmid DNA

PIA – *Pseudomonas* Isolation Agar

PLS-DA – Partial Least Squares Discriminant Analysis

RFU – Relative Fluorescence Units

Sm – Streptomycin

SNVs – Single Nucleotide Variants

ssDNA – Single-Stranded DNA

Tc – Tetracycline

TEM – Transmission Electron Microscopy

TIMS – Trapped Ion Mobility Separation

tracrRNA – trans-activating CRISPR RNA

TU – Tanscriptional Unit

wt – Wild-type

Figures index

Figure 1- β-Myrcene, a monoterpene hydrocarbon, is depicted in its molecular structure. It consists of a hydrocarbon backbone with a double bond and several methyl branches. This chemical compound serves as a versatile starting material for several industrial applications. 25

Figure 2 - Illustration of Illumina technologies methodologies. The DNA is firstly fragmented, and each fragment is linked to an adaptor, allowing amplification cycles of the fragment. After amplification, the first sequencing cycle begins, allowing to identify the first base. The second cycle allows to identify the second base, and so on. During sequencing cycles, a DNA polymerase adds nucleotides complementary to the sequence given and each type of dNTP (A, T, G or C) is marked with different fluorescence thus each time one nucleotide is added, a different fluorescence will be captured by a detector, producing data that is transcribed to reads. When sequencing is ready, the reads have to be reconstructed to the whole genome, in a process called assembly. Bioinformatic pipelines are employed at this point. Image based on BioRender. 27

Figure 3 - Illustration of Oxford-Nanopore MinION procedure. Oxford-Nanopore technologies rely on an electrically resistant polymer membrane containing several nanoscale protein pores (nanopores). This membrane is embedded in an electrolyte solution, and upon sequencing, a constant voltage is applied to produce an ionic current through the nanopores. The single stranded DNA or RNA negatively charged are then driven to the pole that is positively charged through the nanopores. Upon the passage of the DNA or RNA, each base will change the ionic current and the change in current will be different for each base. Methylated bases will change the current differently than non-methylated bases, being possible to determine the methylome with this technology. The speed of translocation in controlled by a motor protein, also responsible to unwind double stranded DNA or RNA complexes into single stranded molecules, having helicase activity. Image based on BioRender 28

Figure 4 – Illustration of CRISPR-Cas9 system as a defence mechanism against viruses in bacteria. Upon viral infection, the Cas proteins integrate short segments of viral DNA, known as spacers, into the CRISPR array within the bacterial genome. After subsequent infection with the same virus, the CRISPR array is transcribed into precursor CRISPR RNA (pre-crRNA) molecules, along with trans-activating CRISPR RNA (tracrRNA). These pre-crRNA together with tracrRNA are processed yielding mature guide RNAs (gRNAs), which binds with the Cas9 ribonucleoprotein. This complex actively searches for segments of the viral genome that are complementary to the gRNA sequence. Upon locating these target sequences, Cas9 induces precise cleavage. Image based on BioRender. 34

Figure 5 - The four clades for the *Pseudomonas* genus are represented in green, orange, red and blue. Buqing Yi and collaborators reported the description of each clade based in general characteristics: the clade 1 (green) is the one with the highest number of species, falling into this clade reported human and animal pathogens as *P. aeruginosa* and *P. putida*; the clade 2 (blue) is comprised of species isolated from different environmental sources; the clade 3 (red) is comprised of species related with plants and the clade 4 (orange) are mostly plant pathogens such as *P. syringae*. All the species labelled have potential biotech..... 39

Figure 6 - Organization of the 28-kb genomic *locus* and Transcriptional Units (TU) predicted in silico by P. Soares-Castro and P. M. Santos. β-Myrcene induced the expression of 22 genes in the genomic island of *Pseudomonas* sp. M1. The genes coloured orange represents the gene cluster identified by Iurescia and collaborators, and the gene in green represents the MyrH enzyme. 44

Figure 7 - TEM microscopy of *Pseudomonas* sp. M1. The cells were imaged after growth in lactate (A) and β-myrcene (B). Rod-shaped gram-negative bacteria, with flagella granting motility. 53

Figure 8 - 16S rRNA phylogenetic tree (A) and MLST reference genes (B) for the type-strains with higher identity with *Pseudomonas* sp. M1 16S rRNA sequence. The identity of the 16S rRNA sequences to M1 are indicated in the brackets. The sequences were aligned using *MAFFT* and the trees calculated through kimura-2 parameter model, using neighbor-joining and bootstrapping. Sequences were retrieved from the genomes of each strain from GenBank. 54

Figure 9 - Genome distance of all *Pseudomonas* strains closely related to *Pseudomonas* sp. M1 through 16S rRNA analysis, calculated with *FASTANI*. *Pseudomonas humi* CCA1 is above the cut-off value (95%) regarding *Pseudomonas citronellolis* type-strain, with 97.11% identity..... 55

Figure 10 - Pangenome of *Pseudomonas citronellolis* species. The core genome region is circled in yellow. The phylogenetic tree was constructed with *FastTree* and alignment of sequenced was performed with muscle (these were the options available during the *Anvi'o* workflow), taking into consideration only the core genome region. The resulting phylogenetic tree, based on the core genome of the *Pseudomonas citronellolis* species, showed the degree of similarity between *Pseudomonas* sp. M1 (red), *Pseudomonas citronellolis* DMS50332, the type-strain of this species (green) and *Pseudomonas humi* CCA1 (blue). 58

Figure 11 - Pangenome of *Pseudomonas citronellolis* species. The accessory genome was divided into 2 subregions represented in A (circled in green) and B (circled in orange). Region A comprises 12800 genes, whilst the region B comprises 3200 genes. The phylogenetic tree was constructed with *FastTree* and alignment of sequenced was performed with muscle (these were the options available during the *Anvi'o* workflow), taking into consideration the respective accessory genome clusters. The resulting

phylogenetic trees, based on each accessory genome of the *Pseudomonas citronellolis* species, showed the degree of similarity between *Pseudomonas* sp. M1 (red), *Pseudomonas citronellolis* DMS50332, the type-strain of this species (green) and *Pseudomonas humi* CCA1 (blue). 59

Figure 12 - PLS-DA model (A) and PCA (B) calculated in R showing the differences regarding the metabolism between *Pseudomonas* sp. M1 and *P. citronellolis*. 61

Figure 13 - Growth assay (A) and LacZ assay time-course (B) for *Pseudomonas* sp. M1 and *P. citronellolis* DSM50332 bacteria in Benzoate. The specific growth rate was $0.55h^{-1}$ and $0.62h^{-1}$ respectively. For the LacZ assay, *Pseudomonas* sp. M1 and *P. citronellolis* were conjugated with the vector pPR9TT containing the fusion between the promoter Pa (regulates the expression of catechol 1,2-dioxygenase – PhcA, necessary for phenol and benzoate mineralization) and the LacZ promoter probe. The time-point chosen for this assay was 24h. Data is calculated in Miller units. The experiment was performed until 24h since previous time-points did not have enough signal to be measured. In fact, the promoter Pa is very much induced by phenol, but when Benzoate is being used, the signal will appear after some hours of accumulation, since this compound is also consumed by other pathways, not having a very strong inducing effect. 62

Figure 14 - 16S rRNA phylogenetic tree for the soil isolates similar to *Pseudomonas* sp. M1 and 3 different *P. citronellolis* strains. The sequences were aligned using *MAFFT* and the trees calculated through kimura-2 parameter model, using neighbor-joining and bootstrapping. Sequences were retrieved from the genomes of each strain from *GenBank*. 64

Figure 15 - Setup for Adaptation Laboratorial Experiment. One single colony was used as starting material for all the independent cultures. All cultures were grown in Minimal Media; 5 cultures were supplemented with β-myrcene and 5 were supplemented with lactate. Cells were propagated every 24h and the experiment was repeated until 600 generations. 76

Figure 16 - Number of generations reached within 24h for lactate adapting cultures (blue) and for β-myrcene adapting cultures (red). The graphic represents the average number of generations for β-myrcene or lactate evolving cultures within 24h through all the experiment (until a total of 600 generations). 78

Figure 17 - Percentage of detected intergenic (dark blue) vs coding mutations (light blue) in Myr cultures (A) and Lac cultures (B). 81

Figure 18 - COGs function for genes mutated in Myr cultures (Blue) and Lac cultures (red). Each letter stands for A - RNA processing and modification; B - Chromatin structure and dynamics; C - Energy production and conversion; D - Cell cycle control, cell division, chromosome partitioning; E - Amino acid

transport and metabolism; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme transport and metabolism; I - Lipid transport and metabolism; J - translation, ribosomal structure and biogenesis; K – Transcription; L - Replication, recombination and repair; M - Cell wall/membrane/envelope biogenesis; N - Cell motility; O - Posttranslational modifications, protein turnover, chaperones; P - Inorganic ion transport and metabolism; Q - Secondary metabolites biosynthesis, transport and catabolism; R - General function prediction only; S - Function unknown; T - Signal transduction mechanisms; U - Intracellular trafficking, secretion, and vesicular transport; V - Defence mechanisms; W - Extracellular structures; X - Mobilome: prophages, transposons; Y - Nuclear structure; Z – Cytoskeleton. 84

Figure 19 - COGs function for genes mutated for each culture in Lac conditions (A) and Myr conditions (B). Each letter stands for A - RNA processing and modification; B - Chromatin structure and dynamics; C - Energy production and conversion; D - Cell cycle control, cell division, chromosome partitioning; E - Amino acid transport and metabolism; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme transport and metabolism; I - Lipid transport and metabolism; J - translation, ribosomal structure and biogenesis; K – Transcription; L - Replication, recombination and repair; M - Cell wall/membrane/envelope biogenesis; N - Cell motility; O - Posttranslational modifications, protein turnover, chaperones; P - Inorganic ion transport and metabolism; Q - Secondary metabolites biosynthesis, transport and catabolism; R - General function prediction only; S - Function unknown; T - Signal transduction mechanisms; U - Intracellular trafficking, secretion, and vesicular transport; V - Defence mechanisms; W - Extracellular structures; X - Mobilome: prophages, transposons; Y - Nuclear structure; Z – Cytoskeleton. 85

Figure 20 - Fold change (FC) of promoter P5 activity after induction with β-myrcene, estimated as the ratio of fluorescence units between β-myrcene grown cells and lactate grown cells, subtracting the contribution of intrinsic fluorescence. In blue are the isolates with significant increase in fold-change when compared with the ancestor strain. In grey are represented the ancestor strain as well the isolates with similar fluorescence levels to the ancestor strain. * - p-value >0.05. 92

Figure 21 - Area in square centimetres of the growth halo after 24 hours growth in plates, for swimming motility assay, in all evolved isolates vs ancestor strain. The LB plates used contained only 0.3% agar, the standard Agar percentage for Swimming experiments. M1C2, M3C30 and M4C15 were used as controls to the experiment since mutations in fleQ gene were not detected. M2C19 and M3C22 significantly decreased motility when compared to the ancestor strain. *- pvalue <0.05. 94

Figure 22 - Growth monitoring of all the evolved isolates without hypermutant genotype and ancestor strain. The specific growth rates calculated for each strain were: M1C2- 0.55h^{-1}, M2C19- 0.71h^{-1}, M3C22- 0.67h^{-1}, M3C30- 0.56h^{-1}, M4C15- 0.61h^{-1}, Ancestor- 0.61h^{-1}. 95

Figure 23 - Area in square centimeters of the growth halo after 24 hours growth in plates, for swimming motility assay, in M2C19 and M3C22 evolved isolates, M1 *wt* with *fleQ* gene deleted and ancestor strain. The LB plates used contained only 0.3% agar. *p-value <0.05. 96

Figure 24 - Fold change (FC) of promoter P5 activity after induction with β-myrcene, estimated as the ratio of fluorescence units between β-myrcene grown cells and lactate grown cells, subtracting the contribution of intrinsic fluorescence, measured overtime. The assay was performed for the ancestor strain (blue), M1ΔFleQ mutant (orange), M2C19 adapted isolate (grey) and M3C22 adapted isolate (yellow). 97

Figure 25 - Area in square centimetres of the growth halo after 24 hours growth in plates, for swimming motility assay. The plates used were composed by Minimal Medium with 0.3% agar, supplemented with β-myrcene or lactate as carbon sources. The data in blue, including the plate, are from β-myrcene supplementation and data in red including the plate, are from lactate supplementation. 98

Figure 26 - Fluorescence microscopy of M1 *wt* cells (A), M1ΔFleQ (B) and the evolved isolates M3C22 (C) and M2C19 (D) containing the plasmid pSEVA637 with the promote-probe P5-GFP, after growth in MM supplemented with β-myrcene as only carbon source. The image is magnified by a factor of 400x. 99

Figure 27 - Percentage of COGs function for proteins identified as at least 5 times more abundant after growth with β-myrcene in comparison with lactate (Fold-change > 1.5) – A – and for proteins identified as at least 5 times less abundant after growth with β-myrcene in comparison with lactate (Fold-change < -1.5) – B. Each letter stands for A - RNA processing and modification; B - Chromatin structure and dynamics; C - Energy production and conversion; D - Cell cycle control, cell division, chromosome partitioning; E - Amino acid transport and metabolism; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme transport and metabolism; I - Lipid transport and metabolism; J - translation, ribosomal structure and biogenesis; K – Transcription; L - Replication, recombination and repair; M - Cell wall/membrane/envelope biogenesis; N - Cell motility; O – Posttranslational modifications, protein turnover, chaperones; P - Inorganic ion transport and metabolism; Q - Secondary metabolites biosynthesis, transport and catabolism; R - General function prediction only; S - Function unknown; T - Signal transduction mechanisms; U - Intracellular trafficking,

secretion, and vesicular transport; V - Defence mechanisms; W – Extracellular structures; X - Mobilome: prophages, transposons; Y - Nuclear structure; Z – Cytoskeleton. 110
Figure 28 - Percentage of COGs function Number of proteins at least 5 times more abundant (FC > 1.5) – A – and 5 times less abundant (FC < -1.5) – B – in the evolved isolates (M2C19 and M3C22) in comparison with the ancestor strain. after growth with β-myrcene.–A - RNA processing and modification;–B - Chromatin structure and dynamics;–C - Energy production and conversion;–D - Cell cycle control, cell division, chromosome partitioning;–E - Amino acid transport and metabolism;–F - Nucleotide transport and metabolism;–G - Carbohydrate transport and metabolism;–H - Coenzyme transport and metabolism;–I - Lipid transport and metabolism;–J - translation, ribosomal structure and biogenesis; K – Transcription;–L - Replication, recombination and repair;–M - Cell wall/membrane/envelope biogenesis;–N - Cell motility;–O - Posttranslational modifications, protein turnover, chaperones;–P - Inorganic ion transport and metabolism;–Q - Secondary metabolites biosynthesis, transport and catabolism;–R - General function prediction only;–S - Function unknown;–T - Signal transduction mechanisms;–U - Intracellular trafficking, secretion, and vesicular transport;–V - Defence mechanisms;–W - Extracellular structures;–X - Mobilome: prophages, transposons;–Y - Nuclear structure; Z – Cytoskeleton. 116
Figure 29 - Predicted number of methylations (A) and predicted number of methylated genes (B) for *Pseudomonas* sp. M1 cells growing in β-myrcene or lactate. In blue are the predicted numbers for 5mC methylation and in orange the predicted numbers for 6mA methylation. 131
Figure 30 - Predicted number of 5mC methylations in genes (dark blue) or intergenic regions (light blue) in β-myrcene and lactate growing cells. 134

Tables index

Table 1 - Some of the relevant traits of *Pseudomonas* species with relevance in the bioremediation, agriculture, and industrial fields. 40

Table 2 - Examples of companies using *Pseudomonas* spp. enzymes or the bacteria as a whole in its products formulations. 41

Table 3 - Major highlights of *Pseudomonas* sp. M1 discoveries over time. 43

Table 4 - Products accumulated by *Pseudomonas* sp. M1 *wt* during the catabolism of different terpenes. Products identified by GC-MS[12,6]. 45

Table 5 - Genome distance of *Pseudomonas* sp. M1 strain against the 5 most similar strains. Calculated through *FastANI* and *GBDP*. 56

Table 6 - Genome distance of *Pseudomonas* sp. M1 strain against the soil isolates collected by Pedro Soares-Castro and collaborators. Calculated through *FastANI* and *GBDP*. 63

Table 7 - Total number of reads before and after trimming. The program used was *Trimmomatic* with sliding window of 20 and read length of 40 b.p. The total number of reads after trimming decreased between 50 and 60%. The cultures are numbered from 1 to 10. The numbers 1 to 5 are the Myr cultures and 6 to 10 are the Lac cultures by order. 79

Table 8 - Coverage for each culture before and after trimming. Coverage was calculated using the formula C=LN / G; where C - coverage, L - read length; N - number of reads and G - genome size. The cultures are numbered from 1 to 10. The numbers 1 to 5 are the Myr cultures and 6 to 10 are the Lac cultures by order. 80

Table 9 - Overall mutations number for Lac and Myr cultures at 200, 400 and 600 generations. 81

Table 10 - Number of synonym vs non-synonym mutations in Lac and Myr cultures over generations. 82

Table 11 - Percentage of mutations detected with 100% frequency, above 10% frequency and below 10% frequency, for all the cultures at 200, 400 and 600 generations. 83

Table 12 - Genes mutated in all 4 Lac cultures and with 100% frequency in at least one culture. The table represents the mutated amino acid after translation to protein; the frequency of the mutation in all Lac cultures, the COG and the closest identity match through *Blastp*. In the mutation column, the first letter is the amino acid present in the ancestor strain, the numbers are the position where the mutation occurred, and the second letter is the amino acid present in the mutated protein. 86

Table 13 - Genes mutated in all 4 Lac cultures without taking into consideration mutation frequency. The table represents the mutated amino acid after translation to protein; the frequency of the mutation

in all Lac cultures, the COG and the closest identity match through Blastp. In the mutation column, the first letter is the amino acid present in the ancestor strain, the numbers are the position where the mutation occurred, and the second letter is the amino acid present in the mutated protein. 87

Table 14 - Genes mutated only in Myr cultures and with 100% frequency. The table represents the mutated amino acid after translation to protein; the frequency of the mutation in all Lac cultures, the COG and the closest identity match through *Blastp*. In the mutation column, the first letter is the amino acid present in the ancestor strain, the numbers are the position where the mutation occurred, and the second letter is the amino acid present in the mutated protein. + represents an insertion and Δ a deletion. 88

Table 15 - Proteins affected during evolution in at least two Myr cultures. The table represents the mutated amino acid after translation to protein; the frequency of the mutation in all Lac cultures, the COG and the closest identity match through *Blastp*. In the mutation column, the first letter is the amino acid present in the ancestor strain, the numbers are the position where the mutation occurred, and the second letter is the amino acid present in the mutated protein. + represents an insertion and Δ a deletion. 89

Table 16 - Resistance test of β-myrcene and lactate cultures evolved until 600 generations, and ancestor strain to 30 μg/mL and 60 μg/mL of gentamycin, 50 μg/mL and 100 μg/mL of kanamycin, and 50 μg/mL and 100 μg/mL of tetracycline. The cells were plated in different concentrations, the first drop being without dilution and every other drop diluted 10x consecutively. 91

Table 17 - Proteins affected in the 2 evolved isolates selected after 600 generations. The table represents the mutated amino acid after translation to protein, which mutant has the mutation and the closest identity match through *Blastp*. In the mutation column, the first letter is the amino acid present in the ancestor strain, the numbers are the position where the mutation occurred, and the second letter is the amino acid present in the mutated protein. + represents an insertion. 93

Table 18 - Number of proteins identified for each strain and condition. Each strain was grown in MM supplemented with β-myrcene or lactate and grown until an OD_{600nm} 0.5 or 0.8. 107

Table 19 - Number of proteins at least 5 times more abundant in growth with β-myrcene than in lactate (FC > 1.5) or 5 times less abundant in growth with β-myrcene than in lactate (FC < -1.5)..... 108

Table 20 - Number of proteins at least 5 times more abundant in growth with β-myrcene than in lactate (FC > 1.5) or 5 times less abundant in growth with β-myrcene than in lactate (FC < -1.5) and p-value lower than 0.1..... 109

Table 21 - Top 10 proteins with at least 5 times more abundance in β-myrcene in comparison to lactate (FC > 1.5) or 5 times less abundance in in β-myrcene in comparison to lactate (FC < -1.5) for OD_{600nm} 0.5 and 0.8, and respective statistical significance, for the ancestor strain. In bold are the proteins belonging to the β-myrcene GI. * - p-value 0.05; ** - p-value between 0.05 and 0.1; - statistically non-significative. 111

Table 22 - Top 10 proteins with at least 5 times more abundance in β-myrcene (FC > 1.5) or 5 times less abundant (FC < -1.5) in β-myrcene in comparison to lactate for OD_{600nm} 0.5 and 0.8 for the isolate M2C19, and respective statistical significance. In bold are the proteins belonging to the β-myrcene GI. * - p-value 0.05; ** - p-value between 0.05 and 0.1; - statistically non-significative. 113

Table 23 - Top 10 proteins with at least 5 times more abundance in β-myrcene (FC > 1.5) or 5 times less abundant (FC < -1.5) in β-myrcene in comparison to lactate for OD_{600nm} 0.5 and 0.8 for the isolate M3C22, and respective statistical significance. In bold are the proteins belonging to the β-myrcene GI. * - p-value 0.05; ** - p-value between 0.05 and 0.1; - statistically non-significative. 114

Table 24 - Number of proteins at least 5 times more abundant (FC > 1.5) and 5 times less abundant (FC < -1.5) in the evolved isolates (M2C19 and M3C22) in comparison with the ancestor strain. 115

Table 25 - Fold-change of proteins which expression is reported as to be affected by the FleQ transcriptional regulator, between the adapted isolates/constructed mutant and the ancestor strain and respective statistical significance. The results shown are for the cells grown in β-myrcene, as only carbon source. * - p-value 0.05; ** - p-value between 0.05 and 0.1; - statistically non-significative. 118

Table 26 - Essential proteins to catabolize β-myrcene with higher abundance in the isolates (M2C19, M3C22) and mutant M1ΔFleQ in comparison with the ancestor strain (Fold-change > 1) for OD_{600nm} 0.5. The mutant M1ΔFleQ did not have any protein identified with higher abundance than the ancestor strain. * - p-value 0.05; ** - p-value between 0.05 and 0.1; - statistically non-significative. 120

Table 27 - Essential proteins to catabolize β-myrcene with higher abundance in the isolates (M2C19, M3C22) and mutant M1ΔFleQ in comparison with the ancestor strain (Fold-change > 1) for OD_{600nm} 0.8. * - p-value 0.05; ** - p-value between 0.05 and 0.1; - statistically non-significative. 121

Table 28 - DNA methyltransferases identified for *Pseudomonas* sp. M1 ancestor strain by mass spectrometry and respective fold-change based in the relative abundance of proteins in β-myrcene vs lactate growing cells. 130

Table 29 - Correlation between dispersion and gene methylation of the 25 proteins most abundant in β-myrcene growing cells vs lactate growing cells (FC abundance from Chapter IV) and respective fold-

change for 5mC and 6mA methylation. Significance was calculated based in the p-value. *- p-value between 0.05 and 0.1; **- p-value lower than 0.05. 133

Table 30 - Fold-change of 5mC and 6mA methylation predicted for proteins specific for β-myrcene catabolism, belonging to the genomic island, respective transcriptional unit and statistical significance. *- p-value between 0.05 and 0.1; **- p-value lower than 0.05. 135

Contributions

This thesis was developed in the context of the Ph.D. grant PD/BD/146184/2019, funded by Fundação para a ciência e tecnologia, Portugal (FCT).

Posters communications:

Filipa Soares, Pedro Soares Castro, Pedro M. Santos. "Genomic-guided evaluation of long-term adaptation of *Pseudomonas* sp. M1 in the presence of terpenes." GenomePT, Vairão, 2019.

Filipa Soares, Pedro Soares Castro, Pedro M. Santos. "β-myrcene-enhanced evolution of *Pseudomonas* sp. M1" BioIberoAmerica conference, Braga, 2022.

Filipa Soares, Pedro Soares Castro, Pedro M. Santos. "β-myrcene-enhanced evolution of *Pseudomonas* sp. M1" FEMS conference of Microbiology, Belgrade, 2022.

Filipa Soares, Pedro Soares Castro, Pedro M. Santos. "The uncertainties of a correct species-level taxonomic classification of a *Pseudomonas* sp.", IV SPG meeting, Braga, 2023.

Oral communications:

"Evolution and adaptation in *Pseudomonas* driven by plant-volatiles." In the DPAEM seminars 2020 and 2021.

"The uncertainties of a correct species-level taxonomic classification of a *Pseudomonas* sp." pitch presentation in the IV SPG meeting, Braga, 2023.

Internships:

Internship at Otto Von Guericke University, Magdeburg, Germany, in the Bioprocessing Engineering group, under the supervision of professor doctor Dirk Bendorff, where I worked for 1 month learning all the procedures of protein sample preparation and run in the LC-MS/MS equipment and subsequent data analysis in the proteomics context.

Co-authorships:

Articles

Soares-Castro P., **Soares F.**, Santos PM. Current Advances in the Bacterial Toolbox for the Biotechnological Production of Monoterpene-Based Aroma Compounds. Molecules. 2020 Dec.

Soares-Castro P., **Soares F.**, Reis F., Lino Neto T., Santos PM. Bioprospection of the bacterial β-myrcene-biotransforming trait in the rhizosphere. Appl Microbiol Biotechnol. 2023 Aug.

Posters

Pedro Soares Castro, **Filipa Soares**, Francisca Reis, Teresa Lino Neto and Pedro M. Santos. "Bioprospection for novel functional modules towards plant monoterpene valorisation." – GenomePT, Vairão, 2019.

Rafaela Roque, Andreia Breia, **Filipa Soares**, Diego Carreira-Flores, Marcos Rubal, Pedro Gomes, and Pedro Santos. "Towards the Characterization and Exploitation of the Mesophotic Microbiome of NW Iberian Peninsula." IV SPG meeting, Braga, 2023.

Andreia Breia, **Filipa Soares**, Pedro M. Santos. "Towards the characterization of light-dependent methylome dynamics in *Pseudomonas aeruginosa*." IV SPG meeting, Braga, 2023.

Co-supervisor experience

Between May and June 2022, I co-supervised the final third year projects of the students Ana Moreira, Claudia Costa, Patricia Fernandes and Vanessa Rodriguez in the context of their Biochemistry bachelor.

Between May and June 2023, I co-supervised the final third year projects of the students Rafaela Roque and Ricardo Oliveira in the context of their Biochemistry bachelor.

Laboratory supervisor experience

From 2019 to 2023, I was the appointed supervisor of the Molecular Biotechnology Lab (LBM) at CBMA, ECUM, University of Minho. In this role, my primary responsibility was to maintain the standards of safety and proper functioning rules. I oversaw the compliance with established protocols (e.g. residues management protocols), ensuring the smooth functioning of the lab and the timely report of issues and possible solutions.

Outreach activities

Organizing committee of "V Simpósio PDBMA"

Chapter I – General Introduction

1.1. Exploitation of Terpenes as feedstock for biotechnological processes

Terpenes are secondary metabolites synthesized by plants, and mainly found in essential oils. The synthesis of terpenes occurs through the condensation of isoprene units (C5) and the combination of different numbers of basic five-carbon units originates several types of terpenes, including monoterpenes with 10 carbon, sesquiterpenes with 15 carbon, diterpenes with 20 carbon and so on. In nature, terpenes and their derivatives (terpenoids), are constituents of hormones, photosynthetic pigments, structural components in membranes, among others. The diversity of terpenes derivatives is achieved through the addition of functional groups by enzymes to the terpene backbone, being hydroxylation, oxidation, reduction, acylation and glycosylation the most frequent modifications[1–4]. Some of these compounds, such as monoterpenes, give intense fragrances to the plants and fruits in which they are produced, because of their high volatility. Monoterpenes and monoterpenoids are described to play important roles in the plant physiology, being produced in response to several environmental stimuli, acting as defense mechanism against predators and other microorganisms, attractors for pollinators and communication. Monoterpenes, derived from two isoprene units, have been used since the nineteenth century as flavor and fragrance agents. They also hold significant value in the nutraceutical market due to their functional diversification and market demand[1–4]. Despite their historical use, monoterpenes have yet to be fully exploited as raw materials in industrial processes, regardless of their abundance in waste streams from industries like wood processing. Efforts are ongoing to develop sustainable platforms for converting terpenes into high-value industrial chemicals. These efforts include developing new terpene sources, sustainable chemical transformations, scaling up conversion processes, and understanding the technical, environmental, and economic factors associated with terpene-based manufacturing technologies[5,6]. Among the monoterpenes studied, β-myrcene stands out as a key compound for industrial applications. This monoterpene has been extensively used as a precursor for polymers, insect repellents, and popular flavors and fragrances[7–9]. With its hydrophobic nature and reactive diene structure, β-myrcene is well-suited for industrial processes (Figure 1).

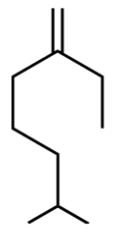


Figure 1- β-Myrcene, a monoterpene hydrocarbon, is depicted in its molecular structure. It consists of a hydrocarbon backbone with a double bond and several methyl branches. This chemical compound serves as a versatile starting material for several industrial applications.

Furthermore, β-myrcene safety profile is established, including a GRAS status, which increases the attractiveness of this compound for various applications in the biotechnology and industrial sectors. In this context, the efforts to modify bacteria into using natural terpene reservoirs to produce value-added compounds, which can serve as valuable chemical precursors, holds considerable significance[8,10,11].

1.2. Contribution of high-throughput technologies for biotechnology

Biotechnology arises as the use of biology to develop new products, methods and organisms with application for human society. For instance, the improvement, enhancement, or engineering of new traits within microorganisms, enabling them to efficiently convert several compounds, including terpenes, into industrially relevant precursors, fall within the biotech context. In recent years, breakthroughs in high-throughput sequencing, screening technologies, molecular biology, and the accompanying bioinformatics tools have revolutionized our ability to uncover novel microorganisms, enzymes, and metabolic pathways. This surge in discovery not only expands our understanding of biological systems but also enables precise customization tailored to the demands of various biotechnological applications. The advent of sequencing technologies has revolutionized our ability to access the full genomes of several microorganisms rapidly and cost-effectively[12,13]. High-throughput sequencing has enabled the simultaneous sequencing of multiple genomes. Illumina sequencing was one of the market leaders due to its capacity to generate vast amounts of data at minimal expense. Illumina sequencing operates by employing a DNA polymerase that incorporates fluorescently labelled nucleotides, enabling the detection of each added nucleotide through distinct fluorescence signals. Parallel sequencing techniques fragment DNA and amplify the resulting fragments with adaptors, producing multiple copies for sequencing. Each sequencing cycle identifies a single base, gradually constructing reads. Moreover, barcoding facilitates the simultaneous sequencing of multiple genomes by tagging DNA fragments with specific identifiers. Following sequencing, computational advances in recent years have aided genome assembly, albeit often resulting in multiple contigs rather than a single molecule. Bioinformatic pipelines aid in reconstructing the genome from these contigs, enhancing data quality. Despite these advancements, sequencing errors such as single nucleotide variants (SNVs) and insertion-deletion errors (INDELs) persist. Various tools are available to address these errors, including adaptor removal, read filtering based on quality or length thresholds, and error correction through nucleotide frequency analysis (Figure 2)[14,15].

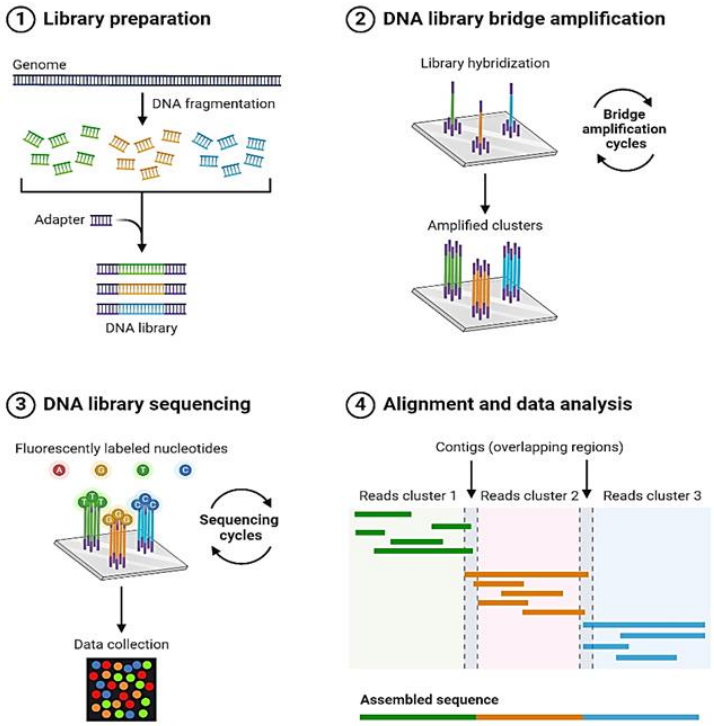


Figure 2 - Illustration of Illumina technologies methodologies. The DNA is firstly fragmented, and each fragment is linked to an adaptor, allowing amplification cycles of the fragment. After amplification, the first sequencing cycle begins, allowing to identify the first base. The second cycle allows to identify the second base, and so on. During sequencing cycles, a DNA polymerase adds nucleotides complementary to the sequence given and each type of dNTP (A, T, G or C) is marked with different fluorescence thus each time one nucleotide is added, a different fluorescence will be captured by a detector, producing data that is transcribed to reads. When sequencing is ready, the reads have to be reconstructed to the whole genome, in a process called assembly. Bioinformatic pipelines are employed at this point. Image based on BioRender.

Looking ahead, third-generation sequencing technologies, such as Oxford-Nanopore sequencing, offer promising solutions by sequencing DNA fragments with high molecular weights, surpassing previous limitations and opening new possibilities for genomic exploration. Since 2014, Oxford-Nanopore technologies has been at the lead of sequencing innovation, introducing the third generation of sequencers characterized by portability and the ability to generate long reads. The MinION device, featured in this study, has facilitated sequencing by its portable and user-friendly design, enabling individual laboratories to conduct real-time sequencing with minimal setup requirements. These compact instruments simply require connection to a computer and a small amount of high-molecular-weight DNA to sequence entire bacterial genomes efficiently. The key advantage of nanopore sequencing lies in its ability to produce long reads, mitigating the challenges posed by repetitive regions encountered with Illumina technology. This cutting-edge approach relies on an electrically resistant polymer membrane containing nanopores, through which single-stranded DNA or RNA molecules, negatively charged, are

driven by a constant voltage. As each base passes through the nanopores, it modulates the ionic current, allowing for base identification. Notably, nanopore sequencing can discern methylated bases, providing insights into the methylome alongside genomic sequencing. A motor protein regulates the speed of translocation, unwinding double-stranded DNA or RNA complexes into single-stranded molecules, a process facilitated by its helicase activity. Computational algorithms, such as *Guppy*, decode the electric signals into reads, a process known as base calling, which can be performed in real-time or post-sequencing (Figure 3).

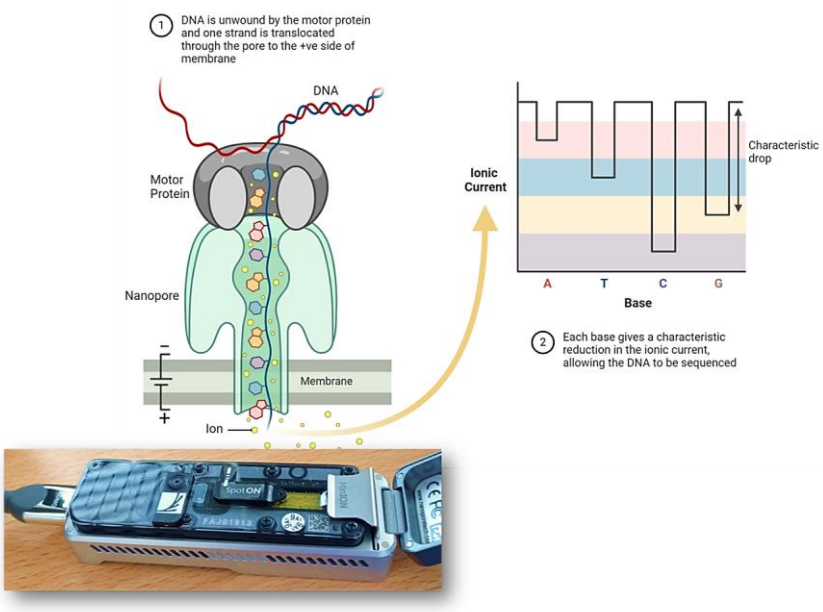


Figure 3 - Illustration of Oxford-Nanopore MinION procedure. Oxford-Nanopore technologies rely on an electrically resistant polymer membrane containing several nanoscale protein pores (nanopores). This membrane is embedded in an electrolyte solution, and upon sequencing, a constant voltage is applied to produce an ionic current through the nanopores. The single stranded DNA or RNA negatively charged are then driven to the pole that is positively charged through the nanopores. Upon the passage of the DNA or RNA, each base will change the ionic current and the change in current will be different for each base. Methylated bases will change the current differently than non-methylated bases, being possible to determine the methylome with this technology. The speed of translocation in controlled by a motor protein, also responsible to unwind double stranded DNA or RNA complexes into single stranded molecules, having helicase activity. Image based on BioRender

Besides nanopore sequencing boasts higher average read quality compared to traditional methods, specialized tools like *Megalodon* enable the identification of DNA and RNA modifications (e.g. methylations), expanding the utility of nanopore sequencing beyond genome sequencing to epigenetic studies[16–19].

1.2. Contribution of multi-omics approaches for biotechnology

Multi-omics approaches are becoming increasingly important in the biotechnology field, providing an overall view of biological systems. For instance, the knowledge gathered from omics fields allows to analyze certain metabolic pathways and its regulation aiding the development of synthetic biology tools and the advancement of metabolic engineering studies necessary for the establishment of a robust and well-characterized biomanufacturing platform. In this context, proteins serve as the cellular effectors, governing essential processes within the cell. While mutations in the genome can impact gene expression, the translation of these genetic changes into phenotypic effects is often modulated by regulatory mechanisms operating at translational and post-translational levels. Thus, protein synthesis represents the culmination of gene expression, being the analysis of the microorganisms proteomes of great importance. Translational control begins with mRNA capping, a process crucial for mRNA stability and protection against degradation by 5′ - 3′ exonucleases. Under normal growth conditions, primary transcripts possess a triphosphate group at their 5′ end, providing protection[20,21]. However, in stress conditions, alternative capping mechanisms, such as Np4A alarmones in *E. coli*, have been identified, highlighting the dynamic nature of mRNA regulation[20,22]. Post-transcriptional modifications, including polyadenylation and methylation, further modulate mRNA stability and accessibility to ribonucleases. RNases, responsible for RNA degradation, undergo modulation by cofactors such as RraA and RraB, which regulate the activity of RNase E, a core component of the bacterial degradosome[23,24]. Additionally, post-translational modifications and proteolysis contribute to protein regulation and turnover, influencing cellular function. Following translation, proteins undergo further regulation through modifications like phosphorylation, acetylation, and methylation, impacting their activity and stability[25,23]. Protein folding, often facilitated by co-factors, is critical for functional activation. Furthermore, proteases mediate protein degradation, maintaining cellular homeostasis[26]. The identification and quantification of proteins are pivotal for understanding cellular processes and pathways affected by evolutionary or engineering interventions. Mass spectrometry (MS) techniques, such as electrospray ionization and Time-of-flight MS, offer high sensitivity and accuracy in protein analysis[27]. Coupling MS with liquid chromatography enables the separation and identification of complex protein mixtures, facilitating comprehensive proteomic profiling[28–30]. Data acquisition methods like data dependent acquisition (DDA) and data independent acquisition (DIA) enhance the sensitivity and specificity of protein detection and quantification. While DDA prioritizes highly abundant peptides, DIA captures all detectable analytes, making it ideal for complex biological samples[31]. The employment of MS methodologies enables the comprehensive analysis of protein content, allowing for comparative studies between wild-type and engineered bacteria. Integration

of proteomics with other omics disciplines, such as genomics, is essential for elucidating microbial phenotypes and optimizing biotechnological applications. A holistic approach, encompassing genome sequencing, phenotypic characterization, and proteomic analysis, forms the cornerstone of successful biotechnological endeavors. A multi-omics perspective facilitates the optimization of microbial strains for enhanced production of target compounds. Moreover, sequencing can also aid in the discovery of novel proteins or post-translational modifications, which can then be validated and characterized using proteomics approaches. This synergy between genomics and proteomics enables a deeper understanding of cellular mechanisms and paves the way for the development of innovative biotechnological solutions.

1.3. Engineering microorganisms for enhanced production of desired compounds

1.3.1. Laboratorial adaptation and evolution

The two fundamental concepts that support contemporary evolutionary biology are the connection of all species through a common ancestor and the principle of natural selection as the mechanism driving evolutionary change. Natural selection is a process by which organisms possessing advantageous traits that promote survival are more likely to pass on those traits to subsequent generations. This leads to the gradual accumulation of beneficial adaptations within a population over successive generations. Such adaptations confer a selective advantage, enhancing the fitness of the organism[32–34]. Evolution, broadly defined, encompasses the dynamic process of trait alteration within biological populations across generations compared to their ancestors. Mutations play a central role in driving evolutionary processes. Mutational events, arising in the genome during replication or in response to environmental stress, introduce genetic variation. Natural selection stands out as the primary driver of adaptive evolution, shaping populations by favoring traits that promote survival in a given environment[35,36]. Organisms exhibiting the most advantageous traits, often referred to as "fitter", are better suited to survive in their respective environments, thereby contributing to the propagation of those advantageous traits within the population[33]. During evolution, all the processes governing the spontaneous mutational events allow the accumulation of not only beneficial but also neutral and deleterious mutations. Beneficial mutations confer a fitness advantage to strains after adaptation and serve as the foundation for adaptive evolution. Neutral mutations have no discernible effect on fitness, while deleterious mutations decrease fitness during adaptation[37,38]. Beneficial mutations, crucial for evolutionary adaptation, are subject to genetic drift and clonal interference phenomena. In experimental evolution, populations typically undergo exponential

or logistic growth, marked by periodic bottleneck events. Genetic drift refers to the loss of rare beneficial mutations due to random fluctuations in population sizes. Additionally, when multiple beneficial mutations arise simultaneously within a population, they engage in competition known as clonal interference. This competition among isolates harboring different beneficial mutations, particularly in large populations, can prolong the time required for fixation[39,40]. The adaptation mechanisms may also be transient, not implying the mutation of the genome. Transient adaptations, such as changes in gene expression, provide immediate responses to environmental challenges but may not necessarily result in heritable changes. Epigenetics is a great contributor for this type of adaptation[41].

The study of evolution spans various biological populations and can be approached at different levels of organization, from molecular genetics to ecological systems. While early studies relied on observational methods to infer evolutionary processes, advances in sequencing technologies have enabled researchers to delve into the molecular foundations of evolution. This integration of molecular and phenotypic data has deepened our understanding of how genetic changes translate into observable traits and behaviors. Microevolution, focusing on changes in allele frequencies and traits within populations over time, provides insights into the mechanisms driving evolutionary change on a smaller scale. Conversely, macroevolution explores patterns of morphological divergence and speciation events over longer timescales, offering a broader perspective on evolutionary history[42].

In laboratory settings, controlled experiments allow researchers to investigate evolutionary processes in a controlled environment. These experiments often involve subjecting populations to selective pressures that mimic natural conditions, such as exposure to antibiotics or environmental stressors. Through these experiments, scientists can observe the emergence and spread of adaptive traits within populations, shedding light on the mechanisms of adaptation and evolutionary change. By employing evolution experiments and sequencing technologies, researchers can calculate mutation rates encompassing all mutation types, including those influenced by natural selection. However, this parameter alone may not offer insights into the adaptive processes at play, as rare beneficial mutations are indistinguishable from neutral or deleterious ones[43–48]. To discern the driving forces behind evolution experiments, performing parallel evolution may aid the identification of beneficial mutations relevant for the adaptation and evolution process. When a single species undergoes evolution independently, under identical conditions and over the same duration, the gene expression profiles of both populations are expected to evolve in tandem. This parallel evolution suggests adaptation and provides valuable insights into the underlying mechanisms driving their adaptive changes[49,50].

Microorganisms, particularly bacteria, have become helpful models for studying evolutionary processes due to their rapid generation times, large population sizes, and genetic flexibility. Evolution experiments with microorganisms offer insights into the mechanisms of adaptation and the dynamics of evolutionary change in real-time[51,52]. Understanding the principles of evolution may be important to seize the adaptive potential of microorganisms to address several challenges, such as antibiotic resistance or to understand and improve metabolic pathways for increased production of metabolites with industrial application. Long-term adaptation experiments can be employed to develop strains with increased desirable traits, such as increased tolerance to stress or enhanced metabolic capabilities[53–59].

1.3.2. Genetic engineering techniques

Advancements in DNA manipulation techniques and the widespread availability of whole genome sequences have paved the way for constructing improved bacterial strains through genetic engineering. Various approaches can be employed to induce mutations in bacterial strains, each with its unique mechanisms and applications. One of the earliest methods for generating chromosomal gene variant libraries involves exposing organisms to DNA-damaging forces or mutagenic compounds that affect the entire chromosome. UV mutagenesis, for instance, is a straightforward and efficient physical mutation method widely used. UV radiation, particularly around 255 nm, which coincides with the DNA absorption spectrum of bacteria, induces the formation of pyrimidine dimers in DNA. These dimers can introduce errors during DNA replication, leading to mutations. Additionally, UV radiation, including UVA and UVB, can indirectly promote the formation of oxidized DNA base damage[60–62]. Chemical mutagenesis, notably using EMS (ethyl methane sulfonate), is another commonly employed method. EMS selectively alkylates guanine bases, causing DNA polymerase to preferentially insert a thymine residue opposite an O-6-ethyl guanine during replication, resulting in random point mutations[63,64]. In addition to chemical and UV mutagenesis, transposons are powerful tools for bacterial mutagenesis. Transposon mutagenesis involves the use of transposable genetic elements that integrate into the genome, generating random insertion mutations. Transposons like Tn5 and Tn10 were commonly employed in bacterial mutagenesis. These elements, containing specific sequences known as inverted repeats, are mobilized by the enzyme transposase, which catalyzes their excision from their original location and insertion into a new site in the genome. This process can disrupt genes, leading to their inactivation or altering their expression levels at the insertion site. Following mutagenesis, specific screening methods are employed to identify mutants of interest. Whole genome sequencing is then used to identify mutations of interest in the

selected mutants. This comprehensive approach allows researchers to engineer bacterial strains with desired traits by manipulating their genomes randomly[65–67].

Depending on the purpose of each study, the use of targeted mutagenesis may be more useful than random mutagenesis, if prior knowledge of the target gene or genomic region already exists. The most used technique for targeted mutagenesis nowadays is the CRISPR-Cas9 system. CRISPR, an acronym for Clustered Regularly Interspaced Short Palindromic Repeats, is a feature found within bacteria, intricately linked with a group of proteins known as Cas (CRISPR-associated) proteins. Among these, Cas9 is recognized for its ability to precisely target and cleave DNA sequences. Originally discovered as part of the bacterial immune system, CRISPR-Cas9 serves as a defense mechanism against invading viruses. When a virus infects a bacterium, the viral DNA is fragmented into distinctive spacers. Cas9, guided by a specific DNA sequence known as the protospacer adjacent motif (PAM), precisely cleaves the viral DNA at the designated site. Subsequently, these DNA fragments, or spacers, are incorporated into the bacterial genome within the CRISPR array, effectively "memorizing" the invader for future encounters. Upon a repeat infection, the bacterium transcribes these stored spacer sequences into precursor CRISPR RNAs (pre-crRNAs). These pre-crRNAs, along with a trans-activating CRISPR RNA (tracrRNA), undergo processing to form mature CRISPR RNAs (crRNAs), also referred to as guide RNAs (gRNAs). This process involves the enzyme RNase III which recognizes specific stem-loop structures within the pre-crRNA and cleaves these structures to release the individual crRNA units (the guide RNAs or gRNAs). These gRNAs serve as molecular guides, directing the Cas9 protein to the complementary sequence on the invading viral DNA. Through the formation of a ribonucleoprotein complex between Cas9 and the gRNA, the bacterial machinery precisely targets the foreign DNA, culminating in the cleavage of the DNA strands at the targeted site. All this process is illustrated in Figure 4.

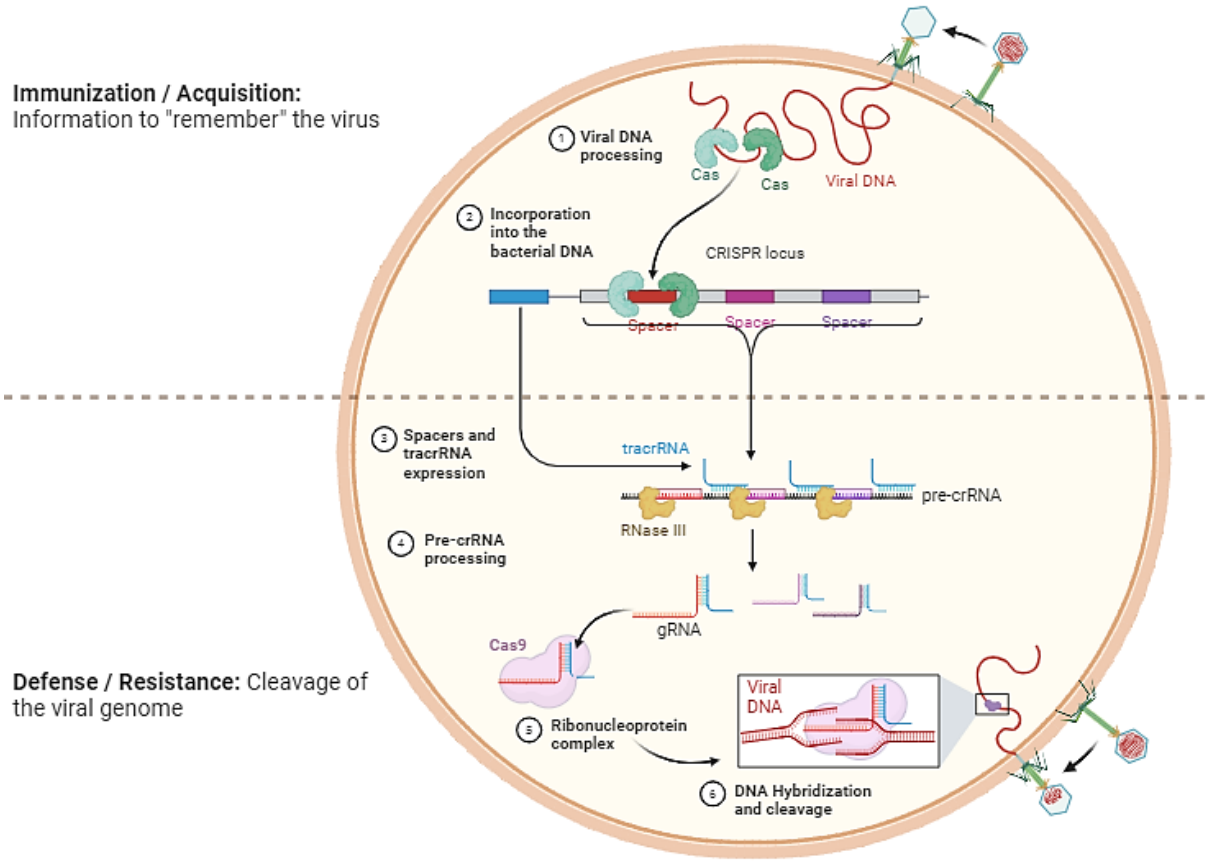


Figure 4 – Illustration of CRISPR-Cas9 system as a defence mechanism against viruses in bacteria. Upon viral infection, the Cas proteins integrate short segments of viral DNA, known as spacers, into the CRISPR array within the bacterial genome. After subsequent infection with the same virus, the CRISPR array is transcribed into precursor CRISPR RNA (pre-crRNA) molecules, along with trans-activating CRISPR RNA (tracrRNA). These pre-crRNA together with tracrRNA are processed yielding mature guide RNAs (gRNAs), which binds with the Cas9 ribonucleoprotein. This complex actively searches for segments of the viral genome that are complementary to the gRNA sequence. Upon locating these target sequences, Cas9 induces precise cleavage. Image based on BioRender.

This disruption initiates cellular DNA repair mechanisms, primarily involving two pathways: non-homologous end joining (NHEJ) and homology-directed repair (HDR). In the NHEJ pathway, the broken DNA ends are directly ligated, often resulting in small insertions or deletions (indels) that can cause frameshift mutations and lead to gene knockout. Conversely, HDR repair is activated in the presence of a homologous DNA template, facilitating precise repair via homologous recombination. By supplying an artificial homologous sequence, researchers can exploit this mechanism to introduce targeted mutations with high precision. In essence, CRISPR-Cas9 not only serves as a guardian of bacterial genomes against viral threats but also stands as a potent tool for precise genome editing, offering unprecedented capabilities for genetic manipulation and research[68–70]. Furthermore, the CRISPR-Cas9 system serves as

an effective counter-selection tool for homologous recombination. Homologous recombination begins with a double strand break in the DNA, often triggered by DNA damage or errors during replication. Introducing targeted mutations in a specific region is possible due to the presence of synthetic single-stranded DNA (ssDNA) with complementary ends to the target region, stimulated by the action of the recombinase. Despite its widespread use, this method suffers from inherent low efficiency, making the identification of mutated clones challenging without selection. To address this issue, Victor de Lorenzo and collaborators developed a technique using the CRISPR-Cas9 system as a counter-selection tool after spontaneous homologous recombination, specifically for *Pseudomonas*. This system consists of three vectors: pSEVA421-Cas9tr containing the Cas9 protein, pSEVA2316 for the cloning of the desired spacer, and pSEVA658-ssr for expressing the homologous recombinase. Additionally, synthetic ssDNA must be provided. In this system, the operator-designed spacer targets the gene or gene region undergoing mutation. As a result, cells where homologous recombination did not occur have their DNA cut by Cas9, leading to cell death. This increases the likelihood of obtaining a mutant when selecting recombinants. Specifically designed and tested in *Pseudomonas*, this technology enables the insertion/deletion of genes or targeted mutations with precision[71].

The genomics revolution has not only boosted the optimization of traits in known microorganisms of biotechnological significance but has also been instrumental in the discovery and exploration of previously unknown microorganisms with relevance to the biotechnology sector. The emergence of metagenomics stemmed from the recognition that prokaryotic diversity surpassed previous estimates, and that the microbial population harbored huge potential for biotechnological applications, pushing beyond the constraints of conventional culture-based approaches. Metagenomics involves the habitat-based analysis of mixed microbial populations at the DNA level, encompassing the genetic material of all microorganisms within a population. This technology not only identifies the microbial species present but also offers insights into their functional roles and metabolic activities. By integrating metagenomics with function-based activity, researchers can efficiently uncover new functional genes from uncultured microbes. While ideally, a metagenomic library would contain clones representing the entire genetic repertoire of a single habitat, the completeness depends on the efficiency of DNA extraction and cloning methods. Metagenomics has emerged as a powerful tool for discovering and characterizing a diverse array of biocatalysts, shedding light on the natural diversity of enzymes and factors influencing their functions, thereby enabling the detection and optimization of biocatalysts for specific processes. A primary focus of metagenomic research has been its application in the search for novel potential biocatalysis processes. The direct sequencing of genetic material from soil sediments and the comparison of sequences with

databases, or through functional analysis screening for specific activities, led to uncover new enzymes such as esterases, nitrile hydratases, alcohol reductases, amidases, cellulases, α-amylases, 1,4-α-glucan branching enzymes, and pectate lyases, underscoring the vast potential of metagenomics in biotechnological innovation[72,73].

1.4. Importance of Taxonomy in the biotech context

Taxonomy provides the essential background for categorizing and identifying microorganisms present in complex environmental samples discovered through metagenomics. By organizing microbes into distinct taxa, researchers gain insights into the diversity within microbial communities and discern the roles played by different groups in ecosystem functions. Taxonomic insights inferred from metagenomic data enable researchers to predict the functional capacities of microbial communities. By linking taxonomic groups with known functional traits, such as enzyme production or metabolite synthesis, metagenomics facilitates the identification of potential biotechnological applications within microbial populations. Traditionally, taxonomy depended heavily on physical traits for species identification and classification. However, the advent of omics technologies has revolutionized our understanding and classification of living species, particularly microorganisms. Accurate identification and classification of microorganisms are vital for understanding their behavior, optimizing study conditions, and protecting against potential risks, especially with pathogenic species[74,75]. In the context of taxonomy, Bergey's Manual has been a cornerstone resource, offering guidelines and rules for microbial classification since its establishment in 1923. Traditionally, taxonomists used several morphological, anatomical, physiological, and ecological characteristics to organize microorganisms[76]. However, with the evolution of genomics, taxonomic characterization expanded to include the 16S rRNA sequence. The 16S rRNA gene, coding for the small ribosomal subunit in prokaryotes, emerged as a primary method for taxonomic identification. This gene contains highly conserved regions interspersed with hypervariable regions, allowing for species differentiation within a genus[77]. By amplifying and sequencing the 16S rRNA gene from environmental samples, known as DNA barcoding, researchers can identify microorganisms present in the sample, underscoring the importance of this sequence in accessing biological diversity. The biggest advantage of 16S rRNA gene is the availability of primers designed to the conserved region of the gene, allowing the identification of different species upon comparison with existing databases[78,79]. Advancements in genomics have facilitated the development of additional techniques to aid taxonomic classification, particularly within the genus. Multilocus Sequence Typing (MLST), introduced in 1998, allows for the precise characterization of bacterial isolates using sequences from multiple housekeeping genes, traditionally 7

housekeeping genes. MLST focuses on analyzing the sequences of internal fragments of multiple housekeeping genes, typically around seven *loci*, selected for their stability and presence across bacterial genomes. These loci are distributed throughout the bacterial chromosome and are chosen to provide a representative picture of genetic diversity. Each locus contains a variable number of alleles, which are defined as distinct sequence variants at that particular *locus*. For example, a typical MLST scheme might target seven housekeeping genes, each with tens to hundreds of known alleles within a given bacterial species. These alleles are numbered sequentially for each locus, and together they form a unique allelic profile for each isolate. The allelic profiles obtained through MLST provide a highly informative fingerprint for bacterial strains, allowing for unambiguous characterization and comparison of isolates across different laboratories and studies. By analyzing the sequence data from multiple loci, MLST captures the evolutionary relationships between isolates and provides insights into the population structure, genetic diversity, and transmission dynamics of bacterial species. This approach considers differences between alleles regardless of the number of nucleotide variations, providing a robust framework for bacterial characterization and taxonomy within the biotechnology field. One of the key conceptual advancements of MLST is the recognition that bacterial populations do not always adhere to a clonal population structure. Instead, MLST acknowledges the importance of genetic exchange events, such as recombination, in shaping bacterial genomes. Therefore, MLST considers both point mutations and recombinational replacements when defining alleles, ensuring a comprehensive understanding of genetic variation within bacterial populations. MLST has improved our ability to characterize bacterial isolates with increased resolution, providing a standardized and portable method for bacterial taxonomy. In the biotechnology field, MLST facilitates the identification of strains with desirable traits for industrial applications, as well as the monitoring of microbial populations in various environmental and clinical settings[80,81]. In addition to MLST, the advancements in high-throughput sequencing have prompted the development of alternative methods for assessing genome similarity, notably Average Nucleotide Identity (ANI). Unlike MLST, which relies on the sequences of specific housekeeping genes, ANI algorithms assess the similarity between the entire coding regions of two genomes. Whole-genome ANI has emerged as a robust measure for estimating genetic relatedness between genomes, particularly within the same species, where organisms typically exhibit ≥95% ANI similarity. This metric calculates the average nucleotide identity of all orthologous genes shared between two genomes, providing a slightly increased resolution between strains of the same species, often with a cutoff value of 94–96% ANI similarity. ANI analysis is adaptable to draft genome sequences obtained from environmental samples using metagenomic or single-cell techniques. These methods, although lacking universally conserved genes,

encode a sufficient number of shared genes, enabling comprehensive genome comparisons on a large scale. ANI values are robust against variations in evolutionary rates and horizontal gene transfer events, as they consider the overall similarity of orthologous genes rather than individual genetic elements[82]. Notably, ANI values exhibit a strong linear correlation with experimental DNA-DNA hybridization (DDH) values, a classic technique used since the 1960s to assess bacterial relatedness[83]. Traditionally, DDH experiments involved the use of radioactive probes to detect DNA hybridization on nitrocellulose filters. However, with the advent of high-throughput sequencing, in silico methods have largely replaced these labor-intensive procedures. Modern algorithms efficiently identify high-scoring segment pairs (HSPs) or maximally unique matches (MUMs) between genomes, allowing for accurate distance calculations and the conversion of these distances into percent-wise similarities, similar to DDH[60,61]. Tools like the Genome-To-Genome Distance Calculator (GGDC) facilitate the application of these computational techniques, providing researchers with a streamlined approach to assess genome relatedness and classification[84].

1.5. *Pseudomonas* genus overview and its relevance to the biotech context

The *Pseudomonas* genus stands out as one of the most taxonomically diverse bacterial genera, with a rich history dating back to its first description in 1894 by Migula, based on morphological studies. Over time, the criteria for *Pseudomonas* taxonomy have been refined, with an exhaustive list of methods detailed in the current edition of Bergey's Manual of Systematic Bacteriology[85]. Among these methods, genomic-based procedures have emerged as the most precise for classification purposes. While the 16S rRNA gene remains a cornerstone of bacterial classification, allowing differentiation at the genus level, its discriminatory power may be limited at the species level within certain phylogenetic groups of *Pseudomonas*[86]. Hence, complementary techniques such as MLST, ANI, and in silico DDH are often employed to enhance species-level resolution. Currently, there are 330 accepted species within the *Pseudomonas* genus, with many having their genomes partially or fully sequenced, as documented by the National Centre for Biotechnology Information (NCBI). Notably, the emerging field of pangenomics has gained importance in understanding the genomic diversity within the *Pseudomonas* genus. Pangenomics explores the complete gene repertoire of a species or genus, comprising a core genome shared by all individuals and an accessory genome that varies between individuals[87,88]. Analysis of the *Pseudomonas* pangenome by Buqing Yi and collaborators revealed that only a small fraction (1.62%) of genes constitutes the core genome, underscoring the high diversity within this genus. The vast majority (98.38%) of genes belong to the accessory genome, highlighting the heterogeneity of *Pseudomonas* species. Using pangenomics data, phylogenetic trees based on core genome genes have elucidated four

distinct clades within the *Pseudomonas* genus (Figure 5). Similarly, phylogenies based on ANI showed concordant results, further validating the evolutionary relationships identified through pangenomics analysis[89].

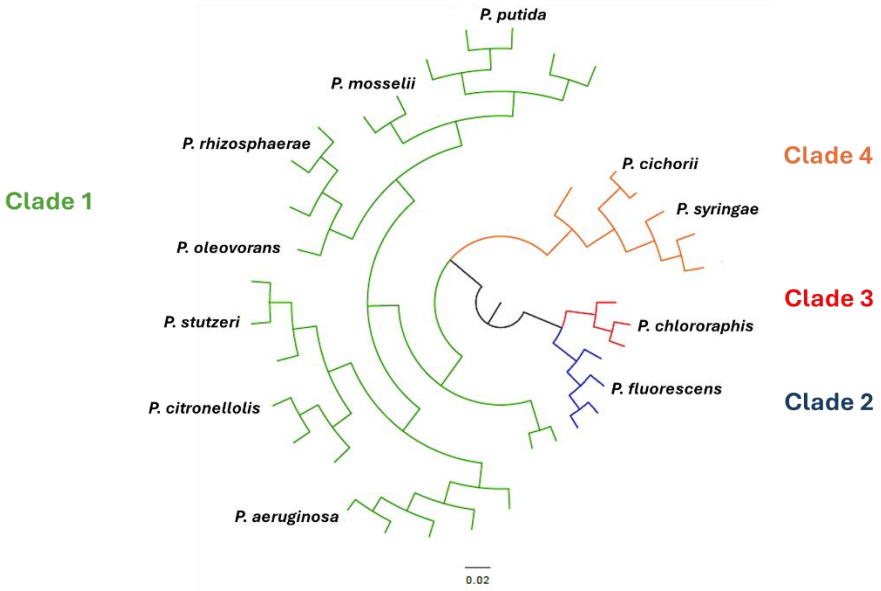


Figure 5 - The four clades for the *Pseudomonas* genus are represented in green, orange, red and blue. Buqing Yi and collaborators reported the description of each clade based in general characteristics: the clade 1 (green) is the one with the highest number of species, falling into this clade reported human and animal pathogens as *P. aeruginosa* and *P. putida*; the clade 2 (blue) is comprised of species isolated from different environmental sources; the clade 3 (red) is comprised of species related with plants and the clade 4 (orange) are mostly plant pathogens such as *P. syringae*. All the species labelled have potential biotech.

Despite the advances in sequencing technologies, phenotypic-based methods still play a crucial role in characterizing and identifying *Pseudomonas* strains. Strains isolated from non-clinical sources are usually analyzed using Biolog Microbial Identification Systems, which provide a rapid and reliable approach by assessing microbial metabolic activity across various carbon sources. Additionally, chemotaxonomic characteristics such as quinone systems and fatty acid profiles (FAME) are routinely employed for bacterial characterization, with FAME analysis proposed as an effective identification method[90,91]. The *Pseudomonas* genus represents a rich reservoir of genetic diversity and metabolic versatility, being an attractive target for biotechnological applications across diverse industries. The extensive metabolic capacity of *Pseudomonas* species allows them to thrive in a wide range of environments, from soil and water to plant and animal hosts. Table 1 shows several relevant *Pseudomonas* strains with biotechnologically relevant traits. This adaptability is particularly advantageous in bioremediation efforts aimed at cleaning up environmental pollutants and contaminants. *Pseudomonas* spp. play a crucial role

in breaking down and detoxifying hazardous substances. These bacteria are adept at metabolizing a variety of organic pollutants, including hydrocarbons, pesticides, and industrial chemicals[92]. Their ability to degrade complex molecules into simpler, non-toxic compounds has been exploited in the cleanup of contaminated sites, such as oil spills and industrial waste sites[93,94].

Table 1 - Some of the relevant traits of *Pseudomonas* species with relevance in the bioremediation, agriculture, and industrial fields.

Pseudomonas strain	***Relevant traits***	***Reference***
	Bioremediation	
Pseudomonas sp. SA3	Biosurfactant production	95
Pseudomonas acidovorans CA28	Degradation of 3-chloroaniline after immobilization in calcium alginate	96
Pseudomonas putida	Resistant to cadmium/Precipitation of copper/ Degradation of phenols and aromatic compounds	97,98,99
Pseudomonas aeruginosa	Remove cadmium from solutions	100
	Agriculture	
P. aeruginosa, P. fluorescens, P. putida, and P. vancouverensis	Suppression of plant pathogens through production of antagonistic compounds (cell wall degradation enzymes, antibiotics etc.) counteracting plant disease and infection.	101,102,103,104
Pseudomonas fluorescens	Triggers the induced systemic resistance of plants through secretion of siderophores and phytohormones	105
Pseudomonas stutzeri A1501	Nitrogen fixation	106
Pseudomonas soil isolates (agricultural fields)	Mineralize $Ca_3(PO_4)_2$, producing organic phosphate that can be used by plants.	107
Pseudomonas putida GR12-2	Produce of phytohormone indoleacetic acid (IAA), hormone that stimulates cell elongation and cell division in plants	108
	Industry	
Pseudomonas sp. BCNU 106, P. stutzeri and P. putida	Trehalose producers (used to preserve products from desiccation)	109
Pseudomonas ficuserectae, P. fluorescens, and P. putida	Express anti-freeze proteins	110

In agriculture, *Pseudomonas* species are known for their beneficial interactions with plants, including promoting growth, enhancing nutrient uptake, and protecting against pathogens. Certain strains of *Pseudomonas* have been developed as biofertilizers and biocontrol agents, offering sustainable

alternatives to chemical fertilizers and pesticides. By harnessing the plant growth-promoting and disease-suppressing properties of these bacteria, farmers can improve crop yields while reducing reliance on synthetic compounds[111]. In the industrial sector, *Pseudomonas* spp. are valued for their ability to produce a wide range of valuable biomolecules, including enzymes, biosurfactants, and chemicals. Their metabolic versatility allows them to use a variety of substrates, including waste materials from agricultural and industrial processes, for the production of high-value products[112,113]. Thus, several biotech industries are already using *Pseudomonas* or molecules derived from *Pseudomonas* species, for instance enzymes in its formulations to produce value-added compounds. Some examples are shown in Table 2.

Table 2 - Examples of companies using *Pseudomonas* spp. enzymes or the bacteria as a whole in its products formulations.

Product	Application	Strain	Company	Reference
Lumafast®, Lipomax®	Lipase P production. Application in detergent industries	*Pseudomonas* spp. Lipases	Amano Pharmaceutical Co. Ltd., Gist-Brocades	114
Rhamnolipids production	Detergents industry	*Pseudomonas putida* KT2440	Evonik	115
Howler® **Cedomon®** **Spot-less®** **Ateze®** **Cerall®**	Fungicides	*Pseudomonas chlororaphis* strain AFS0 *P. chlororaphis* MA 342 *P. aureofaciens* Tx-1 *P. chlororaphis* MA342	AgBiome, BioAgri AB, Turf Science Laboratories, Koppert Biological Systems	116,117
Zequanox®	Biocides	*Pseudomonas fluorescens* CL 145A	Marrone Bio Innovations	118
2-phenyllactic acid	Production of aromatic compound	*Pseudomonas gladioli*	IFF	119
Munox®	Water treatment and bioremediation of petroleum hydrocarbons	*Pseudomonas* spp.	Phibro Biotechnics	120

Overall, the diverse metabolic capabilities and environmental adaptability of *Pseudomonas* species make them invaluable assets in biotechnology. By exploiting the unique properties of these bacteria, researchers and industries can develop innovative solutions for environmental remediation, agricultural

sustainability, and industrial production, contributing to a more sustainable and environmentally friendly future.

1.6. *Pseudomonas* sp. M1 as a biotechnologically relevant strain

Among the diverse array of *Pseudomonas* species renowned for their biotechnological utility, *Pseudomonas* sp. M1 stands out as a well-characterized bacterium with biotechnological utility. Discovered in 1999 by Iurescia and collaborators, this gram-negative, rod-shaped bacterium drew attention for its exceptional ability to metabolize aromatic and hydrophobic monoterpenes[120]. As research on *Pseudomonas* sp. M1 progressed, attempts were made to taxonomically classify the bacterium within the *Pseudomonas* genus, with initial suggestions leaning towards its alignment with *Pseudomonas citronellolis* due to a 95% genetic similarity[121]. Since its identification, several studies have delved into unravelling the particularities of this organism. A brief timeline with the major cornerstones about *Pseudomonas* sp. M1 is depicted in Table 3.

In 2002, Santos and co-workers reported on the catabolism of phenol and chlorophenols by *Pseudomonas* sp. M1, shedding light on the interplay between these pathways and glutathione S-transferases (GST). Despite the absence of genes encoding GST proteins within the phenolic operons, the researchers identified two GST proteins — GST_M1A and GST_M1B — whose activity arose only during phenol catabolism. This induction of GST was particularly pronounced at toxic phenol concentrations, suggesting a regulatory link between the phenol metabolism and GST activity. Furthermore, the authors explored the impact of various organic acids on GST activity when phenol was introduced, revealing increased enzyme activity even in the presence of less favorable carbon sources like pyruvate[122]. This underscored GST role as a responsive element to specific metabolic stresses induced by the phenol metabolism. In 2007, a quantitative proteomics approach contributed to the understanding of Pseudomonas sp. M1 intricate adaptive response to phenol[123]. In the same year, the *phc* gene cluster responsible for phenol catabolism in *Pseudomonas* sp. M1 was elucidated. This cluster comprised genes such as catechol 1,2-dioxygenase (*phcA*) and a multicomponent phenol hydroxylase (phcKLMNOP), along with the transcriptional regulator PhcR. Furthermore, the study revealed the capability of *Pseudomonas* sp. M1 to metabolize benzene, displaying shared genetic machinery for phenol and benzene catabolism[121].

Table 3 - Major highlights of *Pseudomonas* sp. M1 discoveries over time.

Year	Event	Reference
1999	• Isolation of *Pseudomonas* sp. M1. • Characterization as bacteria able to completely mineralize the monoterpene β-myrcene. • First steps into the characterization of β-myrcene catabolic pathway.	124
2002	• Report of *Pseudomonas* sp. M1 ability to catabolize phenol and chlorophenols.	122
2007	• Establishment of the adaptive response of *Pseudomonas* sp. M1 to phenol. • First taxonomic classification attempt of *Pseudomonas* sp. M1 suggesting high similarity to *Pseudomonas citronellolis* species.	123
2009	• Characterization of *Pseudomonas* sp. M1 proteome in the presence of β-myrcene (hints about the adaptation response towards β-myrcene)	125
2015	• *Pseudomonas* sp. M1 fully sequenced genome. • Discovery of the 28-kb genetic *locus* necessary for the catabolism of β-myrcene. • *In-silico* prediction of 8 transcriptional units in the 28-kb gene *locus*. • Transcriptomic profile of β-myrcene grown cells.	126
2017	• Experimental validation of the transcriptional units predicted *in-silico*. • Metabolite profile of *Pseudomonas* sp. M1 when challenged with β-myrcene and other monoterpenes.	127

The significance of β-myrcene, a hydrophobic monoterpene, in industrial applications prompted investigations into its degradation pathway in *Pseudomonas* sp. M1. Iurescia and collaborators initiated this journey in 1999, isolating M1 from the Rhine River in Netherlands. Through transposon mutagenesis, the researchers unveiled the *myrA*, *myrB*, *myrC*, and *myrD* genes orchestrating β-myrcene mineralization[124]. Subsequent studies, particularly in 2015, deciphered the complete β-myrcene catabolic pathway and associated genes, shedding light on the genetic machinery for terpene degradation in this bacterium. The genome of *Pseudomonas* sp. M1 underwent full sequencing through Illumina MiSeq technology, complemented by data from other methodologies. Through RNA sequencing, the researchers pinpointed specific gene *loci* crucial for β-myrcene catabolism, revealing 52 genes exhibiting a tenfold

upregulation. Among these, 22 were clustered within a 28-kilobase genomic *locus*. Notably, fifteen of these genes displayed the highest expression levels across the entire transcriptome when cells were exposed to β-myrcene. Employing in-silico predictions, the genomic island (GI) was subdivided into eight transcription units (TUs), a delineation corroborated by mapping RNA-seq reads against the genome corresponding to the 28-kilobase GI (Figure 6).

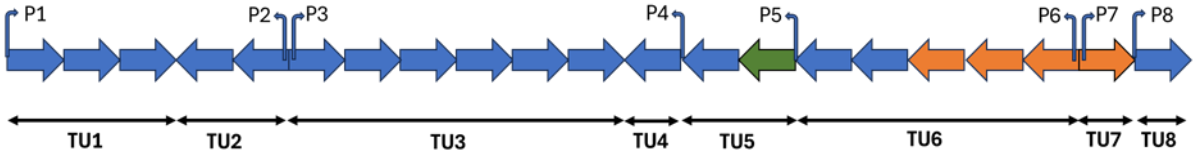


Figure 6 - Organization of the 28-kb genomic *locus* and Transcriptional Units (TU) predicted in silico by P. Soares-Castro and P. M. Santos. β-Myrcene induced the expression of 22 genes in the genomic island of *Pseudomonas* sp. M1. The genes coloured orange represents the gene cluster identified by Iurescia and collaborators, and the gene in green represents the MyrH enzyme.

TUs 1, 3, 5, 6, and 7 comprised catabolic genes, suggesting potential polycistronic transcription, while TUs 4 and 8 were implicated in regulatory functions. TU2 was linked to a chemosensory system likely involved in β-myrcene recognition and uptake. Notably, a monooxygenase within TU5, later identified as a hydroxylase (MyrH), exhibited a significant increase in expression, playing a pivotal role in the initial oxidation step of beta-myrcene into myrcen-8-ol[128]. The authors also proposed the independent evolution of catabolic and regulatory genes within the GI, supported by likelihood scores and phylogenetic analyses. Continued research, performed by Pedro Castro and collaborators in 2017, shed further elucidation into the genetic and functional characteristics of *Pseudomonas* sp. M1. Using a promoter probe engineered to fuse predicted promoters with Green Fluorescent Protein (GFP), the researchers observed the highest fold-change fluorescence levels after 5 hours of β-myrcene induction, particularly with promoters P2, P5, and P7 responsible for the expression of TUs 2, 5, and 7. Subsequently, through targeted transposon mutagenesis aimed at the GI, two mutants — M1C19 and M1C38 — were generated. The promoter activity profile of these mutants revealed significant findings. The M1-C19 mutant, with an interrupted LuxR family regulator gene, exhibited all promoters switched off, evidenced by similar fluorescence ratios between β-myrcene challenged and unchallenged cells, ultimately inhibiting growth in β-myrcene. Conversely, the M1-C38 mutant, featuring inactivated promoters P2, P3, and P6 responsible for regulating certain catabolic steps, demonstrated a loss of the myr+ phenotype, underscoring the essential role of the MyrH protein in β-myrcene catabolism. To validate the functional modules within the GI,

metabolite analysis of the β-myrcene pathway was conducted in both wild-type (*wt*) and mutant strains, including the N22 mutant from previous research by Iurescia and collaborators. In the *wt* strain, myrcen-8-ol was identified as the initial major metabolite after 30 minutes of β-myrcene induction, consistent with the N22 mutant inability to produce subsequent pathway intermediates. Furthermore, myrcenoic acid levels increased over time, accompanied by the appearance of derivatives such as 4-methyl-3-hexenoic acid and 4-methylhexanoic acid after 3.5 hours of growth. Conversely, mutants M1-C19 and M1-C38 failed to produce the expected pathway metabolites, although without impacting cell viability. In-depth analysis of *Pseudomonas* sp. M1 cell membrane lipid profiles following exposure to saturating β-myrcene levels revealed significant changes, primarily involving the cis-to-trans isomerization of unsaturated fatty acids. Additionally, M1 cells accumulated derivatives when challenged with various terpenes beyond β-myrcene, including β-citronellol, Citral, Linalyl acetate, Limonene, and β-pinene (Table 4).

Table 4 - Products accumulated by *Pseudomonas* sp. M1 *wt* during the catabolism of different terpenes. Products identified by GC-MS[12,6].

Substrate	Product
β-Myrcene	Myrcenal
	Myrcen-8-ol
	4-Methylhexanoic acid
	4-Methyl-3-hexenoic acid
	Myrcenoic acid
Linalyl acetate	β-Linalool
	Geraniol
Limonene	Limonene-1,2-diol
β-Pinene	α-Terpineol
	Citronellic acid

Interestingly, different monoterpenes prompted different levels of induction in different transcription units (TUs) within the GI, with β-myrcene exhibiting the highest induction levels among them[129].

Today, with its genome fully sequenced and its biotechnological potential increasingly recognized, *Pseudomonas* sp. M1 stands as a testament to the power of microbial adaptation and genetic resilience in tackling complex environmental challenges. Despite its extensive characterization, efforts persist to classify this remarkable bacterium at the species level, underscoring the ongoing quest to unlock its full potential and expand the frontiers of biotechnology.

1.6. Outline of this thesis

This thesis focuses on the remarkable ability of *Pseudomonas* sp. M1 to catabolize monoterpenes, specifically β-myrcene, producing compounds with high industrial value (terpenoids). The exploitation of this bacterium for biotech applications requires accurate description of its genotypic and phenotypic characteristics. Thus, correct taxonomic classification is crucial for understanding its unique properties and potential applications, as well as to implement the appropriate biosafety handling procedures and facilitating regulatory approvals. Thus, Chapter II of this thesis focuses on the validation of *Pseudomonas* sp. M1 as belonging to the *P. citronellolis* species.

Chapter III of this thesis focuses on enhancing the ability of this bacterium to bio transform terpenes into terpenoids, leading to relevant biotransformation yields for biotechnological applications. This was accomplished though ALE. The holistic approach adopted for this thesis, relating genomics (Chapter III), proteomics (Chapter IV) and methylomics (Chapter V) data, allowed to understand the mechanisms governing the adaptation of this bacterium to β-myrcene, as well as some of the mechanisms regulating mutational events.

In Chapter IV of this thesis, the proteomic characterization of *Pseudomonas* sp. M1 enabled the relative quantification of proteins before and after ALE. This analysis corroborated the improved efficiency of the adapted isolates in catabolizing the monoterpene and provided insights into the protein modifications acquired during the adaptation, thus elucidating the adaptation process occurred in this bacterium.

In Chapter V of this thesis, a preliminary analysis of the methylome dynamics of *Pseudomonas* sp. M1 provided insights into the correlation of genome methylation and mutational events that occurred during the ALE experiment.

The Chapter VI integrates all the knowledge acquired during this thesis raising some pertinent questions regarding the future of *Pseudomonas* sp. M1 as a potential biocatalyst for the production of terpenoids.

Chapter II - Revisiting the taxonomy context of *Pseudomonas* sp. M1 as an isolate with great biotech potential

2.1. Introduction

Pseudomonas sp. M1, initially isolated in 1999 by Iurescia and collaborators, stands as an exemplar among bacteria able to metabolize hydrophobic substrates, including monoterpenes, phenol and benzene[122,123,126,127,130–132]. Notably, its capacity to metabolize monoterpenes, exemplified by its complete mineralization of β-myrcene, highlights its metabolic versatility. The metabolism of phenol and benzene by *Pseudomonas* sp. M1 is possible through the use of analogous metabolic pathways, although with specialized genetic machinery to handle benzoate. To elucidate the phenol degradation pathway, researchers employed random transposon mutagenesis, leading to the identification of genes pivotal for phenol catabolism. The isolation of a phenol-growth-deficient mutant, *Pseudomonas* sp. 22J, which also lost its capacity to use benzene, facilitated the sequencing and subsequent delineation of eight essential ORFs forming the *phc* cluster, governing the upper phenol degradation pathway. Investigations into the regulation of the *phc* cluster involved the construction of four LacZ promoter fusions. Two constructs featured promoters Pk and Pa controlling LacZ expression, while the other two incorporated the *phcR* gene under the control of its native promoter, fused upstream of the Pk and Pa promoter regions along with the *lacZ* gene. The PhcR regulator was found to be necessary, as both promoters exhibited minimal activity in its absence. Moreover, the activity of both promoters significantly increased upon exposure to phenol or benzene, proving the shared genetic machinery for phenol and benzene catabolism in the *Pseudomonas* sp. M1 strain[121].

Furthermore, *Pseudomonas* sp. M1 appears to have similar strains scattered on different soil environments. Despite being initially isolated from the Rhine River sediments in the Netherlands, Pedro Soares-Castro and collaborators reported the isolation of bacterial isolates similar to *Pseudomonas* sp. M1. In this study, different isolates capable of using β-myrcene as only carbon source were isolated from two distinct locations in Portuguese soil and sequenced. The microorganisms isolated encompassed several different species, however nine isolates — UMC76, UMC631, UMC3103, UMC3106, UMC3129, UME83, UMA601, UMA603, and UMA643 — were found to harbor the 28-kb *locus* necessary for the catabolism of β-myrcene present in *Pseudomonas* sp. M1. This *locus* exhibited 100% nucleotide identity and alignment coverage, along with a similar structural organization to the *Pseudomonas* sp. M1 *locus*. These isolates also accumulated the same intermediates as *Pseudomonas* sp. M1, upon identification through gas chromatography coupled with mass spectrometry (GC-MS).

Given the potential biotech applications of *Pseudomonas* sp. M1 and the available curated genomic data, it was decided to clarify its taxonomic classification. Already in 2007, an attempt was made to

taxonomically classify this bacterium using partial 16S sequencing. However, the species assignment criteria were below the threshold to allow a clear species assignment.

2.2. Material and Methods

2.2.1. Phylogenetic analysis – 16S rRNA

The 16S rRNA sequence was retrieved from *Pseudomonas* sp. M1 genome, Genbank accession GCF_022760135.1. The sequence was used in NCBI nucleotide *Blastn* to search for the type-strains with higher identity percentage with M1. The strains retrieved were: *P. citronellolis* DSM50332 (Genbank accession: GCF_004745455.1), *P. delhiensis* RLD-1 (Genbank accession: GCF_900187975.1), *P. jinjuensis* NBRC103047 (Genbank accession: GCF_900103845.1), *P. knackmussii* B13 (Genbank accession: GCF_009911755.1), *P. panipatensis* CCM7469 (Genbank accession: GCF_900099785.1), *P. aeruginosa* DSM50071 (Genbank accession: GCF_000006765.1), *P. humi* CCA1 (Genbank accession: GCF_001748265.1), *P. guariconensis* LMG27394 (Genbank accession: GCF_002864915.1), P. benzenivorans DSM8628 (Genbank accession: GCF_900100495.1), *P. oleovorans* 1225T (Genbank accession: GCF_000297075.2), *P. monteilii* B5 (Genbank accession: GCF_003671975.1), *P. otitidis* PAM1 (Genbank accession: GCF_011397855.1), *P. azotifigens* DSM17556 (Genbank accession: GCF_000425625.1), *P. alcaligenes* NEB585 (Genbank accession: GCF_001597285.1), *P. nitroreducens* LMG21614 (Genbank accession: GCF_029457395.1) and P. mosselii 17497 (Genbank accession: GCF_002736065.1).

2.2.2. Phylogenetic analysis – Multi locus sequencing typing (MLST)

To support the 16S rRNA data, seven genes typically used for strain typing were chosen: *atpD*, *guaA*, *gyrB*, *recD*, *rpoB*, *rpoD* and *recA*. These genes were previously used to stablish the phylogeny of *Pseudomonas* genus[133]. MLST schemes were also reported through the years for *P. aeruginosa*, *P. fluorescens* and *P. putida* [134–136]. This approach proved to be useful and precise. The sequences were retrieved from NCBI from the genomes of the type-strains indicated in the 16S rRNA result. These seven genes were totally sequenced for all the indicated strains. For each strain, the genes were concatenated into 1 single file.

2.2.3. Phylogenetic analysis – Maximum likelihood tree

To construct the phylogenetic tree from 16S rRNA, all the sequences were normalized for the length 1230b.p., since some strains had this gene only partially sequenced. For the MLST genes, normalization was not performed since all genes chosen were totally sequenced in all strains. *MAFFT* version 7[137] was used to align all the sequences followed by phylogenetic tree construction on *MEGA* version 11.0.13[138]. For the tree, 100 bootstrapping[139] was used, as well as kimura-2 parameter[140] as model and maximum likelihood[141] as statistical method, chose upon finding the best model in *MEGA 11*.

2.2.4. Genome-based analysis – Average Nucleotide Identity (ANI) and In-silico DNA-DNA hybridization (DDH)

ANI is the average of the nucleotide identity of orthologous genes shared between two selected genomes. To calculate this parameter, *FastANI* version 2.0[142] was used. This tool uses the algorithm *Mashmap*[143] to locally align and map two selected genomes, finding the orthologous genes between them. Afterwards, the nucleotide identity between those regions is calculated. ANI was calculated between *Pseudomonas* sp. M1 and all the type-strains from *Pseudomonas* genus. The strains with higher ANI value were: *P. humi* CCA1 (Genbank accession: GCF_001748265.1), *P. citronellolis* DSM50332 (Genbank accession: GCF_004745455.1), *P. delhiensis* RLD-1 (Genbank accession: GCF_900187975.1), *P. jinjuensis* NBRC103047 (Genbank accession: GCF_900103845.1), *P. knackmussii* B13 (Genbank accession: GCF_009911755.1) and *P. panipatensis* CCM7469 (Genbank accession: GCF_900099785.1).

To complement ANI values, DNA-DNA hybridization was predicted in-silico using the tool Genome-to-Genome Distance Calculator (GGDC) version 3.0[144]. This tool infers whole-genome distances using different algorithms. The strains with higher DDH value were the ones indicated in ANI calculation. Both ANI and in silico DNA-DNA hybridization were employed according to Margarita Gomila and collaborators which reported the clear separation of different *Pseudomonas* species using these tools[145].

2.2.5. Genome-based analysis – Pan-genome analysis of *Pseudomonas citronellolis*

The pangenome of *Pseudomonas citronellolis* species was constructed using all the available datasets for *P. citronellolis* isolates in NCBI. The genomes of *Pseudomonas* sp. M1 and *P. humi* CCA1 were also added to this analysis in order to verify its relatedness with this species. The total number of sequences used were 27, with the following Genbank accessions: GCF_001446495, GCF_00144699, GCF_001586155, GCF_001654435, GCF_001748265, GCF_002091555, GCF_004745455,

GCF_016009275, GCF_022760135, GCF_024170125, GCF_024170235, GCF_024170345, GCF_024170385, GCF_024170415, GCF_024170445, GCF_024170585, GCF_024170725, GCF_024496105, GCF_025398795, GCF_026625105, GCF_027569775, GCF_029335315, GCF_029335335, GCF_035658415 and GCF_900112375. For this analysis Anvi'o[146] platform was used in the same way as Buqing Yi et al. to construct the pangenome of *Pseudomonas* genus[89].

2.2.6. Metabolic Fingerprint – GENIII MicroPlate™ test panel Biolog

Phenotype microarrays were employed through GENIII MicroPlate™ test panel from Biologs patented redox technology. This technology uses cell respiration (NADH production) as a universal reporter. If the phenotype is strongly "positive", the cells respire actively, reducing the tetrazolium dye and forming a strong color. If it is weakly positive or negative, respiration is slowed or stopped, and the result is less color or no color. Biologs are long used to characterize the metabolic profile of the new species being a precise technology when it comes to metabolic profiles of different species[147]. All the cells were grown firstly overnight in Minimal Media (MM) prepared as described previously by Hartmans et al. (1989)[148] supplemented with lactate. In the morning, the cells were refreshed for the same conditions and grown until an OD_{600nm} of 0.3 for the cells to be in the exponential phase growth. The cells were suspended in the inoculating fluid (IF) provided by the manufacturer in a ratio allowing 95% transmittance. The 96-well plates were inoculated with 100 μl using a multichannel pipette and grown at 30°C, protected from the light. The optical density was measured at 6, 24, 32 and 48 hours in a microplate reader at 590nm to measure the growth kinetics. Afterwards, the area under the curve from all the growth curves was calculated. Statistical analysis was done using two-way *ANOVA* in *GraphPad Prism* version 8.3.0, and Partial Least Squares – Discriminant Analysis (PLS-DA) was done in *R* version 4.0.3[149], tool present in the package mdatools[150] - for the PLS-DA statistics, cross-validation was used since this model may be prone to overfitting – as well as Principal Component Analysis (PCA) in *R* version 4.0.3[14], tool present in package ggbiplot, based on ggplot2 package[151].

2.2.7. Growth evaluation of *Pseudomonas* sp. M1 and *Pseudomonas citronellolis* in benzoate

To evaluate the ability of *Pseudomonas* sp. M1 and *P. citronellolis* cells to grow in benzoate, both strains were grown in this carbon source. Firstly, one colony of each strain was incubated overnight in MM supplemented with lactate and grown at 30°C with agitation. In the morning, the cells were refreshed for

MM supplemented with benzoate in a OD_{600nm} around 0.1 and grown for 24h with measurements each 2 hours.

2.2.8. Verification of phenol metabolism in *Pseudomonas citronellolis* using the β-galactosidase assays

Although *Pseudomonas* sp. M1 is able to catabolize phenol resorting to the *phc* cluster, this feature was not yet verified in *Pseudomonas citronellolis*. The vector pPR9TT containing the fusion between the promoter Pa (regulates the expression of catechol 1,2-dioxygenase – *phcA*, necessary for phenol and benzoate mineralization) and the LacZ promoter probe[1] was inserted into *Pseudomonas citronellolis* through conjugation. The cells were plated in x-gal upon transformation. Afterwards, a miller-units assay was employed for both strains – *P. citronellolis* and *Pseudomonas* sp. M1. Firstly, the cells were grown overnight in Minimal media supplemented with lactate. The overnight cells were refreshed in the morning and grown until an OD_{600nm} 0.6. Upon this growth the cells were refreshed once again to Minimal Media supplemented with lactate or Benzoate (0.01%). After 2h growth, the LacZ assay started. Each 2 hours, the cells were collected and its OD_{600nm} measured. From the inocula, 600 μL of culture was collected and centrifuged to pellet. These cells were resuspended in cold Z-Buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgCl_2$) containing 10 mM DTT and 50 μL 0.1% SDS and 100 μL chloroform were added. The permeabilized cells were incubated for 3 minutes in a 30ºC water bath. Afterwards 200 μL of a ONPG 4 mg/mL solution was added and the cells were incubated in a 30ºC water bath for 10 minutes (until the yellow color appears). At this point, 500 μL of Stop Solution (1 M Sodium carbonate) were added. All the samples were centrifuged for 5 minutes, and the OD_{420nm} was measured in a quartz cuvette. Miller units were calculated following the formula:

1000 x [(OD420 - 1.75 x OD550)] / (T x V x OD600).

2.2.9. Transmission Electron Microscopy (TEM)

For TEM microscopy, *Pseudomonas* sp. M1 cells were negatively stained with 0.5 % uranyl acetate on copper grids, and after air drying the grids were examined with TEM. The cells were harvested at mid-exponential phase. The procedure used was the one usually used for the description of *Pseudomonas* species[152].

2.3. Results and Discussion

2.3.1. *Pseudomonas* sp. M1 morphology

As elucidated by Iurescia and collaborators., *Pseudomonas* sp. M1 exhibits the characteristic traits of an oxidase-positive, gram-negative, motile, rod-shaped bacterium[153]. This description found further validation through transmission electron microscopy (TEM), which revealed a bacillus-like structure with flagella at its extremities, thus reinforcing the established profile of this strain. For the imaging, *Pseudomonas* sp. M1 cells were grown in lactate (Figure 7A) and β-myrcene (Figure 7B). Intriguingly, even in the presence of the hydrophobic monoterpene β-myrcene, the morphology of *Pseudomonas* sp. M1 remained unaltered, as depicted in Figure 7. Despite the distinct properties of these carbon sources, including the hydrophobicity and acidic pH induced by β-myrcene, no significant differences were discernible in the cellular morphology between cultures grown on different carbon substrates.

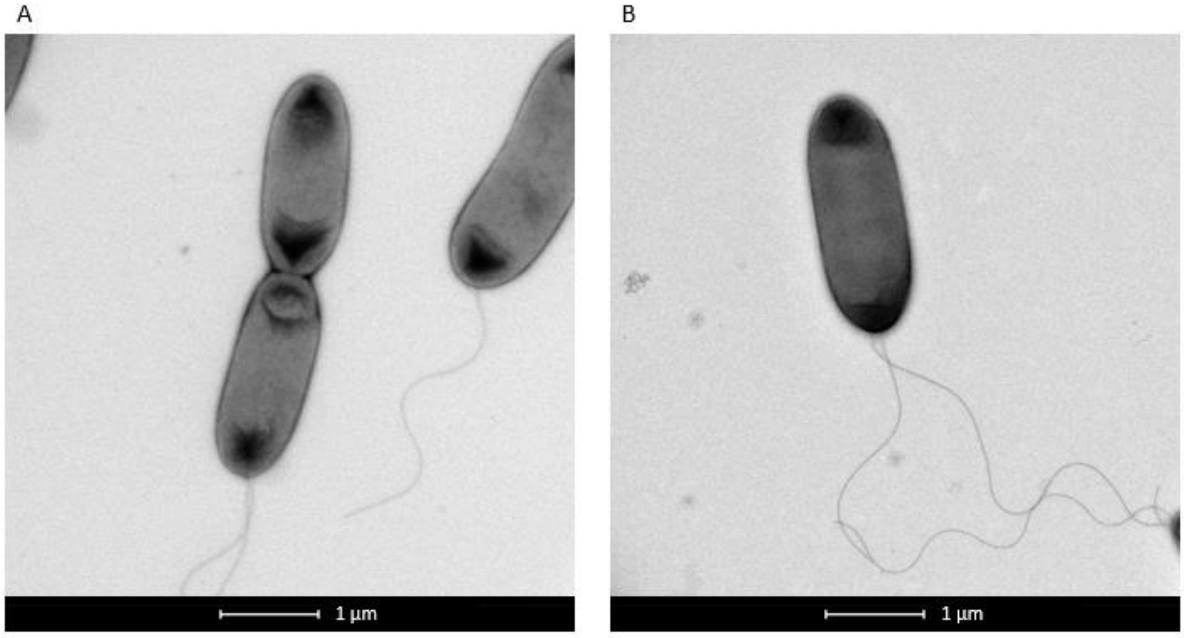


Figure 7 - TEM microscopy of *Pseudomonas* sp. M1. The cells were imaged after growth in lactate (A) and β-myrcene (B). Rod-shaped gram-negative bacteria, with flagella granting motility.

Collectively, these observations highlighted the conformity of *Pseudomonas* sp. M1 morphology with the typical attributes observed across the *Pseudomonas* genus[154].

2.3.2. M1 strain is closely related to *P. humi* and *P. citronellolis* through phylogenetic analysis

As previously indicated, a 2007 analysis based on 16S rRNA sequences suggested the potential connection between *Pseudomonas* sp. M1 and *Pseudomonas citronellolis*. While 16S rRNA analysis

remains fundamental in taxonomic classification, advancements in genomics have introduced additional techniques such as Multi-Locus Sequence Typing (MLST), Average Nucleotide Identity (ANI), *in silico* DNA-DNA hybridization (DDH), and pangenome analysis. These methodologies enhance the accuracy of species classification[155].

Initially, the complete 16S rRNA gene of *Pseudomonas* sp. M1 was utilized to identify closely related type strains using the NCBI nucleotide *Blast* tool (*Blastn*). The retrieved type strains, all belonging to the *Pseudomonas* genus as expected, exhibited identity percentages ranging from 99.3% to 95.3%. Notably, *Pseudomonas citronellolis* emerged as the most closely related species with 99.3% identity, followed by *Pseudomonas humi* with 99.1% identity (Figure 8A). In tandem with this analysis, MLST analysis was conducted using the 16S rRNA gene sequence and seven sequences of highly conserved genes (*atpD*, *guaA*, *gyrB*, *recD*, *rpoB*, *rpoD*, and *recA*). These genes, characterized by relatively slow evolution rates, were publicly available and completely sequenced for all type strains closely related to the *Pseudomonas* sp. M1 strain identified through 16S rRNA analysis. Sequence alignment was performed using *MAFFT*, and phylogenetic trees were constructed using the kimura-2-parameter model with 100 bootstrapping and neighbor-joining statistical method, yielding results closely aligned with reality (Figure 8B).

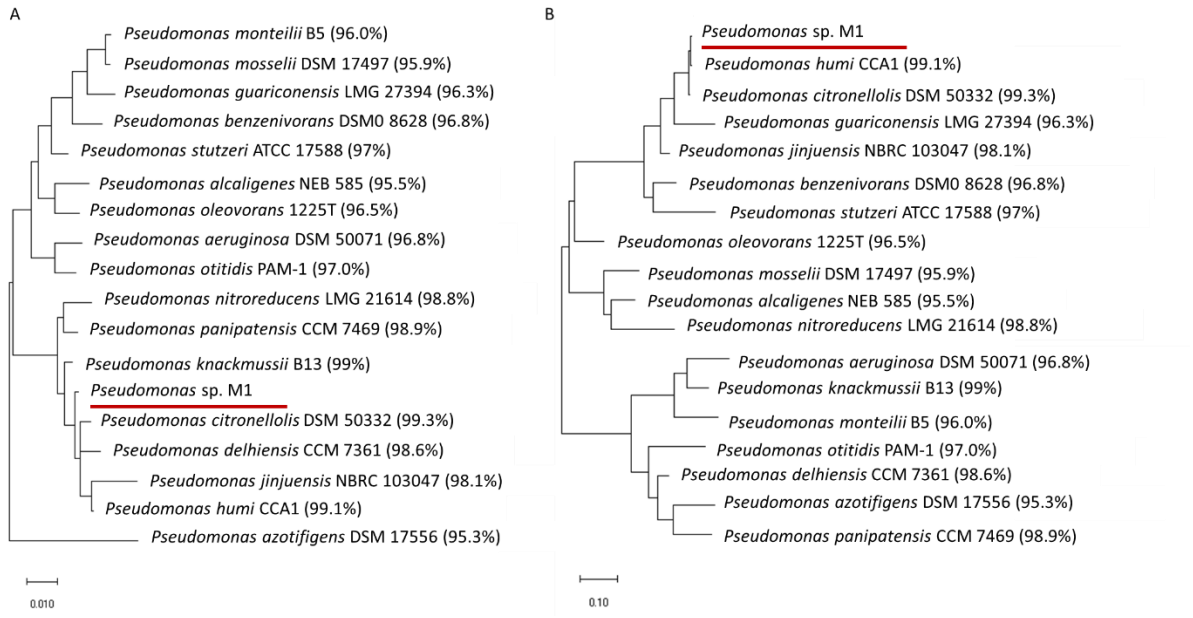


Figure 8 - 16S rRNA phylogenetic tree (A) and MLST reference genes (B) for the type-strains with higher identity with *Pseudomonas* sp. M1 16S rRNA sequence. The identity of the 16S rRNA sequences to M1 are indicated in the brackets. The sequences were aligned using *MAFFT* and the trees calculated through kimura-2 parameter model, using neighbor-joining and bootstrapping. Sequences were retrieved from the genomes of each strain from GenBank.

Phylogenetic analysis based on both the 16S rRNA gene and MLST genes revealed the close relationship of the M1 strain with *Pseudomonas humi* and *Pseudomonas citronellolis* type strains. Taking into

consideration the advancements in sequencing technologies, the classification of several strains may undergo fluctuations. It has been reported that approximately 25% of *Pseudomonas* genomes cataloged in GenBank are inaccurately classified, largely due to the isolation of strains preceding taxonomy updates aided by novel methodologies[145]. Therefore, it is imperative for this study to reevaluate the taxonomy of the type strains most closely associated with the *Pseudomonas* sp. M1 strain.

To determine the correlation among the identified type strains, the Average Nucleotide Identity (ANI) was computed. ANI serves as a crucial metric for assessing the genetic relatedness or similarity between bacterial or archaeal genomes. By delineating conserved regions across all genomes through mapping, ANI calculates the nucleotide identity base by base, position by position, for these shared segments. A threshold of 95% is conventionally employed to indicate a taxonomic relationship between strains. If the ANI value falls below this cutoff, the strain under analysis may represent a novel species[143,157]. Moreover, researchers showed that *FastANI* (the tool used in this thesis) estimates were accurate for draft genomes, in the range of 20–100% completeness, suggesting that FastANI can tolerate variable assembly quality and completeness[143].

The analysis revealed that all strains considered indeed belonged to distinct species, with the exception of *Pseudomonas humi* CCA1. Notably, when *P. citronellolis* served as the reference, *P. humi* failed to meet the threshold (Figure 9).

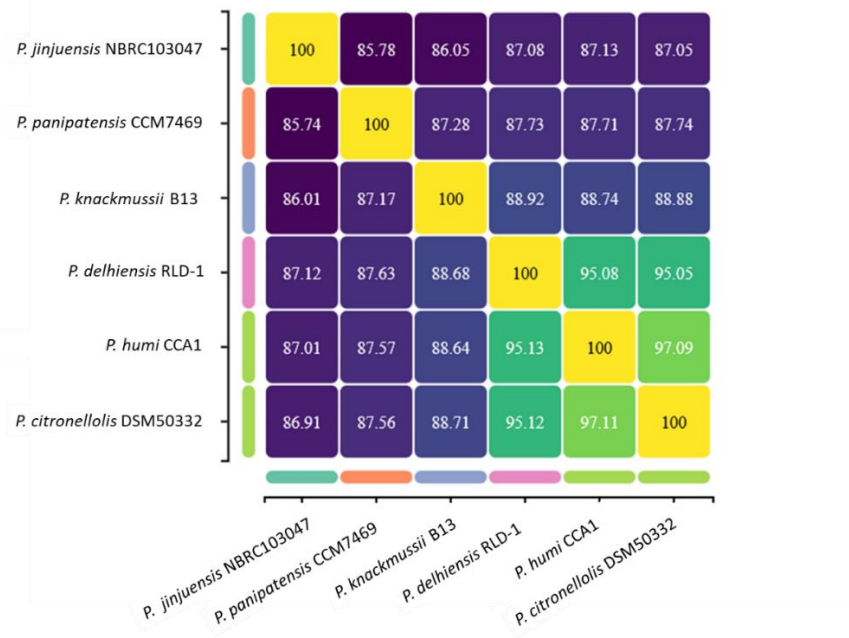


Figure 9 - Genome distance of all *Pseudomonas* strains closely related to *Pseudomonas* sp. M1 through 16S rRNA analysis, calculated with *FASTANI*. *Pseudomonas humi* CCA1 is above the cut-off value (95%) regarding *Pseudomonas citronellolis* type-strain, with 97.11% identity.

Consequently, it is reasonable to consider *P. humi* as a strain belonging to *P. citronellolis* species, given that the latter was named first and remains the most extensively studied and characterized among the two. Therefore, it follows that the *Pseudomonas* sp. M1 likely falls within the *Pseudomonas citronellolis* species.

To refine the classification of *Pseudomonas* sp. M1, *in silico* calculations of Average Nucleotide Identity (ANI) and DNA-DNA Hybridization (DDH) values were estimated. These analyses were performed between the type strains most closely associated with *Pseudomonas* sp. M1 as indicated by 16S rRNA and MLST analyses, namely *P. citronellolis*, *P. delhiensis*, *P. kanackmussii*, and P. *panipatensis*, excluding *P. humi* as this strain seemed to be misclassified. For DDH value determination, the Genome-to-Genome Distance Calculator (GGDC) was employed. This tool identifies orthologous genes between the compared genomes and assesses the similarity of these genes. The established threshold value for DDH analysis is typically set at 70%.

Table 5 - Genome distance of *Pseudomonas* sp. M1 strain against the 5 most similar strains. Calculated through *FastANI* and *GBDP*.

Method	***P. citronellolis*** **DSM50332**	***P. delhiensis*** **RLD-1**	***P. kanackmussii*** **B13**	***P. panipatensis*** **CM7469**
		ANI		
Identity (%)	97.1	95.1	88.8	87.7
Aligned fraction (%)	81.9	87.1	72.4	74.5
		GBDP		
Genome distance	0.0324	0.0564	0.1268	0.1364
Estimated DDH	69.7 - 75.6%	54.6 - 60.1%	30.6 - 35.6%	28.7 - 33.6%

The ANI calculated for *Pseudomonas* sp. M1 when compared to *P. citronellolis* showed an identity of 97.1% and DDH values ranging from 70% to 76% when compared to *P. citronellolis*. When examining its comparison with all the other strains, the ANI and DDH values fell below the established thresholds of 95% and 70%, respectively (Table 5), providing additional evidence for its closeness with the *P. citronellolis* strain. Consequently, congruent with the findings of the phylogenetic analysis, the *Pseudomonas* sp. M1 appears to be a member of the *Pseudomonas citronellolis* species.

2.3.3. Pan-genome of *Pseudomonas citronellolis*

The pan-genome can be described as the entire set of genes present in a determined group of representative strains within a genus or a species[157]. It usually comprises the core genes – the genes shared between all the individuals of a species or genus – the accessory genes – genes shared by some of the individuals (not all) – and unique genes (singletons) – genes present only in one strain. This approach allows to unveil the genetic diversity within a species since a single reference genome cannot represent all the species patterns by itself[87]. *Anvi'o* is a bioinformatics platform that allows the analysis of data from microbiology, including the construction of pangenomes. In its pipeline several tools are used, including *Diamond*[158] and/or *Blastp* to align the protein sequencies from all the genomes against each other. The Markov Cluster Algorithm (MCL)[159], used in this pipeline, allows to identify similar clusters between all the sequences retrieved, thus helping to infer the evolutionary relationships between all the genomes. For this analysis, the 27 genomes available corresponding to the *Pseudomonas citronellolis* species were retrieved, as well as the genomes of *Pseudomonas* sp. M1 and *Pseudomonas humi* CCA1. The 27 genomes were run in *Anvi'o* in order to determine the pangenome of *Pseudomonas citronellolis* species and better correlate *Pseudomonas* sp. M1 and *Pseudomonas humi* to this species. In this analysis, a total of 2725 genes were identified within the core genome, while 16000 genes comprised the accessory genome, with an additional 3455 genes categorized as singletons. Notably, among the isolates examined, *P. citronellolis* SJTE-3 exhibited the highest number of singletons, with 892 genes, followed closely by *P. citronellolis* GW210018_S65, featuring 884 unique genes. Conversely, *Pseudomonas humi* CCA1 and *Pseudomonas* sp. M1 displayed a relatively lower count of singletons, with 271 and 181 genes, respectively. This disparity in singleton counts between *Pseudomonas humi* CCA1, *Pseudomonas* sp. M1, and other *P. citronellolis* strains underscores their classification within the *Pseudomonas citronellolis* species. The diminished presence of exclusive genes in these strains further strengthens their association with *Pseudomonas citronellolis*, as the observed genetic variation is comparatively limited. Moreover, the genes constituting the core genome of the *Pseudomonas citronellolis* species were employed to establish a phylogenetic relationship among all 27 strains (Figure 10).

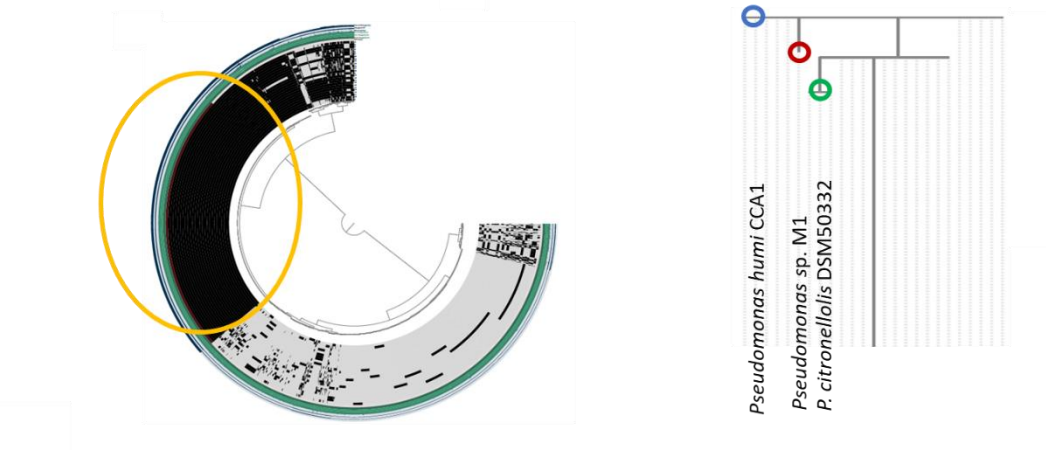


Figure 10 - Pangenome of *Pseudomonas citronellolis* species. The core genome region is circled in yellow. The phylogenetic tree was constructed with *FastTree* and alignment of sequenced was performed with muscle (these were the options available during the *Anvi'o* workflow), taking into consideration only the core genome region. The resulting phylogenetic tree, based on the core genome of the *Pseudomonas citronellolis* species, showed the degree of similarity between *Pseudomonas* sp. M1 (red), *Pseudomonas citronellolis* DMS50332, the type-strain of this species (green) and *Pseudomonas humi* CCA1 (blue).

In *Anvi'o* the phylogenetic tree was constructed resorting to *muscle*[160] for the alignment of the sequences chosen and *FastTree* for the construction of maximum-likelihood tree for large alignments[161].

Considering the most conserved cluster within the pangenome of the *Pseudomonas citronellolis* species — the core genome, encompassing genes shared by all strains — the core genome of *Pseudomonas* sp. M1 aligns with those of other *Pseudomonas citronellolis* strains. This alignment extends to the sequence of the type-strain, *Pseudomonas citronellolis* DSM50332, further reinforcing the classification of *Pseudomonas* sp. M1 as belonging to the *Pseudomonas citronellolis* species. Nevertheless, *Pseudomonas* sp. M1 occupies a distinct branch within the phylogenetic tree, suggesting a potential speciation event unique to this strain. Additionally, *Pseudomonas humi* CCA1 shares the same branch as several other *P. citronellolis* strains, further reinforcing its classification within *the Pseudomonas citronellolis* species. Focusing on the Accessory genome, two distinct clusters emerged within the pangenome of *Pseudomonas citronellolis* species, as depicted in Figure 11A and B. In Accessory genome A, a greater number of genes were shared among all strains.

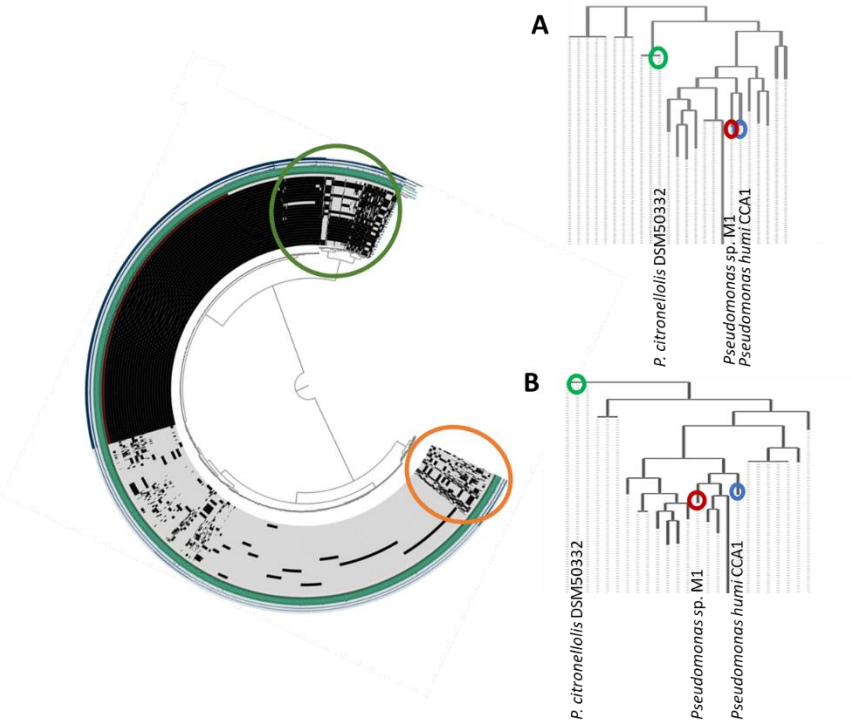


Figure 11 - Pangenome of *Pseudomonas citronellolis* species. The accessory genome was divided into 2 subregions represented in A (circled in green) and B (circled in orange). Region A comprises 12800 genes, whilst the region B comprises 3200 genes. The phylogenetic tree was constructed with *FastTree* and alignment of sequenced was performed with muscle (these were the options available during the *Anvi'o* workflow), taking into consideration the respective accessory genome clusters. The resulting phylogenetic trees, based on each accessory genome of the *Pseudomonas citronellolis* species, showed the degree of similarity between *Pseudomonas* sp. M1 (red), *Pseudomonas citronellolis* DMS50332, the type-strain of this species (green) and *Pseudomonas humi* CCA1 (blue).

The subsequent phylogenetic analysis revealed a clustering of *Pseudomonas* sp. M1 and *Pseudomonas humi* within the same clade, while *Pseudomonas citronellolis* DSM50332 was grouped in a different clade. In contrast, Accessory genome B showcased fewer common genes among genomes, resulting in all three strains being grouped into different clades. However, *Pseudomonas* sp. M1 and *P. humi* CCA1 exhibited closer proximity to each other than to *Pseudomonas citronellolis* DSM50332. Thus, in the context of the accessory genome — comprising genes present only in some strains of the *Pseudomonas citronellolis* species — both *Pseudomonas humi* CCA1 and *Pseudomonas* sp. M1 align with this species, also showing significant similarity between them.

Despite exhibiting some variances in comparison to *Pseudomonas humi* CCA1 and *Pseudomonas citronellolis* DSM50332, *Pseudomonas* sp. M1 maintains a close affinity to both strains. Notably, *Pseudomonas* sp. M1 appears to lean closer to *Pseudomonas humi* CCA1, a trend previously observed. However, uncertainty arose regarding the accuracy of *Pseudomonas humi* CCA1 taxonomic classification.

Despite this uncertainty, *Pseudomonas humi* CCA1 displayed a notably low number of singletons, even fewer than isolates already classified as belonging to the *P. citronellolis* species.

Furthermore, the minimal number of singletons identified in the genome of *Pseudomonas* sp. M1 when clustered with all available *P. citronellolis* genomes indicates its affiliation with this species. Moreover, the *Pseudomonas citronellolis* species appears to be remarkably diverse, a characteristic noted as early as 2003 by Dhruva Bhattacharya and collaborators[162]. This diversity could potentially arise from divergent evolution, possibly driven by environmental constraints. While such divergence may eventually culminate in speciation, the current genomic profile of *Pseudomonas* sp. M1 does not sufficiently deviate to guarantee its separation from the *P. citronellolis* species[163].

2.3.4. Comparison of metabolic profiles between *Pseudomonas* sp. M1 and *P. citronellolis*

While *Pseudomonas* sp. M1 seems to belong to the *P. citronellolis* species, notable metabolic differences have been observed compared to other *P. citronellolis* strains. While *P. citronellolis* DSM50332 can metabolize monoterpenes such as citronellol (from which it was initially isolated) and geraniol, it is unable to grow on β-myrcene[164]. Conversely, *Pseudomonas* sp. M1 cannot grow in citronellol. *Pseudomonas* sp. M1 metabolizes β-myrcene through β-oxidation thanks to the 28-kb genomic *locus* previously referred, whereas *P. citronellolis* DSM50332 uses the *Atu* pathway to metabolize citronellol and other terpenes. Hence, *P. citronellolis* DSM50332 lacks the genetic machinery to catabolize β-myrcene, just as *Pseudomonas* sp. M1 cells lack the machinery to catabolize citronellol[128,164]. However, the metabolic disparities between the two strains extend beyond monoterpenes. When cultured in the organic acid lactate and monosaccharide glucose, both strains exhibit distinct specific growth rates. To further elucidate the metabolic variances between *Pseudomonas* sp. M1 and *P. citronellolis* DSM50332, GenIII Biolog plates — a microbial identification test panel – were employed. Biolog plates have long been utilized for strain characterization and identification. When a strain can grow on different carbon sources and chemicals present in each well of the 96-well plate, the cells respire, causing a color change in the dye present (tetrazolium blue). This well-established method allows for the establishment of a "phenotypic fingerprint" for each strain, facilitating strain-to-strain comparisons. By employing partial least squares discriminant analysis (PLS-DA) and principal component analysis (PCA), it was possible to distinctly separate *P. citronellolis* DSM50332 and *Pseudomonas* sp. M1. The carbon sources contributing most to this separation included Tween 40, Glycyl-L-Proline, D-Gluconic Acid, Glycerol, D-Aspartic Acid, D-Glucose-6-PO4, L-Glutamic Acid, L-Histidine, D-Malic Acid, D-Serine, and Formic Acid. Furthermore, the

covariance attributed to this distribution for the PLS-DA analysis was only 30.9% for component 1 and 37.34% for component 2, and for the PCA analysis was 41.6% for component 1 and 20.9% for component 2. This indicates that while it is possible to phenotypically distinguish between both strains, the difference is not substantially significant (Figure 12).

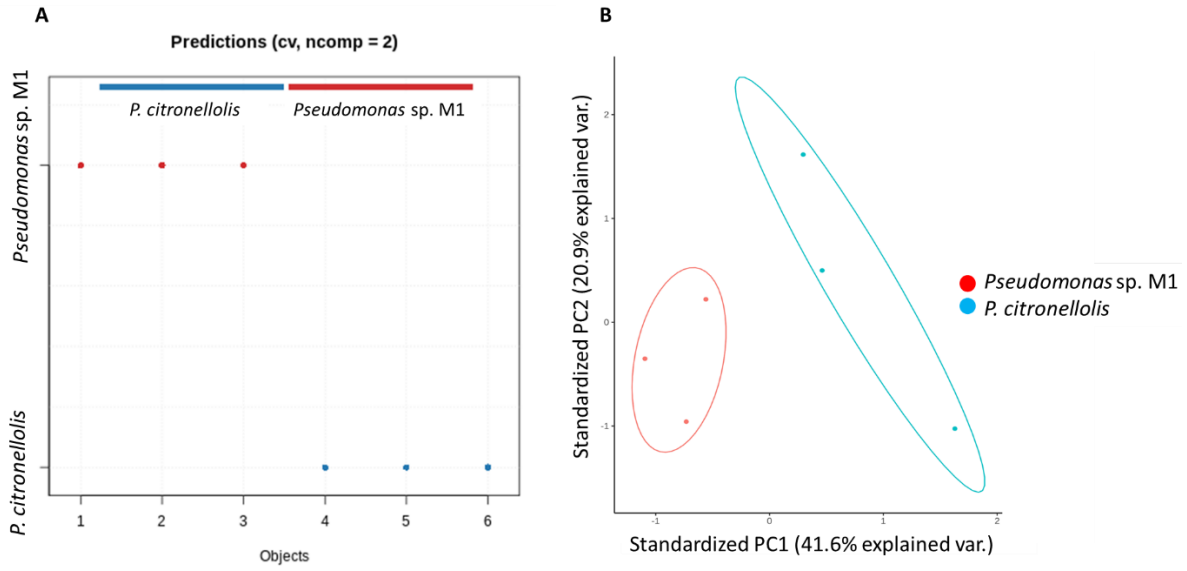


Figure 12 - PLS-DA model (A) and PCA (B) calculated in R showing the differences regarding the metabolism between *Pseudomonas* sp. M1 and *P. citronellolis*.

Although the genomic differences between both strains are not enough to consider *Pseudomonas* sp. M1 as a new species, there are some metabolic differences. However, considering the diversity among the *P. citronellolis* species, the differences in phenotype are not enough to characterize *Pseudomonas* sp. M1 different from *P. citronellolis*.

2.3.5. The genomic machinery to metabolize phenol/benzoate is shared between M1 and *P. citronellolis*

P. citronellolis DSM50332 shares the capacity to metabolize highly hydrophobic substrates, such as phenol, n-dodecane, or naphthalene, akin to *Pseudomonas* sp. M1[123,162,164–166]. However, the precise pathways used by *P. citronellolis* DSM50332 remain elusive. To ascertain the ability of *P. citronellolis* DSM50332 to grow on benzoate, a growth assay was conducted, monitoring the growth of both *Pseudomonas* sp. M1 and *P. citronellolis* DSM50332 in minimal medium supplemented with benzoate over a 24-hour period (Figure 13A). Notably, the specific growth rate when utilizing benzoate as the sole carbon source was found to be similar between *Pseudomonas* sp. M1 and *P. citronellolis* DSM50332.

As previously noted, benzoate induces the activation of genes responsible for phenol catabolism in *Pseudomonas* sp. M1. To investigate whether the catabolic pathways utilized for these compounds are analogous between both strains, the vector pPR9TT containing the fusion between the promoter Pa (which regulates the expression of catechol 1,2-dioxygenase – *phcA*, necessary for phenol and benzoate mineralization) and the LacZ promoter probe, was introduced into the *P. citronellolis* DSM50332 strain.

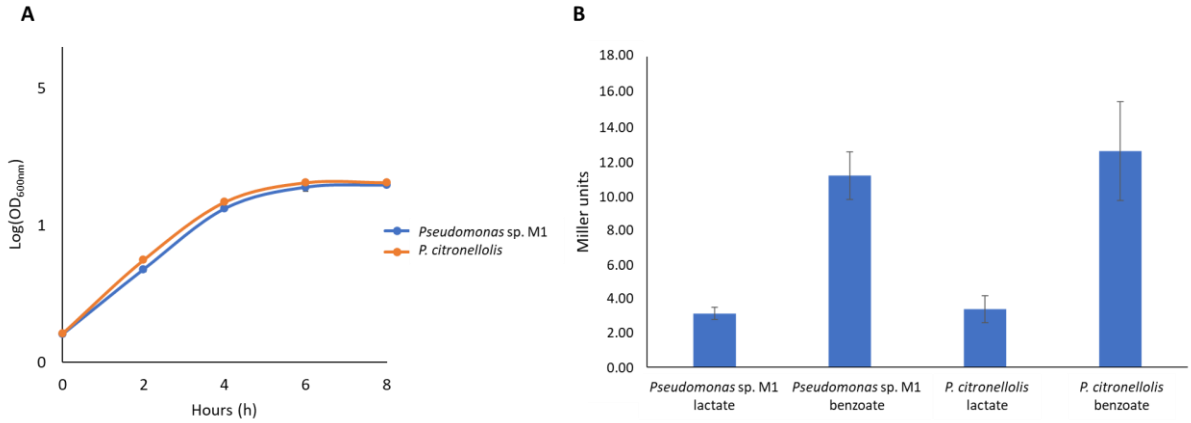


Figure 13 - Growth assay (A) and LacZ assay time-course (B) for *Pseudomonas* sp. M1 and *P. citronellolis* DSM50332 bacteria in Benzoate. The specific growth rate was 0.55h^{-1} and 0.62h^{-1} respectively. For the LacZ assay, *Pseudomonas* sp. M1 and *P. citronellolis* were conjugated with the vector pPR9TT containing the fusion between the promoter Pa (regulates the expression of catechol 1,2-dioxygenase – PhcA, necessary for phenol and benzoate mineralization) and the LacZ promoter probe. The time-point chosen for this assay was 24h. Data is calculated in Miller units. The experiment was performed until 24h since previous time-points did not have enough signal to be measured. In fact, the promoter Pa is very much induced by phenol, but when Benzoate is being used, the signal will appear after some hours of accumulation, since this compound is also consumed by other pathways, not having a very strong inducing effect.

If the promoter Pa of *Pseudomonas* sp. M1 has an analogous in *P. citronellolis* DSM50332, the LacZ operon will be expressed, and the cells will acquire a blue color in the presence of x-gal, when the cells are grown in benzoate. In fact, the *P. citronellolis* DSM50332 transformed acquired a blue color, showing that the genetic machinery used by *Pseudomonas* sp. M1 and *P. citronellolis* DSM50332 for the catabolism of benzene/phenol are similar (Figure 13B). Moreover, the LacZ signal after 24h of accumulation was similar for *Pseudomonas* sp. M1 and *P. citronellolis* DSM50332.

2.3.6. *Pseudomonas* sp. M1 is spread across the soil

As referred previously, Pedro Soares-Castro and collaborators unveiled a breakthrough in microbial isolation, announcing the isolation of several strains capable of using β-myrcene as a carbon source from

soil samples. Among these isolates, nine designated as UMC76, UMC631, UMC3103, UMC3106, UMC3129, UME83, UMA601, UMA603, and UMA643, stood out for harboring the 28-kilobase genomic *locus* originally found in *Pseudomonas* sp. M1 strain. Remarkably, these isolates exhibited a striking resemblance to the *Pseudomonas* sp. M1 *locus*, showing 100% nucleotide identity and alignment coverage, along with a similar structural organization. Additionally, they displayed an identical profile of accumulated intermediates as *Pseudomonas* sp. M1 upon identification via gas chromatography coupled with mass spectrometry (GC-MS)[167].

To confirm the identity of these isolates as *Pseudomonas* sp. M1, ANI and DDH calculations were performed, revealing nearly 100% similarity to the *Pseudomonas* sp. M1 genome (Table 6). This affirmed the lineage of these isolates as *Pseudomonas* sp. M1. Furthermore, phenotypical tests using Biolog plates were conducted on the nine isolates, yielding results remarkably akin to those of *Pseudomonas* sp. M1. Subsequent PLS-DA grouped these isolates together with *Pseudomonas* sp. M1, further underscoring their close relationship. The discovery that these isolates, originated from two distinct locations in Portugal, while *Pseudomonas* sp. M1 was initially isolated in the Netherlands, suggests that *Pseudomonas* sp. M1 has disseminated across various soil environments.

Table 6 - Genome distance of *Pseudomonas* sp. M1 strain against the soil isolates collected by Pedro Soares-Castro and collaborators. Calculated through *FastANI* and *GBDP*.

	***Pseudomonas* sp. M1 - Like**								
Method	**UMC76**	**UME83**	**UMA601**	**UMA603**	**UMA643**	**UMC631**	**UMC3103**	**UMC3106**	**UMC3129**
	ANI								
Identity (%)	100.0	100.0	100.0	100.0	100.0	99.9	100.0	99.8	100.0
Aligned fraction (%)	98.2	98.7	97.9	97.9	97.2	97.3	97.5	96.0	98.1
	GBDP								
Genome distance	0.0027	0.0021	0.0053	0.0063	0.0084	0.009	0.0074	0.019	0.0051
Estimated DDH	100%	100%	99.9-100%	99.9-100%	99.8-100%	99.8-99.9%	99.8-100%	99.2-99.8%	99.9-100%

The presence of numerous individuals closely resembling *Pseudomonas* sp. M1, both phenotypically and genotypically, may prompt speculation that *Pseudomonas* sp. M1 is a subspecies of *P. citronellolis*. While the grouping with the *Pseudomonas* sp. M1 strain via PLS-DA was feasible, the co-variance values for this clustering floated around 30% for component 1 and 33% for component 2. These values, considered low, underscore the striking similarity between *Pseudomonas* sp. M1 and *P. citronellolis*. Moreover, the

phylogeny of these isolates, using the 16S rRNA sequence, showed the clustering of these soil isolates with *Pseudomonas* sp. M1 in the same clade (Figure 14).

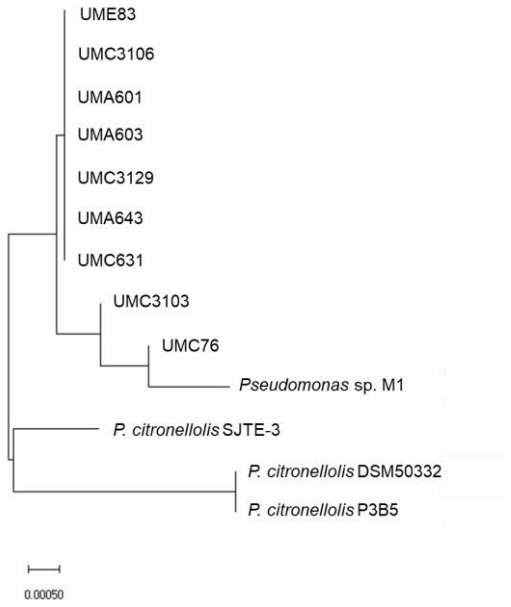


Figure 14 - 16S rRNA phylogenetic tree for the soil isolates similar to *Pseudomonas* sp. M1 and 3 different *P. citronellolis* strains. The sequences were aligned using *MAFFT* and the trees calculated through kimura-2 parameter model, using neighbor-joining and bootstrapping. Sequences were retrieved from the genomes of each strain from GenBank.

Despite these observations, several analyses — ranging from phylogenetic assessments through 16S rRNA gene, MLST, whole-genome approaches (ANI and in-silico DDH), to pangenome analysis — converged on the conclusion that *Pseudomonas* sp. M1 belongs to the *Pseudomonas citronellolis* species. Despite disparities in metabolism, notably *Pseudomonas* sp. M1 capacity to catabolize β-myrcene, and the distinct metabolic fingerprinting established for this bacterium compared to *P. citronellolis*, the genome of *Pseudomonas* sp. M1 bears significant resemblance to that of *P. citronellolis*, failing to meet the criteria for designation as a new species.

In summary, *Pseudomonas* sp. M1 exhibits typical characteristics of a gram-negative, rod-shaped, motile bacterium. Its genome encompasses all the requisite genes for detecting, importing, and catabolizing hydrophobic substrates, particularly showcasing the ability to completely mineralize β-myrcene. The genotype enabling the catabolism of the monoterpene β-myrcene is unique to *Pseudomonas* sp. M1 alone and it is dispersed throughout the soil in other individuals that share 100% genomic similarity with *Pseudomonas* sp. M1 upon comparison. Notably, none of the individuals within the *P. citronellolis* species possess this ability, as they lack the genomic *locus* associated with β-myrcene catabolism. Through rigorous phylogenetic and genome-based in-silico analyses, this bacterium has been firmly classified as

belonging to the *Pseudomonas citronellolis* species, confirming the results from 2007. While *Pseudomonas* sp. M1 genome exhibits some deviations from that of *P. citronellolis*, they fall short of the threshold required for designation as a new species. Moreover, phenotypical differences in core metabolism between *Pseudomonas* sp. M1 and *P. citronellolis* DSM50332 were discerned through the Biolog GENIII plate test. These disparities were also evident among individuals isolated from the soil, resulting in their clustering with the *Pseudomonas* sp. M1 strain following statistical analyses using PLS-DA. However, these phenotypical distinctions do not suffice to classify *Pseudomonas* sp. M1 as a distinct species from *P. citronellolis*. Therefore, *Pseudomonas* sp. M1 belongs to the *Pseudomonas citronellolis* species, confirming the previous suspicions.

Chapter III - Improvement of *Pseudomonas* sp. M1 efficiency towards the catabolism of β-myrcene – Adaptive Laboratorial Evolution (ALE)

3.1. Introduction

Microorganisms can adapt to environmental changes such as exposition to recalcitrant environments and new carbon sources though adaptive evolution[168–172]. In these cases, a strong selective pressure is being applied to the organisms, which will influence their adaptation[173]. One way to adapt is through transient modifications, for instance changing genes expression, without any modifications in its genome[174]. When the selection pressure is applied for enough time, mutations start to accumulate in the genome of the microorganism. Mutations arise spontaneously in the genome and are dependent on the genomic background of the organism. For instance, when a genome is rich in repetitive regions and mobile genetic elements, the number and type of mutations arising in this genome will be different from a genome poor in such elements[175]. When in the presence of a selective pressure, the individuals with accumulated mutations which confer a growth advantage in that environment, will persevere[176,177]. Thus, in a population, several lineages with different mutations in its genome will arise, with different adaptations towards the environment imposed. During evolution, the mutations arising in a population are categorized as beneficial, neutral, deleterious or lethal. Usually, the higher rate of mutations seems to be neutral, while deleterious mutations appear at a much higher rate compared to the beneficial ones. However, when a selection pressure is applied, deleterious mutations tend to be purged overtime, being the adaptation driven by beneficial mutations[178]. When a beneficial mutation arises in a population, the individuals carrying this mutation will grow and reproduce at a higher rate, thus increasing the frequency of this mutation in the population. In asexual populations such as bacteria, as many beneficial mutations arise, its fixation rate is decreased. This happens because several lineages each containing its own beneficial mutations will compete against each other and some beneficial mutations will be lost overtime, thus slowing the accumulation of beneficial mutations to the new populations, in a phenomenon called clonal interference[179]. Evolution is governed by several processes. In a phenom called hitchhiking, neutral or deleterious mutations may also have its frequency increased if linked to the allele with increased frequency (beneficial mutation)[180].

The study of microbial populations in the laboratory are being often used to test evolutionary theories. The older evolution experiment known was started by Lenski and collaborators in 1988 and used *E.coli* as subject. This experiment is still ongoing nowadays[181]. These kinds of experiments start with a culture where only one single colony is cultivated. When the cells reach a high density, the cultures are diluted, and the cell division continues. This cycle is repeated over several generations, allowing the populations to accumulate mutations in their genome. The growth conditions used in these experiments are usually selective, allowing the adaptation of the microorganism towards the environment imposed[182]. The main

advantages of using microorganisms are its short generation time, easy propagation, large population size, easy manipulation of environmental variables and smaller genomes that simplify the analysis and understanding of the adaptation mechanisms[183]. The most common procedure for adaptative laboratorial evolution (ALE) experiments in microorganisms is the serial batch cultivation, where the microorganisms are serial transferred into shake flasks after a fixed period of growth, until reaching the desired generations. The duration of an ALE experiment should be enough for the populations to reach the desired fitness. After that point, if any stronger selection is added, the populations seem to reach an evolutive plateau[184,185]. Usually, several replicates evolving for the same conditions (independently evolving cultures) are used to confer some statistical power to the experiment and aid in the detection of evolutionary changes, difficult to access in only some hundreds of generations, in a process called parallel evolution. The independent evolution of a lineage will provide information about the contribution of beneficial mutations to the evolution. It is expected that independently evolving cultures will accumulate similar beneficial mutations to adapt towards the same environment[50,186,187].

Laboratorial evolution experiments are nowadays being employed in the biotechnology field both to understand how organisms evolve genetically and metabolically and to optimize strains able to produce value-added compounds important in industry (e.g. ethanol, butanol and lipids) as cell factories. This was already employed in several strains, including *E.coli* and *Pseudomonas putida* KT2440[59,188–190].

The study subject of this work is *Pseudomonas* sp. M1, a bacterium known for its biotechnological relevance. This strain can mineralize β-myrcene thanks to the presence of a 28-kb genomic *locus* coding for all the enzymes necessary in this pathway. During the metabolism of this compound, *Pseudomonas* sp. M1 produces several intermediates relevant for industry. In particular, GC-MS analysis of the metabolites produced by *Pseudomonas* sp. M1 and a derived strain (mutated in one of the catabolic genes associated with β-myrcene catabolism) evidenced the effective production and accumulation of industrially relevant intermediates (e.g. the myrcen-8-ol used in scent formulations and the pheromone-like 4-methylhexanoic)[130]. Thus, the purpose of this chapter is to improve the mineralization of β-myrcene by *Pseudomonas* sp. M1 through ALE and understand the contribution of permanent and transient mechanisms for the adaptation of M1 strain towards the monoterpene.

3.2. Material and methods

3.2.1. Bacterial strains and culture conditions

E.coli S17.1 λpir, *E.coli* DH5α and *E.coli* CC118, as well as, *Pseudomonas* sp. M1 strains were routinely maintained in Lysogenic Broth Lennox (LB) medium plates (1% Bacto-tryptone, 0.5% Yeast extract, 0.5%

NaCl, 2% Agar) – liquid or solid – supplemented with antibiotic (gentamicin: 30 μg/mL; kanamycin: 50 μg/mL and streptomycin: 100 μg/mL), when required.

For the laboratorial evolution experiment, *Pseudomonas* sp. M1 wild-type (*wt*) were plated in Luria Bertani Broth (LB) plates (1% Bacto-tryptone, 0.5% Yeast extract, 0.5% NaCl, 2% Agar) and a single colony was used to start the experiment, to decrease the ancestor genimic background. Afterwards, M1 cells were cultivated in 50 mL of mineral medium (also called minimal medium) (MM: 8.9 mM K_2HPO_4, 6.2 mM NaH_2PO_4, 34.2 μM EDTA, 7 μM $ZnSO_4$, 6.8 μM $CaCl_2$, 18 μM $FeSO_4$, 0.8 μM Na_2MoO_4, 0.7 μM $CuSO_4$, 1.7 μM $CoCl_2$, 1.9 μM $MnCl_2$, 15.1 mM $(NH_4)_2SO_4$ and 0.5 mM $MgCl_2$) supplemented with 48 mM lactate or 100 μL of β-myrcene (Merck; CAS number: 123-35-3; density: 0.794 g/mL; log K_{OW}: 4.17; water solubility at 25°C 5.6 mg/L or 41.1 μM) as sole carbon source. Since the water solubility of β-myrcene is very low, the amount used in the experiments assured a constant water phase saturation (130 mg/L, 48 μmol in solution).

The cells were grown in 250 mL Erlenmeyer flasks, at 30°C and with orbital shaking (200 rpm). For Erlenmeyer's containing β-myrcene, rubber stoppers were used to seal the flasks.

For growth monitoring the cells Optical Density was measured at 600 nm (OD_{600nm}) in the spectrophotometer genesis 20. Moreover, pH values were also measured every day in the pHmeter micro2002.

3.2.2. Adaptive Laboratorial Evolution Experiment

A pre-inoculum, prepared using one single colony from *Pseudomonas* sp. M1 *wt* was inoculated in 100 mL of MM, using lactate (0.1%) and β-myrcene (20 μL) as carbon source, and grown as referred previously. After overnight growth, the inoculum was refreshed to MM with β-myrcene as only carbon source (100 μL) or to MM with lactate as only carbon source (0.40% lactate) with an OD_{600nm} of 0.1 and grown overnight at 30°C with agitation. In the following sections of this document, these cultures will be named as "evolving cultures".

The evolving cultures were propagated every day, after approximately 24h of growth, in 50 mL MM supplemented with the respective carbon source and using the proper volume of M1 cells for each starting inoculum reach an initial OD_{600nm} between 0.08 and 1.20. The propagation was carried out for approximately 2 years, in order to reach 600 generations in all evolving cultures.

After each daily transfer, the previously grown cultures were placed in the refrigerator at 4°C and stored during one day as a backup. At each 50 generations, 3 cryovials of each evolving population were stored

at -80°C, with 300 μL of cells and 300 μL of glycerol 30%. Also, 3 Eppendorf's containing 1200 μL of each culture were centrifuged at 7000 rpm for 6 minutes and were stored at -20°C, for total DNA extraction.

To monitor for contaminations, all the evolving cultures were plated every week in LB and *Pseudomonas* Isolation Agar (PIA) plates. Since only *Pseudomonas* grow in PIA plates, a different number of colonies count between LB and PIA, signal possible contamination.

In order to monitor the growth and the duplication time of *Pseudomonas* sp. M1 cells, the CFUs/mL were determined every week. The count of CFUs was performed at 0h, 8h and 24h of growth, to calculate the number of generations reached by the evolving populations in 24h of growth.

3.2.3. Total DNA extraction of evolved cultures, sequencing and datasets treatment

The total DNA of the evolved cultures was extracted with a protocol adapted from Yu and Morrison[191]. A solution containing 10% SDS and 0.5 M EDTA was used to lyse the cells, the RNA was degraded using RNAse enzyme, protein precipitation was performed using 7.5 M ammonium acetate and the DNA precipitation and concentration was performed using isopropanol. After precipitation, the DNA samples were washed using ethanol 70% and rehydrated with Tris-HCl, 5 mM, pH 8.5. After extraction, the integrity and quantity of the DNA samples was assessed by DNA electrophoresis with 0.8% agarose gel prepared in TAE (2 M Tris base; 1 M Acetic acid; 50 nM EDTA). In this protocol, the RNA and protein removal steps were performed previously to the DNA elution, and the percentages of SDS and EDTA were superior to the original protocol.

The extracted DNA was processed according to Illumina instructions to generate Nextera XT paired-end libraries (2x150bp) and sequenced using high-throughput Illumina Nextseq sequencing platform at Instituto Gulbenkian de Ciência (IGC).

For the datasets treatment, the reads were first assessed using *FastQC* program, to check its quality and reads length. Since the quality at the end nucleotides of all reads was variable and the read lengths were also variable, the datasets had to be trimmed. To understand the implications of trimming, the reads were first assembled using the tool *Unicycler* version 0.5.0. Moreover, the reads were mapped against the *Pseudomonas* sp. M1 reference genome already available with *Breseq* version 0.38.1, to know which mutations were found. The reads were then trimmed using several programs - *FastQ-mcf* from *EA-utils* package version 1.1.2, *Trimmomatic* version 0.39 and *Sickle* version 1.33. After assembly with *Unicycler* and mutations calling with *Breseq*, the chosen trimming program was *Trimmomatic*. Thus, the obtained

datasets were firstly trimmed with *Trimmomatic* version 0.39. The program scanned all the reads with a 2-base wide sliding window, cutting when the average quality per base dropped below 20. Moreover, the reads below 150 bases long were dropped. For the scanning of genetic variants between the ancestor strain and the evolved cultures, *Breseq* version 0.38.1 was used. For this analysis, polymorphism-prediction was used since the samples were not expected to be clonal. This mode predicts polymorphic (mixed) mutations.

3.2.4. Isolation and conjugation of evolved M1 strains to β-myrcene

A screening was performed in the β-myrcene evolved populations, at 600 generations. For this, a vector – pSEVA637 – harboring a fusion between Green Fluorescent Protein (GFP) as reporter and the promoter P5 (responsible for the expression of protein MyrH involved in the first step of β-myrcene catabolism), was used. Firstly, 1 mL of each Myr culture at 600 generations was collected. After centrifugation, the cells were diluted in NaCl 0.9% and plated in LB, allowing the isolation of colonies. After growth, a total of 300 colonies (60 of each evolved culture) were collected randomly. The collected colonies were then grown in 3 ml of LB liquid in falcon (overnight growth, 30ºC, 200 rpm).

E.coli S17.1 containing the pSEVA637 vector harboring the promoter probe were grown overnight in LB liquid supplemented with gentamycin at 37ºC, with agitation.

For the conjugation, both *Pseudomonas* sp. M1 evolved isolates and *E.coli* were washed with 1 mL of NaCl 0.9% three times, resuspended and mixed together in the same Eppendorf. After centrifugation, the pellet was plated in a 200 μm filter placed in LB plates. This procedure allows for both strains to be physically together allowing conjugation. After 24h growth at 30ºC, the mixtures were resuspended in 1 mL NaCl 0.9% and plated into *Pseudomonas* Isolation Agar (PIA) plates supplemented with 30 μg/mL gentamycin, allowing the isolation of only *Pseudomonas* with the vector (conjugants). PIA plates were grown for 48h at 30ºC. The conjugation procedure was also performed in the *Pseudomonas* sp. M1 ancestor strain.

3.2.5. Monitoring the fluorescence of the isolated strains

In order to quantify and compare the fluorescence levels of the *Pseudomonas* sp. M1 evolved isolates to the ancestor strain, the conjugated colonies were grown in MM medium supplemented with β-myrcene in 96-well plates. The conjugated cells were firstly resuspended in MM supplemented with lactate for an OD_{600nm} of 0.1 and then plated, in duplicate, in 96-well microplates, in a total volume of 200 μL per well.

These microplates were supplemented with β-myrcene in some empty wells and lids and closed with parafilm. The same volume of the same conjugated cells was plated in parallel in a control microplate, without β-myrcene supplementation. These control plates were considered the blank. All the microplates were grown overnight at 30ºC and 100 rpm. After growth, the OD_{600nm} was measured in a microplate reader Biorad iMark and the fluorescence measured in a fluorimeter Fluorskan Ascent FL using an excitation wavelength of 485nm and an emission wavelength of 538nm.

The fluorescence values of each *Pseudomonas* sp. M1 colony measured were normalized with the OD_{600nm} resulting in relative fluorescence units (RFU). Due to basal fluorescence of *Pseudomonas* sp. M1, the relative fluorescence levels detected after induction with β-myrcene were represented as the ratio between the normalized fluorescence of cells grown with β-myrcene and the normalized fluorescence of cells grown with lactate (fluorescence levels induced by β-myrcene = RFUmyr/RFUlac). Also, the ratio between the fluorescence levels produced by the isolates and ancestor strain were calculated in order to compare changes is the performance of evolved M1 cells that resulted from the experimental evolution, giving some insights about which isolates could possibly be more efficient in the β-myrcene catabolism.

The fluorescence of isolates with the highest fluorescence levels during this assay, were confirmed in Erlenmeyer flasks. These isolates were grown in liquid MM medium supplemented with 48 mM of lactate and grown for approximately 2h at 30ºC, 200 rpm. After reaching an OD_{600nm} of 0.3 the cells were challenged with 100 µL of β-myrcene, inducing the reporter. Each 2h, the fluorescence was measured in a Qubit Fluorimeter. Moreover, the isolates and ancestor counterparts not containing the probe were also grown in parallel and its fluorescence measured in the same time points.

To confirm the presence of the vector in the isolates with higher fluorescence levels, the primers used were the following:

Forward: 5' TATAATTTAAATACTCCCACCCGCCCATTTCC3'

Reverse: 5' ACTGATTTAAATCTGGATTCTCACCAATAAAAAACGC3'

The PCR was run in the thermocycler using the following program: initial denaturation step at 96ºC for 3 minutes, 30 cycles of denaturation at 96ºC for 30 seconds, annealing at 60ºC for 30 seconds and extension at 72ºC for 45 seconds, followed by final extension step at 72ºC for 5 minutes.

3.2.6. Total DNA extraction of isolated strains, sequencing, and datasets treatment

The total DNA of isolates were extracted and sequenced as previously done for the evolved cultures - A solution containing 10% SDS and 0.5 M EDTA was used to lyse the cells, the RNA was degraded using

RNAse enzyme, protein precipitation was performed using 7.5 M ammonium acetate and the DNA precipitation and concentration was performed using isopropanol. After precipitation, the DNA samples were washed using ethanol 70% and rehydrated with Tris-HCl, 5 mM, pH 8.5. After extraction, the integrity and quantity of the DNA samples was assessed by DNA electrophoresis with 0.8% agarose gel prepared in TAE (2 M Tris base; 1 M Glacial acetic acid; 50 nM EDTA).

The extracted DNA was processed according to Illumina instructions to generate Nextera XT paired-end libraries (2x150bp) and sequenced using high-throughput Illumina Nextseq sequencing platform at Instituto Gulbenkian de Ciência (IGC).

The obtained datasets were firstly trimmed with *Trimmomatic* version 0.39. The program scanned all the reads with a 2-base wide sliding window, cutting when the average quality per base dropped below 20. Moreover, the reads below 150 bases long were dropped. For the scanning of genetic variants between the ancestor strain and the evolved isolates, *Breseq* version 0.38.1 was used. For this analysis, the default mode was used since the samples were expected to be clonal.

3.2.7. *fleQ* gene deletion in the ancestor strain

To verify the effect of mutations in the gene *fleQ*, thought to be key for the increased performance of evolved isolates, the ancestor strain was submitted to the targeted deletion of this gene. The procedure used for this step was the same as Aparicio and collaborators[71].

Firstly the *E.coli* CC118 containing the three necessary plasmids – pSEVA2316, pSEVA658-ssr and pSEVA421-Cas9tr were grown overnight in LB liquid media. In the morning of the next day, the three plasmids were extracted. The majority of the plasmids were used for the next steps of the protocol, but E.*coli* DH5α was also transformed with each of these plasmids for future storage. For plasmid extraction, the method used was the classic extraction. From the overnight cultures of *E.coli* cells, 1.5 mL of culture was collected and centrifuged. The pellet was then resuspended with 100 µL of Solution 1 (25 mM Tris-HCL, pH 8.0, 10 mM EDTA and RNAse A to a final concentration of 100 µg/mL). For the next step, the lysis, 200 µl of Solution 2 were added (200 mM NaOH, 1 %SDS) and the lysis was stopped using 150 µL of Solution 3 (3 M KOAc, pH 4.8). The mixture was centrifuged for 15 minutes and the supernatant was recovered. The DNA was precipitated using isopropanol 100% in a proportion of 1:1 and washed after precipitation with ethanol 70%. After drying at room temperature, the DNA was resuspended using 10 Mm Tris-HCl, pH 7.8. After extraction the plasmid pSEVA2316 was cloned with the spacers necessary for the Cas9 endonuclease recognize its target. The spacer sequences chosen were the following:

Oligo-S: 5'AAACGCCTTCACCGGCGCCATCACCAGCCGCGCCG3'

Oligo-AS: 5'AAAACGGCGCGGCTGGTGATGGCGCCGGTGAAGGC3'

For the cloning, Golden Gate protocol was chosen[192]. Firstly, the oligos were phosphorylated with T4 kinase enzyme (10x T4 DNA kinase buffer, 50 µM Oligo forward, 50 µM Oligo reverse, 1 µL T4 polynucleotide kinase in a total volume of 50 µL) for 1 hour at 37°C in a thermoblock. For the annealing, 1.0 M of NaCl were added to the phosphorylated mixture and incubated at 95°C for 3 minutes in a thermocycler. In the golden gate assembly, the steps of digestion with restriction enzymes and ligation to the plasmid happen simultaneously. For this, 10x T4 DNA ligase buffer, 100 nM annealed oligos, 20 nM pSEVA2316, 0.5 µL T4 DNA ligase and 0.5 µL of ECO31I (*BsaI*) restriction enzyme were mixed in a final volume of 10 µL. The reaction mixture was then incubated in the thermocycler following the protocol: 25 cycles of heating step at 37ºC for 3 minutes and cooling down step at 16ºC for 4 minutes, followed by a step of 80ºC for 15 minutes. Afterwards, the cloned plasmid was transformed into *E.coli* DH5α strain. The competent cells were first defrosted on ice. Afterwards 100 ng of the cloned plasmid were added to the mixture. The cells were incubated in 42°C in the thermoblock and immediately transferred to ice and incubated for 10 min. After the heat-shock, the cells were resuspended in 800 µL of liquid LB pre heated at 37°C and incubated for 1h at 37°C, 200 rpm. After recovery, 200 µL the cells were streaked into LB plates containing the resistance marker (50 µg/mL Kanamycin). After growth, several colonies were chosen, its pDNA extracted as previously described and sent to sanger sequencing. The plasmid with the right construct was used for electroporation into *Pseudomonas* sp. M1.

Pseudomonas sp. M1 ancestor cells were grown in MM supplemented with Lactate overnight, as previously described. In the morning, cells were refreshed for the same media and grown until an OD_{600nm} of 0.8, harvested, centrifuged at 5000 rpm for 10 minutes and washed 3 times with washing buffer (300 mM sucrose, 7 mM sodium phosphate, pH 4.7, 1 mM $MgCl_2$) for electroporation (according to the Bacterial Electro-transformation and Pulse Controller Instruction Manual version 8). To the 2-mm gap electroporation cuvette previously chilled, 100 ng of plasmid pSEVA421-Cas9tr was added to its walls followed by the transference of 100 µL of electrocompetent M1 cells. The electroporation pulse was set to 2.5 kV and after the pulse, 900 µL of LB were added. The cells recovered for 2 hours at 30°C with shaking, followed by plating in LB streptomycin. Electroporation was then followed for the plasmid pSEVA658-ssr selecting the transformants by plating in LB Streptomycin (Sm) and Gentamycin (Gm). The expression of the Ssr recombinase from the plasmid pSEVA658-ssr was induced with 3-methylbenzoate (3 MB) which activates the XylS-Pm system for the transcription. Thus, the selected cells containing both plasmids were grown overnight in MM media supplemented with Lactate, 15 µg/ml Gm and 100 µg/ml Sm and grew as previously described. In the morning, 1 mM of 3 MB was added to the

cells and the culture was incubated for 3h more. After, the cells were collected and washed again with electroporation buffer to be electrocompetent.

For the homologous recombination to occur, the presence of the ssDNA containing the 40 nucleotides upstream and downstream homologous to the target-gene to be deleted is necessary. The ssDNA oligos sequence used was:

5'CTTCGGAAGACAAGAGCAAGATGTGGCGTGAAACCAAGATCGGCATGAGCCGCCGCGAGGAGGAAGTGGC GGAGGATTGA3'.

The M1 electrocompetent cells induced with 3 MB were once again electroporated to incorporate the ssDNA oligos and the construction of plasmid pSEVA2316 + Spacer. First, a stock solution of 100 μM ssDNA oligos was prepared. Afterwards, a mixture of 1 μL ssDNA oligo and 100 ng of the construction pSEVA2316 + Spacer were added to the walls of the electroporation cuvette, followed by 100 μL of electrocompetent cells and electroporation was performed as previously described. The transformants were then plated in LB containing kanamycin (Km), gentamycin, and streptomycin.

The presence of mutants was verified through PCR using primers complementary to the regions flanking the target-gene. Thus, amplification indicates that the gene was not deleted. The primers sequences were the following:

Forward: 5'CTTCGGAAGACAAGAGC3'

Reverse: 5'GTTGAGCAGTACCAGTTG3'

Twelve colonies were randomly selected for PCR colony. Firstly, the colonies were stroked in LB plate with a toothpick. This same toothpick was dipped in a microtube with 20 μL of distilled water. The 12 microtubes were then heated at 65°C for 20 minutes to break the cells and make their DNA available. This mixture was used as the DNA template. The PCR was run in the thermocycler using the following program: initial denaturation step at 96ºC for 3 minutes, 30 cycles of denaturation at 96ºC for 30 seconds, annealing at 63ºC for 30 seconds and extension at 72ºC for 1 minute and 30 seconds, followed by final extension step at 72ºC for 5 minutes.

For the curation of the plasmids, the cells were grown for 4 days, in MM supplemented with β-myrcene without any selection marker, being refreshed every 48h growth. The cells were struck into 4 LB plates (without antibiotic, with Km, Sm or Gm) and a PCR using the primers PS1 and PS2 designed by Aparicio and collaborators were used to prove the non-existence of such plasmids in the cells. These primers are complementary to the pSEVA plasmids[71]. The PCR program used was the same as above.

3.2.8. Motility assay

After the deletion of the target gene, the effect on cells motility was verified. LB plates with only 0.3% agar were prepared, as well as MM plates also with 0.3% agar[193]. M1 mutant constructed previously, M1 evolved strains and M1 *wt* were grown in LB medium and MM medium supplemented with β-myrcene, until an OD_{600nm} around 0.3, corresponding to the exponential phase of growth. Afterwards, 3 µL of cells were plated in the center of each plate. After 48h growth at 30°C, the plates were photographed in a Chemidoc and the rings were measured through pixel accumulation.

3.3. Results and Discussion

Pseudomonas sp. M1 cells underwent 600 generations of laboratory-driven evolution and adaptation using β-myrcene or lactate as sole carbon and energy sources in Mineral Media. This media will allow adaptation towards the chosen carbon sources since there are no other carbon sources available. For this experiment one single colony was used to supply all the evolving cultures, thus all the independently evolving cultures have the same genomic background. For each carbon source, 5 independent cultures were established – 5 cultures adapting to β-myrcene and 5 cultures adapting to lactate (Figure 15).



Figure 15 - Setup for Adaptation Laboratorial Experiment. One single colony was used as starting material for all the independent cultures. All cultures were grown in Minimal Media; 5 cultures were supplemented with β-myrcene and 5 were supplemented with lactate. Cells were propagated every 24h and the experiment was repeated until 600 generations.

The volume of β-myrcene used in this experiment was 100 µL. This quantity avoids starvation effects that may impact the experiment. The partition of this compound between gas phase and liquid phase determines the bioavailability of this compound in aqueous environments, thus affecting the biotransformation of this compound. These calculations were already performed in previous studies of the investigation team[194]. The use of several cultures growing and adapting independently, was important to find which mutations may be related with the adaptation to the respective carbon sources. All the

independent cultures were grown at 30°C with agitation for 24h and refreshed by serial transfer after that time. With 24h growth, M1 cells reached the stationary phase growth. Thus, the cells will be adapted not only to the desired carbon source, but also to growth in stationary phase, increasing the difficulty of data analysis since the mutations will not be related only to the adaptation to carbon sources. The reason behind this setup is connected to the increased mutation rate during stationary phase growth. During this phase, mutations in amino acid catabolic processes and stress response sigma factor, *rpoS* are important to survive and get a selective advantage in such nutrient starved conditions[195]. However, the stress responses are mainly controlled by RpoS, able to regulate the mobility of transposable elements and the DNA repair systems[196,197]. Moreover, during this phase, SOS response is activated which increases the low-fidelity DNA polymerase IV expression[198]. Thus, the emergence of a higher number of mutations may be possible, being also possible the mutation of genes such as *mutL* and *mutS* belonging the mismatch DNA-repair system. Increasing the mutation rate of a population may accelerate the evolutionary adaptation, since the accumulation of mutations is higher, including beneficial mutations. However, when the mutation rate is too high, the balance between beneficial, deleterious and neutral mutations is broke, thus leading to populations with decreased fitness, sometimes even less fitness than the ancestor strain, in a long ALE experiment (e.g. 5000 generations)[199,200]. This was reported for *E. coli* where a high mutation rate was considered as 100 to 1000-fold higher (hypermutants) than the ones usually observed for a wild-type *E. coli* strain[201,202].

Moreover, the use of 2 different carbon sources in this work allowed to distinguish between mutations related with growth in stationary phase and random mutations introduced by high mutation rates from beneficial mutations specific for the adaptation to each carbon source.

In the setup used in this thesis, the number of generations reached during the time of the experiment were monitored through CFUs/mL count within 24h (Figure 16). The number of generations for the M1 cells adapting to β-myrcene, had a slight increase in the generations reached in 24h, comparing the beginning of the experiment (average of 3 generations reached) and the 600 generations (average of 4 generations reached). In the Lactate adapting cells, the number of generations reached in 24h had a slight decrease, comparing the beginning (average of 5 generations reached) and the end of the experiment (average of 4 generations reached). However, the Lactate cultures reached the 600 generations faster than β-myrcene cultures.

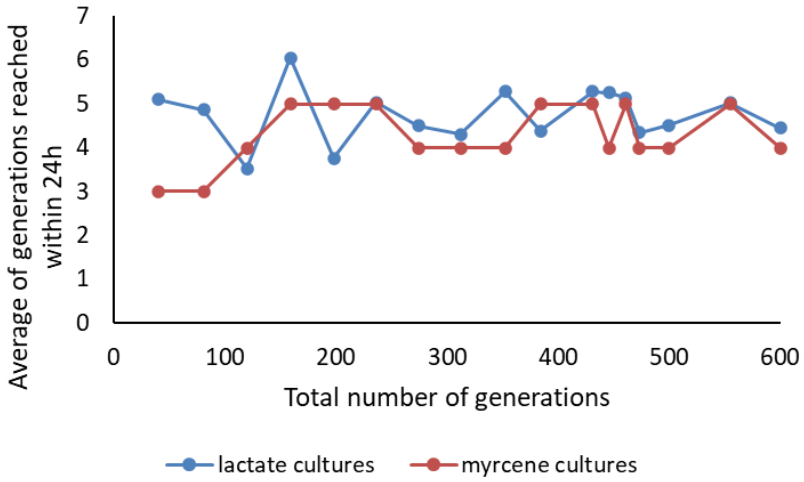


Figure 16 - Number of generations reached within 24h for lactate adapting cultures (blue) and for β-myrcene adapting cultures (red). The graphic represents the average number of generations for β-myrcene or lactate evolving cultures within 24h through all the experiment (until a total of 600 generations).

The cultures were monitored daily, through the measurement of pH and OD_{600nm}. Briefly, the cultures growing in β-myrcene reached an average OD_{600nm} of 1.8 and an average pH of 5.5 and the cultures growing in Lactate reached an OD_{600nm} of 2.5 and an average pH of 8.8. Moreover, the cultures were monitored for contamination though plating every week in LB and PIA plates. Contamination was detected at 550 generations for culture Lactate 5. The data obtained for this culture at 600 generations were not used for this thesis. All the data presented for Lactate cultures from now on will only take into consideration the 4 remaining cultures.

3.3.1. Comparative genomics

After 600 generations of ALE experiment, 1 mL of each culture was retrieved at 200, 400 and 600 generations, the total DNA was extracted, followed by sequencing using Illumina Next-seq. In each culture, the existence of several populations with different mutations is expected. The time of experiment was set taking into consideration the literature existent which states that between 100 and 500 generations adaptation is usually fast, slowing down during the course of the evolution experiment[170,203]. DNA sequencing at the culture level, allowed to know the overall mutations present in each culture and their frequency, using the program *Breseq* in the polymorphism prediction mode that attempts to identify the frequency of each mutation (SNV or Indel) in the sample provided. The program *Breseq* maps the reads from the evolved populations against a reference genome, which, in this case *Pseudomonas* sp. M1, already exists with high quality (Genbank accession: GCF_022760135.1). In the evolution

experiments, several populations will be present in one single culture, harboring different mutations. Thus, the need to determine the frequency of each detected mutation becomes imperative. *Breseq* was already used to identify mutations after laboratory evolution experiments by several authors[204–206]. As reported previously by Daniel E. Deatherage and collaborators, the low quality of reads 3' ends may introduce mismapping which is a main source of errors, resulting in false-positive mutations predicted by *Breseq*. Thus, several trimming programs were tested before the use of *Breseq*. *Trimmomatic* allows the trimming of Illumina adaptors, as well as the trimming of reads by quality and by length. Several trimming programs were tested. Briefly, the reads were assembled using the program *Unicycler* before and after trimming. Reads trimmed with *Trimmomatic* formed less contigs, indicating the higher quality of the trimmed reads. *Trimmomatic* scanned the average quality of each 2 base pairs of all reads. Any read with quality below 20 would be discarded. Moreover, all reads below 40 b.p. were dropped. Reads shorter than 40 b.p. may align in the wrong place of the genome, causing errors during mapping. The number of reads before and after trimming with *Trimmomatic* are depicted in Table 7.

Table 7 - Total number of reads before and after trimming. The program used was *Trimmomatic* with sliding window of 20 and read length of 40 b.p. The total number of reads after trimming decreased between 50 and 60%. The cultures are numbered from 1 to 10. The numbers 1 to 5 are the Myr cultures and 6 to 10 are the Lac cultures by order.

	Total reads before trimming	Total reads after trimming		Total reads before trimming	Total reads after trimming		Total reads before trimming	Total reads after trimming
200G-1	8549816	5949473	**400G-1**	9054495	5240013	**600G-1**	10013247	5359761
200G-2	8139118	5206120	**400G-2**	9429968	5312909	**600G-2**	8566885	5984021
200G-3	7772784	5389055	**400G-3**	7800493	4840155	**600G-3**	8960032	5188595
200G-4	6845837	5035484	**400G-4**	8439986	5535955	**600G-4**	8223152	4960526
200G-5	7270626	5392536	**400G-5**	7457347	5274073	**600G-5**	7759753	5724354
200G-6	7864860	5150780	**400G-6**	7780957	4840519	**600G-6**	6665147	5568950
200G-7	8953394	5051440	**400G-7**	7604141	9154170	**600G-7**	8475496	5380886
200G-8	8176319	5401959	**400G-8**	6780202	5564589	**600G-8**	7670600	4916874
200G-9	10113070	5144191	**400G-9**	7313266	5018933	**600G-9**	10431537	6498845
200G-10	11004728	5328508	**400G-10**	7222615	5182792	**600G-10**	7272575	5150796

After trimming with *Trimmomatic* the number of reads decreased, decreasing the coverage around 50% to 60%. However, the coverage after trimming is still between 100x and 158x (Table 8). The coverage refers to the extent of sequencing for each base pair of the genome.

Table 8 - Coverage for each culture before and after trimming. Coverage was calculated using the formula C=LN / G; where C - coverage, L - read length; N - number of reads and G - genome size. The cultures are numbered from 1 to 10. The numbers 1 to 5 are the Myr cultures and 6 to 10 are the Lac cultures by order.

	Coverage before trimming	Coverage after trimming		Coverage before trimming	Coverage after trimming		Coverage before trimming	Coverage after trimming
200G-1	177	124	**400G-1**	188	109	**600G-1**	208	111
200G-2	169	108	**400G-2**	196	110	**600G-2**	178	124
200G-3	161	112	**400G-3**	162	100	**600G-3**	186	108
200G-4	142	105	**400G-4**	175	115	**600G-4**	171	103
200G-5	151	112	**400G-5**	155	109	**600G-5**	161	119
200G-6	163	107	**400G-6**	162	100	**600G-6**	138	116
200G-7	186	105	**400G-7**	190	158	**600G-7**	176	112
200G-8	170	112	**400G-8**	141	116	**600G-8**	159	102
200G-9	210	107	**400G-9**	152	104	**600G-9**	217	135
200G-10	228	111	**400G-10**	150	108	**600G-10**	151	107

After trimming, the number of reads used for each sample was normalized to 4840519, which was the lowest number of reads obtained for all the datasets and ran in *Breseq*. In this program, firstly, the reads are mapped to the reference genome using *Bowtie 2*, reporting partial alignments and matches to repeat sequences. Afterwards, only the two ends of the identified reads are mapped to the reference genome. If the ends of the reads align perfectly in that location and not in any other location of the reference genome, a variant junction (mutation) is evidenced. *Breseq* also has two modes to identify mutations, the "consensus" mode, and the "polymorphism" mode. During the "consensus" mode, the program assumes that new genetic variants have 100% frequency in each DNA sample sequenced since it is clonal. Thus, this mode can only be used to identify mutations from a single colony, which will be clonal. During the "polymorphism" mode, the program searches for a determined number of reads that reach or exceeds the minimum threshold. The threshold used in this work was the default, which is 3 reads. Moreover, the end of these reads must perfectly align with the reference genome in at least 6 additional base pairs than the "consensus" mode. Afterwards, the program calculates the mutation frequency taking into consideration the number of reads signaling each mutation and the total number of reads that map against the reference genome in that location. All this processing is performed resorting to an algorithm[207]. As output, this tool provides a table containing all the mutations within the reads available upon the sequencing of 1 mL of culture and the percentage of each mutation in the universe of reads

available. In this way, one may have hints about the stage of evolutionary adaptation and the establishment of populations. If a mutation has 100% frequency, it means that this mutation is present in all the reads analyzed, thus, present in all the individuals of that culture.

The data retrieved from the sequencing for all the cultures at 200, 400 and 600 generations allowed a general idea about the mutations accumulated during the evolution experiment. Starting by the overall number of mutations for all cultures, a higher number of mutations was detected in the lactate cultures in comparison to β-myrcene cultures – lactate cultures showed 2.9% more mutations at 200 generations, 4.7% at 400 generations and 5.3% at 600 generations – and the number of mutations increased over generations either in lactate or β-myrcene cultures (Table 9).

Table 9 - Overall mutations number for Lac and Myr cultures at 200, 400 and 600 generations.

Generations Number	Lactate cultures	β-Myrcene cultures
200	331	116
400	743	159
600	1194	224

Over the evolution process, mutations may arise in coding regions of the genome or in intergenic regions including promoters and terminator regions. The percentage of intergenic mutations vs coding mutations is different between cultures adapted to Lactate or to β-myrcene, thus different adaptation strategies were used by M1 cells to adapt to the different carbon sources. The percentage of intergenic mutations is higher than coding mutations in Myr cultures at 200 and 400 generations, being this trend inversed only at 600 generations, whereas the percentage of intergenic mutations are lower than coding mutations in Lac cultures at all generations (Figure 17).

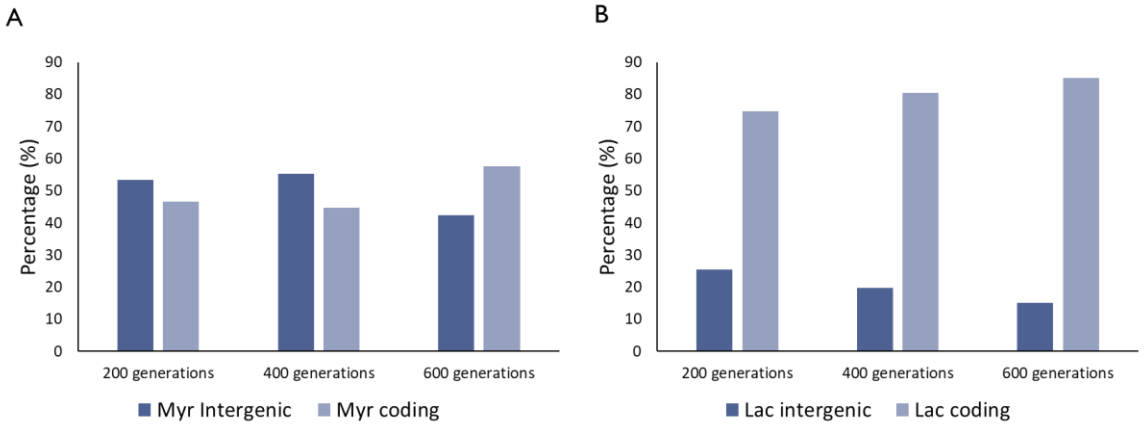


Figure 17 - Percentage of detected intergenic (dark blue) vs coding mutations (light blue) in Myr cultures (A) and Lac cultures (B).

Moreover, the number of non-synonym mutations for Lactate cultures is higher than synonym mutations at all generations, whereas in β-myrcene cultures this only happens at 400 and 600 generations (Table 10). The importance of these data relies in its impact on the proteins coded by the genes affected by such mutations. Synonymous mutations will code for the same amino acid, being the availability of tRNAs the step that may affect the translation of such protein, whilst non-synonymous mutations will effectively code for a different amino acid, which may impact the protein function.

Table 10 - Number of synonym vs non-synonym mutations in Lac and Myr cultures over generations.

	Lactate cultures		**β-Myrcene cultures**	
Generations Number	**Synonymous**	**Non-synonymous**	**Synonymous**	**Non-synonymous**
200	54	193	37	17
400	146	451	15	56
600	279	736	36	93

As referred previously, several lineages with different mutations may arise in one culture. As the program *Breseq* allows the determination of frequency of each mutation detected in the universe of reads obtained, the overall modifications and the fixation degree of a determined mutation can be calculated. When analyzing the frequencies of all mutations in the different cultures (Table 11), a prevalent difference is observed between β-myrcene cultures (M1, M2, M3, M4 and M5) and Lactate cultures (L1, L2, L3, L4 and L5): the β-myrcene cultures, except culture M5, do not have any mutations present in all reads sequenced (100% frequency) in the population over the generations. The culture M5 has some mutations with 100% frequency beginning at 400 generations. However, all Lactate cultures have mutations present in all reads sequenced (100% frequency), even at 200 generations. Moreover, overall β-myrcene cultures have a high percentage of mutations present in less than 10% of the reads sequenced (frequency above 10%), whilst in Lactate cultures this percentage is lower. All these data indicates that the evolution and adaptation process of *Pseudomonas* sp. M1 towards Lactate or β-myrcene carbon sources were different. Analyzing all the mutations detected for each culture, the increased mutations number in lactate cultures relates to the arising of mutations in *mutL* and *mutS* genes, responsible for gene repair. Mutations in such genes allow a higher number of mutations by chance all over the genome and are responsible for hypermutant populations. These mutations appeared as soon as 200 generations in Lac cultures, however only 2 Myr cultures (M2 and M5) showed mutations in *mutL* gene at 600 generations. Mutations in genes responsible for gene repair, increase genome instability[208], being a disadvantage when developing a biocatalyst[209,210]. However, mutations in the DNA repair system may represent an evolutive

advantage, since a higher frequency of mutations was achieved upon the acquisition of mutations in this system – Lactate cultures and M5 culture had the detection of mutations in 100% of the reads sequenced and the same did not happen for the β-myrcene cultures without mutations in the DNA repair system.

Table 11 - Percentage of mutations detected with 100% frequency, above 10% frequency and below 10% frequency, for all the cultures at 200, 400 and 600 generations.

	Mutations Frequency 100%			Mutations Frequency Above 10%			Mutations Frequency Below 10%		
	200 Generations	400 Generations	600 Generations	200 Generations	400 Generations	600 Generations	200 Generations	400 Generations	600 Generations
M1	0.0	0.0	0.0	20.0	20.8	30.2	80.0	79.2	68.3
M2	0.0	0.0	0.0	33.3	15.8	52.7	100.0	84.2	47.3
M3	0.0	0.0	0.0	14.8	22.0	21.7	133.3	76.3	78.3
M4	0.0	0.0	0.0	9.3	18.5	13.9	122.2	77.8	86.1
M5	0.0	5.9	3.7	29.2	43.5	45.0	104.2	56.5	53.2
L1	13.8	0.5	3.0	42.6	47.2	57.0	56.4	51.8	42.6
L2	6.2	6.2	3.9	64.9	49.1	71.1	35.1	50.4	28.7
L3	13.6	6.3	6.1	51.5	63.9	47.5	48.5	35.6	52.2
L4	0.0	2.3	0.7	41.2	43.4	77.2	58.8	56.6	22.1
L5	7.0	5.7	-	61.6	78.6	-	38.4	21.4	-

As referred previously, populations with increased mutation rate may accelerate the evolutionary adaptation, since the accumulation of mutations is higher, including beneficial mutations[202]. A table with all the mutations detected at the different generations for each culture can be found in supplementary material, Table S1 for Lactate cultures and Table S2 for β-myrcene cultures. Since this experiment was performed to increase *Pseudomonas* sp. M1 towards the catabolism of β-myrcene, mutations in the 28-kb *locus* necessary to catabolize β-myrcene were expected. However, in any of the β-myrcene cultures were detected mutations in this cluster. Regarding mutations in this genomic island (GI) for cultures adapting to Lactate, 3 cultures – L1, L3 and L4 – had mutations in the gene MRY70_06120, coding for a protein belonging to the TU3 of the GI only at 600 generations. This was the only mutation inside the GI of all the experiment, despite the hypermutant character of Lac cultures.

Focusing on the populations evolved until 600 generations, for all the coding regions (genes) with mutation (either synonym, non-synonym or Indels), the function of the proteins encoded by those genes was searched in the Clusters of Orthologous Genes (COGs) database. With these data, the major functions affected for each culture may be evaluated, providing an easy comparison between Lac and Myr cultures (Figure 18). In Lactate cultures, the number COGs identified is more diversified than in β-

myrcene cultures. This is expectable since Lac cultures had around 25% more mutations than Myr cultures.

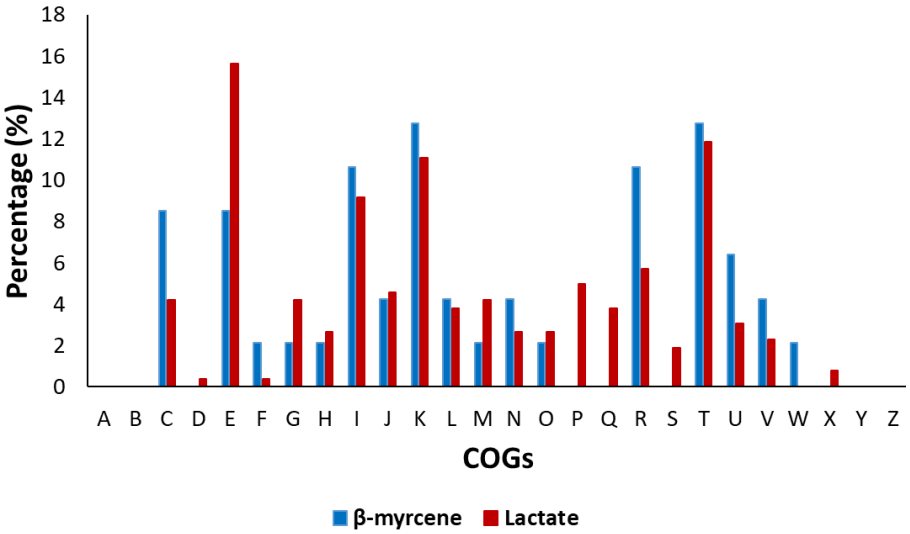


Figure 18 - COGs function for genes mutated in Myr cultures (Blue) and Lac cultures (red). Each letter stands for A - RNA processing and modification; B - Chromatin structure and dynamics; C - Energy production and conversion; D - Cell cycle control, cell division, chromosome partitioning; E - Amino acid transport and metabolism; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme transport and metabolism; I - Lipid transport and metabolism; J - translation, ribosomal structure and biogenesis; K – Transcription; L - Replication, recombination and repair; M - Cell wall/membrane/envelope biogenesis; N - Cell motility; O - Posttranslational modifications, protein turnover, chaperones; P - Inorganic ion transport and metabolism; Q - Secondary metabolites biosynthesis, transport and catabolism; R - General function prediction only; S - Function unknown; T - Signal transduction mechanisms; U - Intracellular trafficking, secretion, and vesicular transport; V - Defence mechanisms; W - Extracellular structures; X - Mobilome: prophages, transposons; Y - Nuclear structure; Z – Cytoskeleton.

The highest percentage of COGs for Lac cultures are in the category E - Amino acid transport and metabolism; I - Lipid transport and metabolism; K – Transcription and T - Signal transduction mechanisms. The genes into COG category E may be related with adaptation to the stationary growth phase since mutations in this category will be important to acquire a selective advantage in starvation. The categories K (transcription) and I (Lipid transport and metabolism) may be related with adaptation to stationary phase, since transcription and lipid transport and metabolism are necessary for the cells to deal with starvation, to regulate for instance the shifts in membrane composition and the redirection of lipid metabolism to the production of signaling molecules that may be used for instance in quorum sensing[211,212]. For Myr cultures, the most prevalent COGs are T - Signal transduction mechanisms; K – Transcription; I - Lipid transport and metabolism and R - General function prediction only. Category R stands for little knowledge acquired for the proteins coded by those genes. For the categories I, K and T

as in Lactate cultures, may be related with adaptation to stationary phase. However, since β-myrcene is a hydrophobic compound, mutated genes in those categories may also be related to the specific adaptation to the compound.

To understand which functions were affected in each culture at 600 generations, despite the overall results, the COGs of the proteins expressed by the mutated genes in each culture were also determined. In fact, the purpose of independently evolving cultures is to compare the major functions affected by each culture in the same condition and gather some hints about the evolution process that occurred. Given the knowledge gathered about the evolution/adaptation processes in the literature, it is expected that during the adaptation to the same condition, independently evolved cultures will gather mutations in the same gene, or affecting similar processes[211,213,214]. Thus, Major functions affected in all cultures of the same condition, may provide hints about the most important functions for adaptation (Figure 19). Lactate cultures had several COGs identified in all 4 cultures, namely the COGs C, E, G, K, L, P, R and T. However, in β-myrcene cultures, the only category in which all 5 cultures have genes mutated is in the category T (signal transduction). Thus, for β-myrcene cultures, mutations in genes coding for proteins in signal transduction category, may be related to the specific adaptation to β-myrcene.

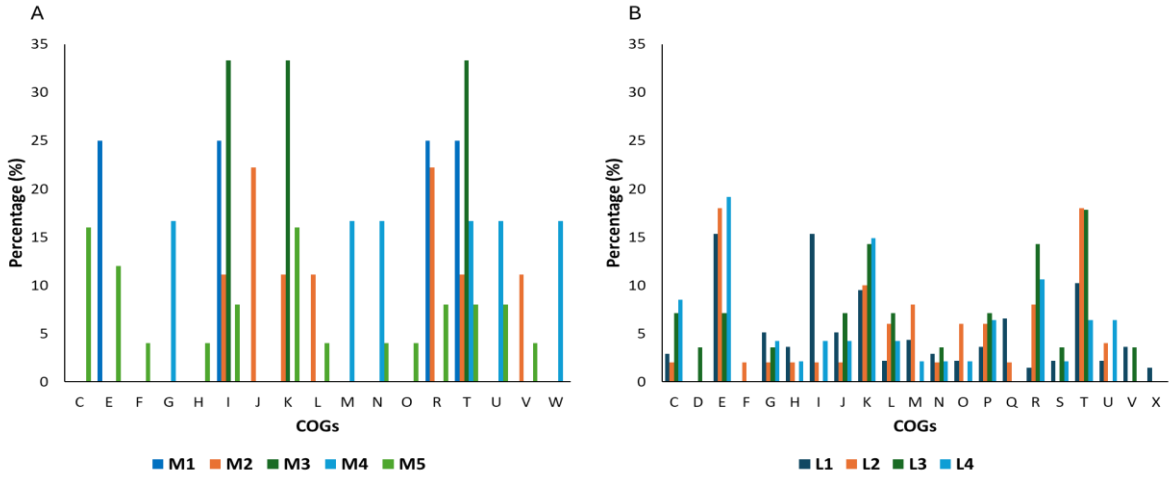


Figure 19 - COGs function for genes mutated for each culture in Lac conditions (A) and Myr conditions (B). Each letter stands for A - RNA processing and modification; B - Chromatin structure and dynamics; C - Energy production and conversion; D - Cell cycle control, cell division, chromosome partitioning; E - Amino acid transport and metabolism; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme transport and metabolism; I - Lipid transport and metabolism; J - translation, ribosomal structure and biogenesis; K – Transcription; L - Replication, recombination and repair; M - Cell wall/membrane/envelope biogenesis; N - Cell motility; O - Posttranslational modifications, protein turnover, chaperones; P - Inorganic ion transport and metabolism; Q - Secondary metabolites biosynthesis, transport and catabolism; R - General function prediction only; S - Function unknown; T - Signal transduction mechanisms; U - Intracellular trafficking, secretion, and vesicular transport; V - Defence mechanisms; W - Extracellular structures; X - Mobilome: prophages, transposons; Y - Nuclear structure; Z – Cytoskeleton.

In order to know the mutations necessary for the adaptation to β-myrcene or lactate, the unique mutations for each carbon source were retrieved. Thus, the mutations between all Lactate and β-myrcene cultures were compared, and the unique genes mutated for each condition were identified. For lactate cultures, the genes mutated with 100% frequency were further analyzed. Moreover, as explained previously, one may seek genes mutated in several independent cultures. Thus, genes mutated with 100% frequency and reported in all 4 Lac cultures were selected. These genes and respective mutations are represented in Table 12. The first apparent feature is the non-synonymous nature of all mutations detected. Only 3 mutated genes are included in the COGs most prevalent for Lac cultures: E - Amino acid transport and metabolism and I - Lipid transport and metabolism.

Table 12 - Genes mutated in all 4 Lac cultures and with 100% frequency in at least one culture. The table represents the mutated amino acid after translation to protein; the frequency of the mutation in all Lac cultures, the COG and the closest identity match through *Blastp*. In the mutation column, the first letter is the amino acid present in the ancestor strain, the numbers are the position where the mutation occurred, and the second letter is the amino acid present in the mutated protein.

Mutation	L1	L2	L3	L4	Gene	Protein	COGs	Blastp
A64T (GCC→ACC)	22.90%	100%	100%	39.00%	MRY70_00550	**WP_009614025.1**	C	90.69% identity Glucose 1-dehydrogenase *P. citronellolis*
P642S (CCG→TCG)	23.20%	100%	100%	41.80%	MRY70_03680	**WP_009613007.1**	E	89.31% identity two-component system sensor histidine kinase CbrA *P. aeruginosa*
L135P (CTG→CCG)	16.00%	100%	93.20%	40.70%	MRY70_05055	**WP_009616023.1**	H R	100% identity alpha/beta hydrolase Pseudomonas
G297E (GGG→GAG)	31.70%	100%	100%	33.90%	MRY70_08825	**WP_009623594.1**	H	99.68% identity D-2-hydroxyacid dehydrogenase *P. citronellolis*
A100T (GCC→ACC)	24.20%	100%	100%	32.90%	MRY70_09890	**WP_009616984.1**	K	91.24% identity HTH-type transcriptional regulator RutR *Stenotrophomonas maltophilia*
M202T (ATG→ACG)	20.20%	100%	100%	43.50%	MRY70_11235	**WP_009623357.1**	Q	100% identity 2-oxopent-4-enoate hydratase *Pseudomonas*
V556A (GTG→GCG)	25.20%	100%	100%	38.60%	MRY70_14050	**WP_009622345.1**	J	95.20% identity Putative protease ydcP *P. knackmussii* B13
N579S (AAC→AGC)	24.60%	100%	100%	38.50%	MRY70_20930	**WP_009621351.1**	I	100% identity fatty acid oxidation complex subunit alpha FadB *Pseudomonas*
G108D (GGC→GAC)	28.60%	100%	100%	39.10%	MRY70_21170	**WP_009621272.1**	N	98.47% identity fimbrial protein *Pseudomonas*
T49I (ACC→ATC)	25.20%	100%	100%	42.60%	MRY70_21475	**WP_009621090.1**	N	100% identity flagellar biosynthesis protein FlhA *Pseudomonas*
R251C (CGC→TGC)	24.00%	100%	100%	40.30%	MRY70_24925	**WP_009617826.1**	M	100% identity UDP-N-acetylglucosamine 1-carboxyvinyltransferase *Pseudomonas*
D197G (GAC→GGC)	26.80%	100%	100%	42.10%	MRY70_26145	**WP_009617422.1**	C E	82.12% identity homoisocitrate dehydrogenase *P. nitroreducens*

The protein WP_009613007.1 was identified as the two-component system sensor histidine-kinase CbrA. CbrA seems to be related with L-Histidine transport in *P. putida* and when coupled with the CbrX product is able to respond to carbon availability. The protein WP_009621351.1 was identified as the subunit alpha FadB from the fatty acid oxidation complex, which is involved in the aerobic and anaerobic degradation of long-chain fatty acids via beta-oxidation. The protein WP_009617422.1 was identified as

a homoisocitrate dehydrogenase. These enzymes are usually associated with the lysine biosynthesis pathway. All these mutated genes may be related to adaptation of growth in stationary phase, as stated previously when analyzed the E and I COGs for Lac cultures. The proteins WP_009623594.1, identified as D-2-hydroxy-acid dehydrogenase, is directly involved with lactate metabolism, catalyzing the conversion of lactate into pyruvate. Moreover, the protein WP_009614025.1, identified as Glucose-1-dehydrogenase, may be related with Lactate adaptation in order to decrease the glycolysis metabolism and favor lactate metabolism. The proteins WP_009621272.1 and WP_009621090.1 were identified as fimbrial protein and flagellar biosynthesis protein, being related with motility. These proteins may also be related with the adaptation to lactate metabolism since motility allows the cells to move into more favorable environments with high concentration of carbon source. On the other hand, the motility apparatus may impose an energetic and catabolic burden on the cell. Thus, its inactivation may be advantageous for adaptation. The protein WP_009617826.1 was identified as UDP-N-acetylglucosamine 1-carboxyvinyltransferase peptidoglycan biosynthesis in bacteria which is vital in several stages of bacterial cell cycle. Thus, genes directly involved with lactate metabolism were mutated in all Lactate cultures. Besides this analysis by frequency, mutated genes specific for Lac cultures and detected in all 4 cultures, were also retrieved. In this case, 2 genes were identified and for the gene MRY70_14975, different mutations occurred for different independent cultures (Table 13).

Table 13 - Genes mutated in all 4 Lac cultures without taking into consideration mutation frequency. The table represents the mutated amino acid after translation to protein; the frequency of the mutation in all Lac cultures, the COG and the closest identity match through *Blastp*. In the mutation column, the first letter is the amino acid present in the ancestor strain, the numbers are the position where the mutation occurred, and the second letter is the amino acid present in the mutated protein.

Mutation	L1	L2	L3	L4	Gene	Protein	COG	Blastp
H52R (CAC→CGC)	42.00%	42.20%	66.40%	37.70%	MRY70_06715	**WP_009616513.1**	K	100% identity LysR family transcriptional regulator *Pseudomonas*
R959C (CGT→TGT)	27.70%		73.20%	37.60%	MRY70_14975	**WP_009622050.1**	Q	100% identity non-ribosomal peptide synthetase *Pseudomonas*
P3532L (CCG→CTG)	52.90%							
L5767L (CTG→CTA)		35.10%						

Through *Blastp* the protein WP_009622050.1 was identified as a non-ribosomal peptide synthetase. Non-ribosomal peptides have several functions in the cell, being secondary metabolites synthesized in the cell without the need for ribosomal machinery or mRNA. These secondary metabolites have diverse

functions, being pigments, antibiotics, signaling molecules for quorum-sensing and biofilm formation, toxins, among others, some of the functions of these metabolites. Interestingly, the mutation in this gene for culture L2 is a synonymous mutation, contrary to all others. Thus, the function of the protein coded will not be altered in culture L2. For the protein WP_009616513.1 not much information was found. In M1 genome, the gene coding for this protein is flanked by genes with unknown function.

Focusing in β-myrcene cultures, a different approach was chosen due to the lower number of mutations and the existence of only 1 culture (M5) with mutations for 100% frequency. First, the unique mutations for Myr cultures were identified. Afterwards, from the unique mutations available, the ones with 100% frequency were selected (Table 14).

Table 14 - Genes mutated only in Myr cultures and with 100% frequency. The table represents the mutated amino acid after translation to protein; the frequency of the mutation in all Lac cultures, the COG and the closest identity match through *Blastp*. In the mutation column, the first letter is the amino acid present in the ancestor strain, the numbers are the position where the mutation occurred, and the second letter is the amino acid present in the mutated protein. + represents an insertion and Δ a deletion.

Mutation	M5	Gene	Protein	COGs	Blastp
Δ 1b.p.	100%	MRY70_07775	**WP_009616863.1**	C	99.47% identity cytochrome B *P. citronellolis*
+G	100%	MRY70_18970	**WP_009619258.1**	-	100% identity hypothetical protein *Pseudomonas*

The mutations with 100% frequency only appeared in M5 culture. Moreover, the 2 genes detected with 100% frequency of mutation for culture M5, were not mutated in any other culture, even considering lower frequencies. The protein WP_009616863.1 was identified as cytochrome B. This protein is a component of the respiratory chain that generates an electrochemical potential (through electrons transference) coupled to the production of ATP. Cytochrome B is involved in the last step of this chain, the reduction of oxygen to water. Mutations in the gene coding for CytB may increase or decrease the ability of cells to produce ATP. For the protein WP_009619258.1 any useful information was provided through *Blastp* or COGS. However, in M1 genome, this protein is flanked by the gene *ptsS* - involved in bacterial phosphate uptake and extracellular assembly of adhesion fibers - and a gene annotated as M14-type cytosolic carboxypeptidase. Cytosolic carboxypeptidases (CCPs) are usually intracellular peptidases involved, among other roles, in the post-translational modifications of tubulin. Thus WP_009619258.1 protein may be related with adhesion interactions. In fact, in wastewater treatments was already observed that with increased cells hydrophobicity, cell-to-cell adhesion and aggregation increases, leading to the

degradation of wastes such as phenol, pyridine and its derivatives, nitrogen, etc. via several metabolic pathways. Although the interaction and uptake of β-myrcene by M1 cells is still not known, mutation in this gene may be related with adaptation to β-myrcene metabolism, since β-myrcene is a hydrophobic compound.

The analysis of genes mutated only in Myr cultures and detected in at least 2 cultures was also employed. In this analysis, 7 genes were identified following these requirements and only 1 gene was mutated in all 5 cultures. Finding the cluster of orthologous genes for each mutated gene, 4 belong to signal transduction mechanisms, 1 to lipid transport and metabolism, 1 to cell wall/membrane/envelope biogenesis and 1 to defense mechanisms (Table 15). The gene mutated in all cultures was the gene MRY70_22200, protein WP_024128246.1 annotated as sigma 54-dependent transcriptional regulator.

Table 15 - Proteins affected during evolution in at least two Myr cultures. The table represents the mutated amino acid after translation to protein; the frequency of the mutation in all Lac cultures, the COG and the closest identity match through *Blastp*. In the mutation column, the first letter is the amino acid present in the ancestor strain, the numbers are the position where the mutation occurred, and the second letter is the amino acid present in the mutated protein. + represents an insertion and Δ a deletion.

Mutation	M1	M2	M3	M4	M5	Gene	Protein	COG	BLASTp
Q410* (CAG→TAG)					31.20%	MRY70_25930	**WP_009617479.1**	I	100% identity fatty acid-CoA ligase *Pseudomonas*
Q106* (CAG→TAG)		28.90%							
D643G (GAC→GGC)					30.20%	MRY70_07785	**WP_009616865.1**	M	100% identity MprF *Pseudomonas*
Δ3 bp			5.90%		15.90%				
+18 b.p.	10.50%								
+TAGGCTG		66.00%				MRY70_30280	**WP_009614484.1**	T P	100% identity TrkA *Pseudomonas*
Δ14 b.p.	11.80%								
A62T (GCC→ACC)					70.40%	MRY70_03390	**WP_009613088.1**	V	84.82% identity General stress protein 18 *P. citronellolis*
H10H (CAC→CAT)		23.60%							
G279E (GGA→GAA)		5.10%				MRY70_26305	**WP_009617364.1**	T	82.77% identity hybrid sensor hystidine kinase LadS *P. aeruginosa*
L438P (CTG→CCG)					26.40%				
L611Q (CTG→CAG)			6.10%						
N230S (AAC→AGC)		60.20%							
G500D (GGC→GAC)		34.80%				MRY70_07840	**WP_024128544.1**	T	76.94% identity GacS *P. aeruginosa*
S480F (TCC→TTC)					72.00%				
D303G (GAC→GGC)			5.10%			MRY70_22200	**WP_024128246.1**	T	91.48% identity FleQ *P. aeruginosa*
V271A (GTC→GCC)					70.40%				
E265K (GAG→AAG)	32.90%								
+GGCGCCGGTGAA	13.60%								
A227T (GCC→ACC)	5.60%								
+ 42 b.p.	37.30%	69.90%		73.20%	19.40%				
S177F (TCC→TTC)			70.80%						

After protein Blast (*Blastp*) in NCBI, the protein coded by the gene mutated in all 5 Myr cultures, showed 91.48% similarity with FleQ transcriptional regulator from *Pseudomonas aeruginosa*. This regulator is necessary to activate the transcription of genes responsible for flagella biosynthesis and

exopolysaccharides in response to c-di-GMP. FleQ is the main effector of c-di-GMP, also known for its role in the transition of planktonic state to biofilm in *P. aeruginosa*. The other mutated genes unique to Myr cultures showed highest identity for histidine-kinase proteins GacS and LadS, both involved in the signalling cascade pathway Gac/Rsm. This pathway regulates the expression of genes related to quorum sensing, secondary metabolism, biofilm formation and motility, being also relevant in the regulation of virulence and antibiotic resistance in pathogenic bacteria. Also, the gene *trkA*, coding for a constitutive potassium transporter, was mutated in 2 Myr cultures. This transporter was also reported to be related with virulence, antimicrobial resistance and biofilm formation. For culture M1, 14 b.p. of this gene were deleted, compromising the function of this gene. For culture M2, an insertion of 7 b.p. was detected. This insertion shifted the reading frame in the last 2 amino acids of the protein translated. Mutations in these 3 proteins may be related to adaptation to β-myrcene, taking into consideration the hydrophobic nature of the compound and the changes that have to occur at the cell membrane level for its uptake, although this process is still unknown. Besides these mutations, a general stress protein was mutated in 2 Myr cultures and the gene *mprF* was mutated in 3 Myr cultures. MprF stands for multiple peptide resistance factor protein, involved in resistance to the CAMP-like antibiotic daptomycin. In this gene, an insertion of 18 b.p. was detected for culture M1 and a deletion of 3 b.p. was detected in culture M3. These 2 genes may be related to adaptation for stationary phase growth.

The identity of the FleQ protein was confirmed in the genome of M1 strain, since this protein belongs to a cluster of flagellar proteins, being flanked by the genes coding for the proteins FliT and FleR, both known to regulate the transcription of flagellar operons, controlling flagellar synthesis.

Given these results, regulation mechanisms like histidine-kinase pathways may be relevant to the adaptation to β-myrcene. Moreover, the adaptation mechanisms to this hydrophobic compound may rely in the ability of this strain to change from a planktonic to biofilm state. For mutations in intergenic regions, 11 regions were mutated in at least three β-myrcene cultures. Interestingly, these mutations were not exclusive to β-myrcene growth conditions, they were also observed in lactate cultures. Therefore, these mutations are likely associated with the background genome rather than being specifically linked to adaptation for β-myrcene growth.

3.3.2. Strains isolated from β-myrcene adapted cultures showed increased performance to catabolize this monoterpene

After reaching the 600 generations, all cultures were tested in plate to check its resistance to three antibiotics – gentamycin, kanamycin, and tetracycline. These antibiotics are the most used as selection

markers in our lab. In fact, several cultures either adapted to β-myrcene or Lactate had their susceptibility to antibiotics decreased (Table 16).

Table 16 - Resistance test of β-myrcene and lactate cultures evolved until 600 generations, and ancestor strain to 30 μg/mL and 60 μg/mL of gentamycin, 50 μg/mL and 100 μg/mL of kanamycin, and 50 μg/mL and 100 μg/mL of tetracycline. The cells were plated in different concentrations, the first drop being without dilution and every other drop diluted 10x consecutively.

	Gm 30 μg/mL	Gm 60 μg /mL	Kan 50 μg /mL	Kan 100 μg /mL	Tc 50 μg /mL	Tc 100 μg /mL
M1	-	-	-	-	-	-
M2	+	+	+	-	-	-
M3	+	+	+	+	+	+
M4	+	+	+	+	+	+
M5	-	-	-	-	+	-
L1	-	-	-	-	+	-
L2	+	+	+	+	+	+
L3	+	+	+	+	+	+
L4	-	-	-	-	+	-
Anc	-	-	-	-	-	-

The purpose of this work was to select evolved isolates with increased efficiency towards the β-myrcene catabolism, thus around 300 isolates were retrieved from 600 generations β-myrcene adapted cultures. A vector carrying a promoter probe containing the promoter responsible for the transcription of TU5 (promoter P5) fused with the green fluorescent protein (GFP) gene, was already constructed, and used in previous works[126]. When the promoter P5 is activated, a fluorescent signal is produced. If the promoter P5 has a higher induction rate in any of the isolates, a stronger fluorescence signal will be detected. Thus, the 300 isolates were submitted to conjugations with this construct. However, only 150 isolates (30 isolates for each β-myrcene culture) were susceptible to the gentamycin selection marker present in this vector, disenabling the selection of properly conjugated isolates. This was possible to visualize thanks to the fluorescence levels reached for these isolates. Only for the 150 isolates referred, the fluorescence signal after induction with β-myrcene was different than the basal level of fluorescence observed for *Pseudomonas* sp. M1. The Relative Fluorescence Units (RFUs) for all the 300 isolates are depicted in supplementary Material Table S3. In this table, the isolates which fluorescence after β-myrcene induction is similar to the basal fluorescence, did not conjugate successfully. From the 150 isolates selected, only 8 isolates were chosen to be further characterized. These isolates were the ones within each culture with

apparent increased fold-change fluorescence in comparison to the ancestor strain (Table S3, Bold). For these 8 isolates, the presence of the vector was verified through the amplification of a region of the vector. Of the 8 isolates chosen, only 6 showed statistically significative higher fold-change of fluorescence when compared with the ancestor strain, when induced with β-myrcene for 5h (Figure 20). These isolates were the M2C7, M2C19, M3C22, M3C30, M5C28 and M5C29.

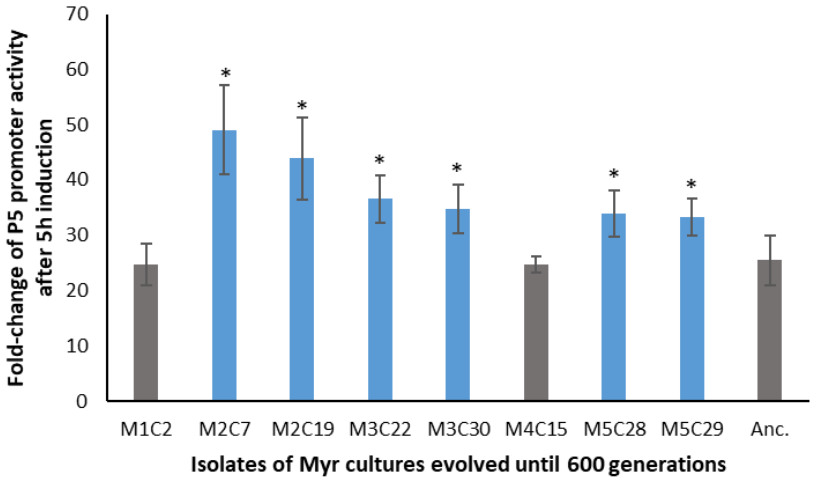


Figure 20 - Fold change (FC) of promoter P5 activity after induction with β-myrcene, estimated as the ratio of fluorescence units between β-myrcene grown cells and lactate grown cells, subtracting the contribution of intrinsic fluorescence. In blue are the isolates with significant increase in fold-change when compared with the ancestor strain. In grey are represented the ancestor strain as well the isolates with similar fluorescence levels to the ancestor strain. * - p-value >0.05.

Pseudomonas sp. M1 has intrinsic fluorescence. In this experiment, the ancestor and adapted counterparts (isolates) not containing the promoter probe were also grown in parallel and its fluorescence measured at the same time-points. Thus, in this experiment, M1 cells containing the promoter probe were grown in lactate and while in 1 inoculum the cells were induced with β-myrcene, the other inoculum was not induced, allowing the calculation of fold-change induction in the cells. Moreover, since the cells have intrinsic fluorescence, the same cells without the promoter probe were grown in parallel and their fluorescence measured. This fluorescence was subtracted to both induced and non-induced cells to withdraw the contribution of intrinsic fluorescence. Since the fluorescence values measured for cells with the promoter probe growing in lactate and the cells without promoter probe were almost the same, the system is almost 100% inducible. The genome of the 8 isolates was extracted and sequenced through Illumina Next-seq. The strains M2C7, M5C28 and M5C29 were not used forward since a hypermutant genotype was present, with mutations in *mutL* or *mutS* genes in the 3 strains. Since this work is about biotechnology and the purpose is to develop a mutant able to be further used for biotechnological applications, the genome instability of a hypermutant is not desirable. When constructing a whole-cell

biocatalyst, microorganisms are engineered to have only essential genes, minimizing its genome, and increasing its stability[215,216]. For instance, Lee J. and collaborators, increased the production of l-threonine by 83% in a *E.coli* strain with minimized genome, as well as the genome stability increased by 10-fold[217]. Moreover, as already referred, several isolates showed decreased antibiotic susceptibility. This effect may be attained with increased mutation frequency, originating mutants that are difficult to work with, which is an undesired trait for biotechnological applications. Thus, the work was proceeded with the non-hypermutant isolates with higher fold-change: M2C19 and M3C22 with 2.5 and 2.4 FC ratio respectively. Only 2 mutations were detected for each isolate, both in coding regions (Table 17).

Table 17 - Proteins affected in the 2 evolved isolates selected after 600 generations. The table represents the mutated amino acid after translation to protein, which mutant has the mutation and the closest identity match through *Blastp*. In the mutation column, the first letter is the amino acid present in the ancestor strain, the numbers are the position where the mutation occurred, and the second letter is the amino acid present in the mutated protein. + represents an insertion.

Mutation	M2C19	M3C22	Gene	Protein	Blastp
+GGTGAAGGCGCC	X		MRY70_22200 ←	**WP_024128246.1**	91.48% identity FleQ *P. aeruginosa*
L211Q (CTG→CAG)		X	MRY70_22200 ←	**WP_024128246.1**	91.48% identity FleQ *P. aeruginosa*
L611R (CTG→CGG)	X		MRY70_26305 ←	**WP_009617364.1**	82.77% identity hybrid sensor histidine kinase/response regulator LadS *P. aeruginosa*
L250Q (CTG→CAG)		X	MRY70_20160 →	**WP_009623403.1**	79.45% identity two-component system sensor histidine kinase/response regulator SagS *P. aeruginosa*

The low number of mutations in these isolates were already expected since they do not possess hypermutant characteristics. In comparison, the isolates M2C7, M5C28 and M5C29 accumulated around 120 mutations each. For M2C19 strain an insertion of 12b.p. in the gene MRY70_22200 - corresponding to the previously identified gene *fleQ* - and a SNV leading to a non-synonymous translation into amino acid in the gene MRY70_26305 were detected. The protein coded by this gene was identified as the hybrid sensor histidine kinase/response regulator LadS with 82.77% identity in *P. aeruginosa*. In *P. aeruginosa*, LadS interacts with the GacS/GacA system related to the motile/sessile lifestyle of the cells. When phosphorylated by LadS, GacS/GacA system acts as a transcriptional activator, upregulating indirectly genes responsible for biofilm sessile state. As for M3C22 strain, the two detected mutations were SNVs leading to a non-synonymous translation detected on genes MRY70_20160 and MRY70_22200. Once again FleQ protein is mutated, being this mutation common between both strains

and appearing in all the adapted cultures, as previously stated. The importance of this gene for the adaptation to β-myrcene, may be related to the ability for the cells switch between mobile and sessile states which may be necessary for β-myrcene detection and uptake. As stated above, the aggregation of cells during wastewater treatment increases the metabolism of hydrophobic compounds. This may also be true for M1 cells during the catabolism of hydrophobic compounds, namely β-myrcene. The gene MRY70_20160 codes for a histidine-kinase protein, with 79.45% identity for the response regulator SagS in *Pseudomonas aeruginosa*. This sensor regulator is essential for the formation of biofilms in *P.aeruginosa*. Olga E. Petrova and collaborators demonstrated that when this regulator is deleted, the attachment of *P.aeruginosa* increases and biofilm formation is abrogated. Since all mutations for both strains hint for the switch between planktonic/sessile state of the cells, simple motility tests were performed for the 2 adapted strains versus the ancestor strain, in LB medium containing only 0.3% agar. In fact, both adapted isolates had its motility decreased when compared with the ancestor strain hinting that the increase in sessile state may be advantageous for the β-myrcene catabolism (Figure 21).

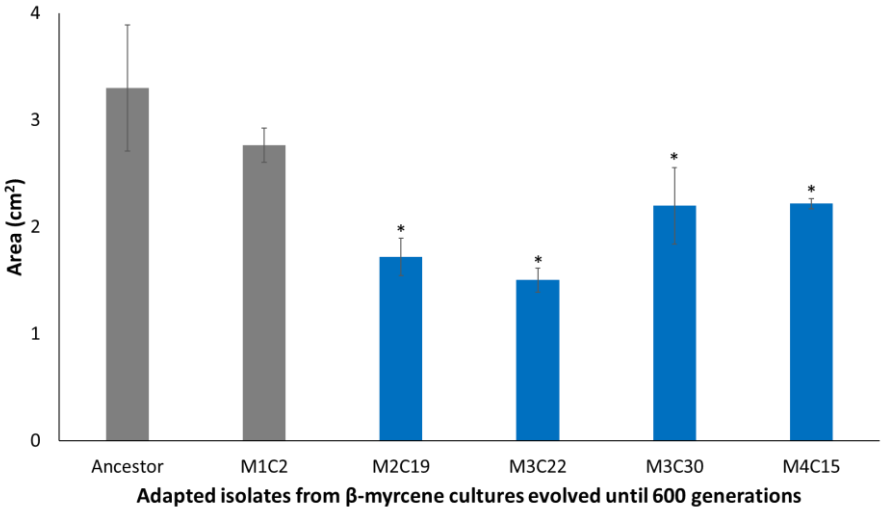


Figure 21 - Area in square centimetres of the growth halo after 24 hours growth in plates, for swimming motility assay, in all evolved isolates vs ancestor strain. The LB plates used contained only 0.3% agar, the standard Agar percentage for Swimming experiments. M1C2, M3C30 and M4C15 were used as controls to the experiment since mutations in *fleQ* gene were not detected. M2C19 and M3C22 significantly decreased motility when compared to the ancestor strain. *- pvalue <0.05.

A higher activity of the GI promoter P5 hints the increased efficiency of the evolved isolates, mainly M2C19 and M3C22 isolates, towards the catabolism of β-myrcene. These were the only 2 isolates containing mutations in the *fleQ* gene. To check if there were increases in the specific growth rate of the discovered isolates due to the adaptation process, mainly in the isolates M2C19 and M3C22, the growth

of all isolates non-hypermutant was monitored, and the specific growth rate calculated for each isolate (Figure 22).

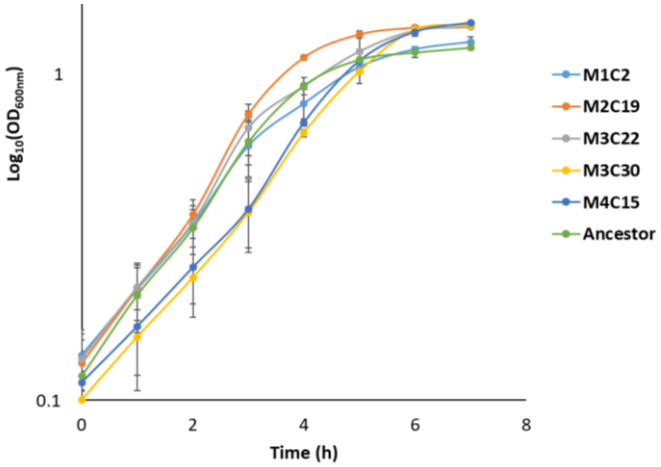


Figure 22 - Growth monitoring of all the evolved isolates without hypermutant genotype and ancestor strain. The specific growth rates calculated for each strain were: M1C2- 0.55h^{-1}, M2C19- 0.71h^{-1}, M3C22- 0.67h^{-1}, M3C30- 0.56h^{-1}, M4C15- 0.61h^{-1}, Ancestor- 0.61h^{-1}.

In fact, the isolates with the highest growth rate were the M2C19 and M3C22 isolates.

To understand whether *fleQ* gene was responsible for this behavior and for the increased P5 promoter activity, the gene was deleted from M1 *wt* genome since loss of motility hinted the loss of function of this gene. The deletion of *fleQ* was accomplished through a new technique developed by Aparicio and colaborators[71] which allows the deletion or targeted mutation of a gene through homologous recombination, targeting the non-recombinants with CRISPR-Cas9 and thus increasing the yield of the method. To execute this technique three plasmids are necessary: pSEVA2316, pSEVA658-ssr and pSEVA421-Cas9tr. The plasmid pSEVA658-ssr harbors the sequence for the expression of a recombinase. Upon the insertion of oligonucleotides complementary to the flanks of *fleQ* gene sequence, the recombinase will allow recombination and thus the deletion of the gene. The plasmid pSEVA2316 will harbor a complementary sequence to the target gene, in this case, the gene *fleQ* and the plasmid pSEVA421-Cas9tr harbors the endonuclease Cas9 that will be guided by the guide-RNA (containing the sequence from the plasmid pSEVA2316) to cut the target genes. Thus, all the cells that did not recombine, will have their DNA cut by the Cas9 endonuclease and will die.

The deletion of *fleQ* gene was well succeeded in the M1 *wt* strain, showing a phenotype similar to the evolved isolates M2C19 and M3C22, with a decreased motility in comparison to the ancestor strain (Figure 23).

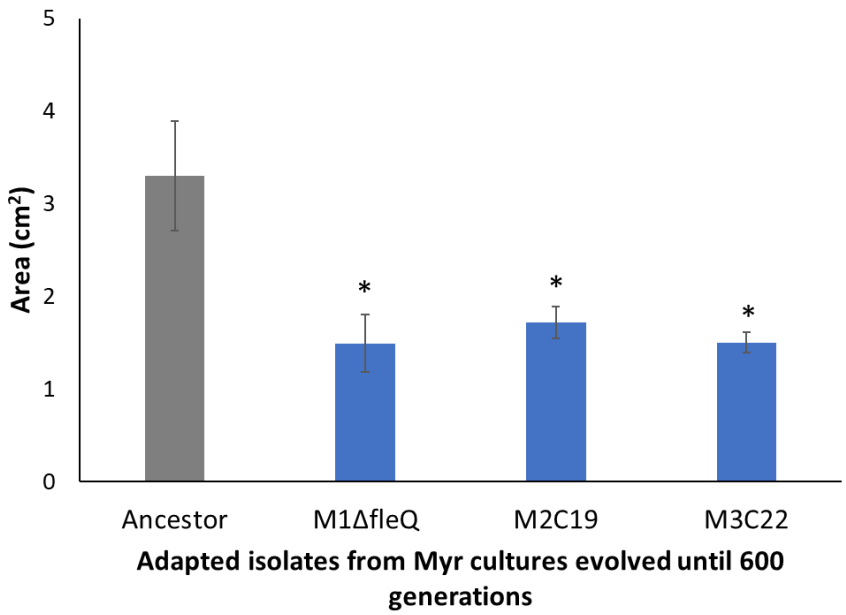


Figure 23 - Area in square centimeters of the growth halo after 24 hours growth in plates, for swimming motility assay, in M2C19 and M3C22 evolved isolates, M1 *wt* with *fleQ* gene deleted and ancestor strain. The LB plates used contained only 0.3% agar. *p-value <0.05.

M1ΔFleQ mutant was also transformed with the promoter probe containing the construction P5 promoter + GFP. At this stage, the accumulation of GFP was measured for 24h for the M1ΔFleQ mutant, both adapted isolates – M2C19 and M3C22 – and the ancestor strain (Figure 24). These results show that after 5 hours of induction the adapted isolates already have around 2.5 times more fluorescence than the ancestor strain, corroborating the previous results. However, in the M1ΔFleQ mutant, the accumulation of fluorescence at 5h induction is not much different from the ancestor strain. Only later in growth, at 24h growth, the mutant shows an increased fluorescence in comparison with the ancestor strain. Moreover, despite the mutant showing higher P5 activity than the ancestor at 24h, this increase is still lower than the adapted isolates. Thus, the deletion of the *fleQ* gene did increase the activity of P5 promoter, showing its importance for the efficient catabolism of β-myrcene, however, the deletion alone did not reach the results showed for the adapted isolates. Thus, the gene *fleQ* seems to have a contribution towards the adaptation of *Pseudomonas* sp. M1 to β-myrcene, however, this contribution may go beyond this transcriptional regulator fully inactivation, and rather rely in other mechanisms.

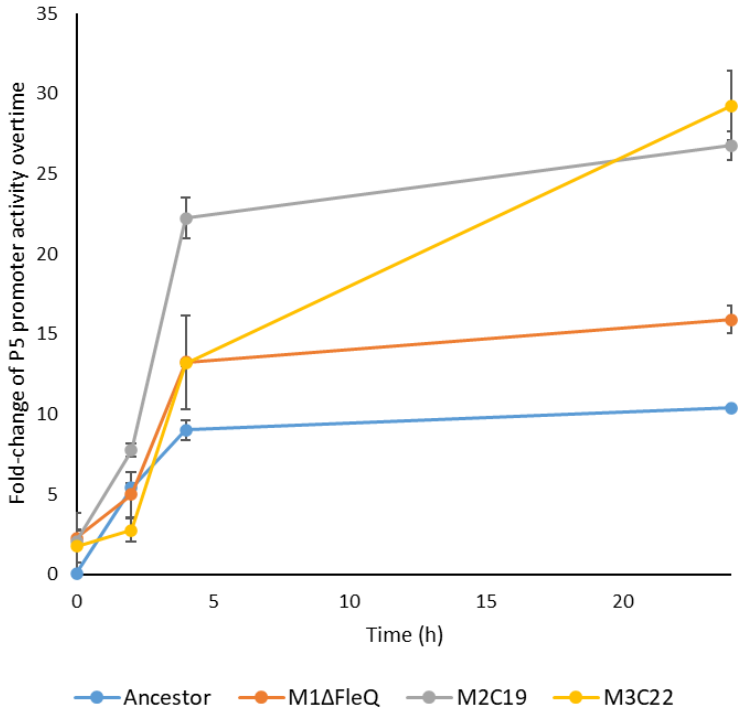


Figure 24 - Fold change (FC) of promoter P5 activity after induction with β-myrcene, estimated as the ratio of fluorescence units between β-myrcene grown cells and lactate grown cells, subtracting the contribution of intrinsic fluorescence, measured overtime. The assay was performed for the ancestor strain (blue), M1ΔFleQ mutant (orange), M2C19 adapted isolate (grey) and M3C22 adapted isolate (yellow).

Two hypotheses arose from these results: i) mutations in the *fleQ* gene have to be specific, instead of completely deactivate its coding protein. Thus, FleQ does not have its activity entirely abrogated and only its interaction with its targets is affected or ii) mutations in the *fleQ* gene must be in syntony with mutations in certain histidine-kinase sensor proteins related with the planktonic/sessile switch of the cells, since both adapted strains had mutations in at least 1 sensor histidine-kinase, despite not being the same gene.

To better understand the swimming capacity of the M1ΔFleQ strain and evolved isolates when β-myrcene is used as only carbon source, plates of Minimal medium with 0.3% agar supplemented with β-myrcene or lactate were done. As expected by the previous results, the motility in lactate for all the mutants is considerably decreased in comparison to the ancestor strain. However, all mutants showed higher motility when in β-myrcene in comparison to lactate (Figure 25). Moreover, the motility of all mutants was higher in plates supplemented with β-myrcene, in comparison to the ancestor strain (Figure 25, blue data). Besides this, the motility in β-myrcene in the evolved isolates vs M1ΔFleQ strain was not significatively different.

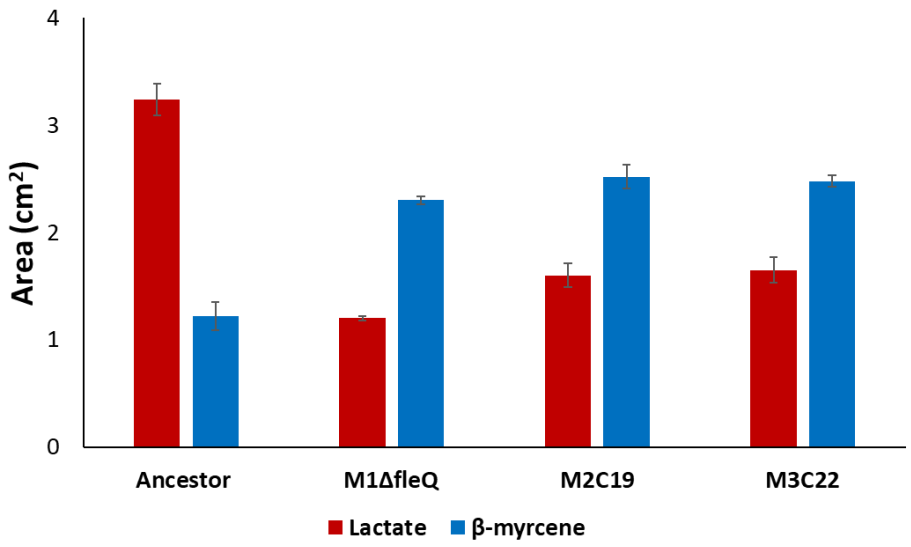


Figure 25 - Area in square centimetres of the growth halo after 24 hours growth in plates, for swimming motility assay. The plates used were composed by Minimal Medium with 0.3% agar, supplemented with β-myrcene or lactate as carbon sources. The data in blue, including the plate, are from β-myrcene supplementation and data in red including the plate, are from lactate supplementation.

To better understand this phenomenon, fluorescence microscopy was performed, since the P5 promoter was coupled to a GFP protein. In the microscope, the behavior of the cells is more easily observed. The behavior of M1 *wt* cells and derived strains, grown using β-myrcene as only carbon source, is depicted in Figure 26. In *Pseudomonas* sp. M1 *wt* cells, no major aggregation of cells is observed, being the planktonic state the preferred one (Figure 26 A). However, when the *fleQ* gene is deleted, the cells form several aggregates, and the individual cells seem to be more elongated (Figure 26 B). When the evolved mutants M2C19 and M3C22 are observed in the microscope, some aggregates are formed, however, in a lower number and with a bigger size than the ones formed by M1ΔFleQ mutant (Figure 26 C and D). The increased expression of P5 promoter in both evolved isolates and in the M1ΔFleQ mutant is most probably related with the switch between planktonic and sessile states since the mutations accumulated by both isolates were only related with this function. To further understand how the mutations in *fleQ* gene influence the protein behavior, the position in which the mutations happened was studied. The insertion +GGTGAAGGCGCC in M2C19 isolate, besides adding 4 amino acids to the protein, happens inside Loop L1 of the protein. In its structure, FleQ protein contains 2 loops designated as the specific σ54-interaction loops. These loops are responsible for ATP binding and hydrolysis necessary for the transcription complex remodeling, so the transcription of the genes associated may happen. Loop L1 has a signature conserved motif in all FleQ proteins – GAFTGA – allowing its identification for this work.

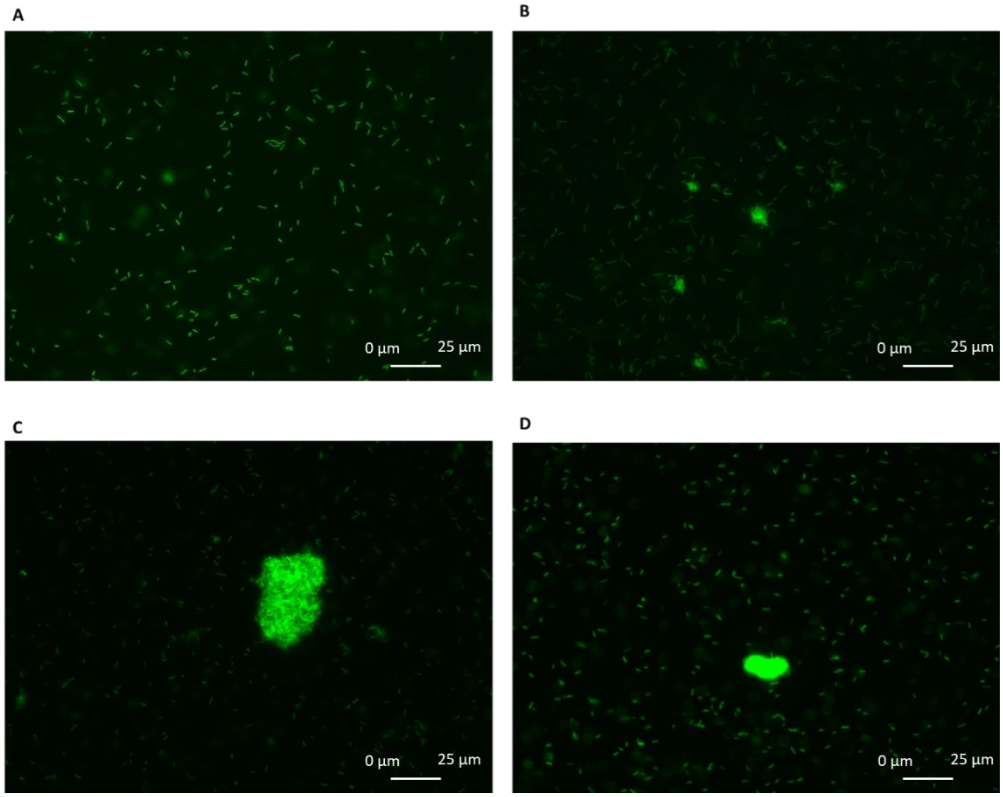


Figure 26 - Fluorescence microscopy of M1 *wt* cells (A), M1ΔFleQ (B) and the evolved isolates M3C22 (C) and M2C19 (D) containing the plasmid pSEVA637 with the promote-probe P5-GFP, after growth in MM supplemented with β-myrcene as only carbon source. The image is magnified by a factor of 400x.

The region where the insertion happened for M2C19 mutant was in the amino acid right before the signature motif of L1. For the isolate M3C22, a non-synonymous SNV occurred also in this loop, only 11 amino acids before the signature motif of L1. Thus, possibly, the behavior of M2C19 and M3C22 evolved isolates is related with the increased or decreased ATP binding and hydrolysis by FleQ protein, increasing or decreasing the transcription of the associated genes, without abolishing its transcription completely. Moreover, the mutations in histidine kinases associated with planktonic/sessile lifestyle may complement the FleQ function and increasing or decreasing the expression of different genes associated with motility upon the detection of β-myrcene particles in the media.

In summary, *Pseudomonas* sp. M1 was subjected to adaptive laboratory evolution (ALE) for 600 generations in two different carbon sources – lactate and the monoterpene β-myrcene. Upon total DNA sequencing, the number of mutations detected for the cultures adapted to lactate was higher than for the cultures adapted to β-myrcene with 2.9% more mutations at 200 generations, 4.7% at 400 generations and 5.3% at 600 generations. From the cultures adapted to β-myrcene after the 600 generations adaptation, several isolates were retrieved and transformed with a promoter probe harboring

a construction of the promoter P5 (which allows the expression of important genes for the β-myrcene catabolism) fused with GFP. Upon RFUs measurement, 6 evolved isolates showed higher activity of promoter P5 in comparison to the ancestor strain, hinting adaptation to β-myrcene. However, only 2 isolates – M2C19 and M3C22 – had a non-hypermutant genotype, thus the work was further focused on these 2 isolates. Interestingly, both isolates accumulated mutations in the gene MRY70_22200, identified as the FleQ transcriptional regulator upon *Blastp*. This regulator is necessary to activate the transcription of genes responsible for flagella biosynthesis and exopolysaccharides in response to c-di-GMP. FleQ is the main effector of c-di-GMP, also known for its role in the transition of planktonic state to biofilm in *Pseudomonas aeruginosa*. In fact, the motility of both evolved isolates was decreased in comparison with the ancestor strain, when grown in LB media with 0.3% Agar. To better understand the role of this regulator for the adaptation to β-myrcene, a mutant lacking the gene *fleQ* was constructed – M1ΔFleQ. Upon transformation of this mutant with the promoter-probe referred previously, the activity of promoter P5 was higher than the ancestor strain after 24h induction with the monoterpene, however lower than the activity of the same promoter in the adapted isolates. Two hypotheses were raised: i) mutations in the *fleQ* gene have to be specific, instead of completely deactivating its coding protein. Thus, FleQ does not have its activity entirely abrogated and only its interaction with its targets is affected or ii) mutations in the *fleQ* gene must be in syntony with mutations in certain histidine-kinase sensor proteins related with the planktonic/sessile switch of the cells, since both adapted strains had mutations in at least 1 sensor histidine-kinase, despite not being the same gene. To further clarify if the evolved isolates and constructed mutant have increased efficiency in comparison to the ancestor strain and why, the next chapter will be about protein identification and quantification for M1 *wt*, M2C19, M3C22 and M1ΔFleQ strains, when grown in minimal media with only β-myrcene as carbon source. In the next chapter, the proteome profile of all isolates will be stablished, allowing to estimate the relative abundance of specific proteins necessary for the catabolism of β-myrcene (proteins expressed by the genes harbored in the 28-kb genomic *locus* – GI), as well as the relative abundance of proteins targeted by the transcriptional regulator FleQ, among other.

Chapter IV – Proteome assessment of *Pseudomonas* sp. M1 *wt* and M1-derived strains – M2C19, M3C22 and M1ΔFleQ

4.1. Introduction

The importance of studying the proteome of an organism is due to the role of proteins as the main effectors of the cells. Thus, the analysis of the proteome may provide a global insight on how a biological system works and adapts to different conditions. The proteome is not constant, differs from cell to cell, depends on the conditions imposed and changes overtime[218,219]. After adaptive laboratorial evolution, changes in the proteome of adapted strains in comparison with the ancestor strain are expected. Thus, proteomics is very helpful to understand the implications of evolution and adaptation in the organisms, elucidating which cellular processes undergo significant alterations and indicating proteins implicated in the targeted biological processes during adaptation. Proteomics enables comparative analysis between different strains or conditions, allowing the identification of common adaptive responses as well as strain-specific adaptations[220]. In the biotechnology context, where living organisms are used to produce modified compounds, proteomics assumes furthest importance in determining the rates of protein production, degradation, and steady-state abundance. This is necessary for maintaining consistent production yields. Typically, proteomics is closely connected with mass spectrometry techniques, particularly liquid chromatography-tandem mass spectrometry (LC-MS/MS)-based approaches, facilitating the identification and relative quantification of proteins[27]. The most prevalent method is data-dependent acquisition (DDA) mass spectrometry. The DDA method requires enzymatically digested proteins into smaller peptides, which are then separated via high-performance liquid chromatography (HPLC) and fed into the mass spectrometer. Within the mass spectrometer, the peptides are ionized, followed by a MS1 scan where their abundance and mass-to-charge ratios (m/z) are recorded. Subsequently, a subset of the most abundant ions is isolated and subjected to fragmentation in the MS2 scan, where the abundance and m/z ratios of the resulting fragments are measured and logged[31,221,222]. DDA has been widely used and optimized in proteomics workflows, making it a reliable and well-established method for protein identification and relative quantification. DDA has earned popularity due to its versatility, ease of setup, straightforward data analysis, and its capacity to detect proteins across a wide dynamic range[221]. In this chapter, the proteome of *Pseudomonas* sp. M1 *wt* and derived strains (M2C19, M3C22 and M1ΔFleQ) will be identified and compared. By identifying and comparing the proteomes of these strains, the relative abundance of the identified proteins will be determined, thus giving insights into the biological processes affected in the adapted/mutant strains. Moreover, the relative quantification of proteins known to be essential for the β-myrcene catabolism will clarify whether these strains exhibit enhanced efficiency in the catabolism of β-myrcene. This analysis will provide valuable insights into the metabolic adaptations and potential improvements in β-myrcene degradation efficiency in the derived strains.

4.2. Material and Methods

In order to verify the effect of mutations in the proteins expressed by the cells, a pipeline for proteomics analysis was followed.

4.2.1. Protein Extraction

The whole cells protein content was extracted for M1 *wt*, two M1 evolved isolates (M2C19 and M3C22) and M1ΔFleQ strain. Firstly, all strains were grown overnight in MM supplemented with lactate or β-myrcene at 30ºC with agitation. In the morning, the cells were refreshed for the same media and grown until OD_{600nm} of 0.5 or OD_{600nm} 0.8. After growth, the cells were centrifuged, and the pellet collected. Cell lysis was done through a combination of chemical lysis and mechanical lysis. For this, cells were resuspended in Urea lysis buffer (9 M urea, 50 mM Tris HCl pH 6.8, 10% glycerol, 2% DTT and 10% SDS) and submitted to 15 cycles of sonication with 10 second pulses.

After lysis, the samples were centrifuged at 4ºC, maximum speed for 10 minutes. The supernatant was retrieved, and the samples were quantified. All this procedure was repeated in triplicate for the different strains and conditions.

4.2.2. Protein Quantification

Proteins were quantified using the modified Lowry method. Firstly, a calibration curve using BSA resuspended in Urea lysis buffer in different concentrations (0, 20, 40, 80, 120, 160, 200 μg/mL) was done. To all the samples with different protein concentrations, DOC 1% (50 μL) and TCA 10% (1 mL) were added to precipitate the proteins in the samples. After centrifugation, the protein pellet was resuspended in 1 mL of solution C (100 parts of Solution A - 2% (w/v) Na_2CO_3, 0.4% (w/v) NaOH, 0.16% (w/v) Na-Tartrate, 1% (w/v) SDS) mixed with 1 part of Solution B - 4% (w/v) $CuSO_4.5H_2O$) and after incubation, 100 μL of Follin reagent (1:2 dilution) was added to all samples. After 30 minutes incubation, the samples were centrifuged for 2 min at 14.000 *xg* and the OD_{750nm} was measured. Thus, the different protein concentrations can be correlated to different OD_{750nm} allowing the construction of a calibration curve.

The same procedure was applied to all the samples, using different dilutions, allowing its correct quantification.

After quantification, 300 μg of protein for each sample were run in SDS-PAGE gels.

4.2.3. SDS-PAGE

SDS-PAGEs were performed in 10% polyacrylamide gels according to the method described by Laemmli[223]. The gels were colored using Coomassie blue G-250. The coloring excesses were removed with acidic water.

4.2.4. LC-MS/MS measurements

All the procedure described above was performed, analyzed and written in collaboration with Professor, Doctor Dirk Benndorf and Patrick Hellwig Ph.D. student in Bioprocesses department of Otto Von Guerrick University in Magdeburg, Germany.

To measure proteins in LC-MS/MS, 25 μg of each protein sample were filtered and washed with Urea Buffer II (8 M Urea, 0.1 M Tris-HCl pH 8.5). Afterwards, the proteins were reduced using 100 μL of 40 mM 1,4-Dithiothreitol (DTT) followed by incubation for 20 minutes at 56°C and in darkness and alkylated using 100 μL of 55 mM Iodoacetamide (IAA) followed by incubation for 20 minutes at room temperature and in darkness. Afterwards the proteins were once again washed with Urea Buffer II and ABC buffer (50 mM Ammonium carbonate – NH_4HCO_3 – pH 7.8). After washing, the proteins were cleaved with Trypsin in an enzyme – substrate ratio of 1:100 and incubated for 2 hours at 37°C with agitation. The samples were finally eluted from the filter using Extraction buffer (50 mM ABC buffer, 5% Acetonitrile MS/MS grade) and LC-MS grade MilliQ water, with centrifugations for 5 min, 10000 *xg* at room temperature. The samples were dried in speed vacuum overnight and resuspended in 75μL of Load A Buffer (0.1% Trifluoracetic acid (TFA) LC/MS grade).

The peptides were analyzed by LC-MS/MS using an UltiMate® 3000 nano splitless reversed-phase nanoHPLC (Thermo Fisher Scientific, Dreieich) coupled online to a timsTOF™ pro 1 mass spectrometer (Bruker Daltonik GmbH, Bremen). The chemicals used were all in LC-MS grade.

The peptides were trapped on a trap column (Dionex Acclaim, nano trap column, 100 μmi.d. x 2 cm, PepMap100 C18, 5 μm, 100 Å, nanoViper) with a flow rate of 15 μL/min loading A buffer (100% water, 0.1% TFA). Afterwards, the peptides were loaded on Dionex Acclaim PepMap C18 RSLC Nano-reversed phase column (2 μm particle size, 100 Å pore size, 75 μm inner diameter, and 500 mm length) for a chromatographic separation. The column temperature was 60°C and a flow rate of the nanoflow with a binary solvent A/B gradient (solvent A: 100% ultrapure water, 0.1% formic acid; solvent B: 100% acetonitrile, 0.1% formic acid) was 400 nL/min. 3 μL of the sample was injected. The separation started with 2% solvent B for 5 min and the proportion of solvent B was linearly increased over a 120 min gradient to 35%. Afterwards, the separation column was washed with 95% solvent B for 10 min and re-equilibrated

with 5% loading B for 20 min. At the same time, the trap column was washed with 95% loading B buffer (100% acetonitrile, 0.1% TFA) for 10 min and equilibrated with 100% loading A buffer (100% water, 0.1% TFA) for 20 min at a flow rate of 45 µl/min.

The positive measurement mode of the MS and a data-dependent MS/MS method were used to measure the peptides. The precursor ions were selected with Trapped Ion Mobility Separation (TIMS) and the acquisition method Parallel Accumulation Serial Fragmentation (PASEF) from 0.6-1.6 V s cm^{-2}. The precursors were trapped in a quadrupole with a resolution of 50.000 and a mass-to-charge ratio (m/z) ranging from 100-1.700. Subsequently, the fragment ions were produced from the ten most intense precursors (charge 1 - 5) of each spectrum (10 PASEF MS/MS scans per cycle) with collision-induced dissociation (CID). The resulting fragment ions were analyzed in a TOF analyzer. The analysis takes place with an active exclusion of the same precursor after 1 spectrum for 24 seconds or a release if intensity/previous intensity exceeded 4.

The timsTOF™ mass spectrometer MS/MS measurement raw data files were processed by the Compass DataAnalysis software (version 5.3.0, Bruker Corporation, Bremen, Germany) and converted into Mascot Generic Files (.mgf). Then, the files were uploaded to MASCOT Daemon (Version 2.6.0)[224] and searched against a customized database:

Fixed modification	Carbamidomethyl (M, +57.021464 Da)
Variable modification	Oxidation (+ 15.994915 Da)
Enzyme	Trypsin
Missed Cleavage	1
Peptide charge	2+, 3+ and 4+
Peptide tolerance	0.02 Da
MS/MS tolerance	0.02 Da
FDR	≤0,01

The search results of each sample were exported as .dat file and uploaded into the MPA software version 3.0. The samples were compared without generating metaproteins[225].

4.2.5. Results analysis

For the statistical analysis, the relative abundance of each protein was normalized taking into consideration the total number of proteins detected for each sample, since this number was not exactly the same for all the analyzed samples. Afterwards, the fold-change (FC) was calculated between the relevant conditions as a ratio. The adjusted p-value was calculated using the Benjamini and Hochberg

approach[226]. For further analysis, the Log_{10}(adjusted p-value) was used, considering as cutoff value, the cutoff of the False Discovery Rate; as well as the Log_2(FC), being considered as positive the values 1.5-fold higher or lower than the reference chosen.

4.3. Results

4.3.1. Overall analysis

In the previous chapter, *Pseudomonas* sp. M1 underwent a 600-generation Adaptation and Laboratorial Evolution (ALE) process. This involved independent growth of cells in MM, using β-myrcene or lactate as sole carbon sources. The primary objective of this experiment was to enhance growth efficiency, particularly targeting β-myrcene, a hydrophobic monoterpene. The outcome of this experiment was the enhancement in efficiency among several isolated colonies by the 600th generation. Among these, the isolates M2C19 and M3C22 emerged as the most promising candidates, showing the highest fold-change in P5 promoter activity compared to the ancestral strain. The P5 promoter plays a crucial role in converting β-myrcene into its alcohol form, which serves as a pivotal step in β-myrcene catabolism. The only mutated gene shared between both isolates was identified as the *fleQ* gene. Consequently, a mutant strain depleted of the *fleQ* gene was engineered, revealing an improvement in growth efficiency on β-myrcene, though not as pronounced as that observed in the evolved isolates. Either way this mutant corroborated the role of the *fleQ* gene in the adaptation of *Pseudomonas* sp. M1 to β-myrcene. To validate the increased efficiency in β-myrcene catabolism observed in these isolates/mutant, and to unravel the mechanisms driving this enhancement, proteomics emerged as the tool of choice. This strategy allows the assessment of relative abundance of proteins present within the cells under specific conditions and growth phases, shedding light on the intricate processes at play.

In this experiment, cell samples for protein extraction were collected at OD_{600nm} of 0.5 and 0.8 across all tested isolates: M2C19, M3C22, M1ΔFleQ mutant and ancestral strain. Additionally, each isolate was cultivated using either lactate or β-myrcene as the sole carbon source, with triplicates considered. This resulted in four distinct conditions for each isolate: OD_{600nm}0.5 β-myrcene, OD_{600nm}0.8 β-myrcene, OD_{600nm}0.5 lactate, and OD_{600nm}0.8 lactate. The analysis of Table 18 reveals a significant disparity in the number of proteins identified for the M1ΔFleQ mutant compared to other isolates. Specifically, at OD_{600nm} 0.5, the protein count for the mutant is approximately 4-fold lower than that of the other isolates, while at OD_{600nm} 0.8 was roughly two-fold lower. Consequently, the results concerning the M1ΔFleQ mutant should be

addressed with caution. Furthermore, due to this divergence, the analysis for this mutant will be exclusively conducted under conditions corresponding to OD_{600nm} 0.8.

4.3.2. Overall fold change of protein abundance between β-myrcene and lactate cultured cells

For the first analysis, the proteins identified when all isolates were cultured in β-myrcene versus when they were grown in lactate were compared. The fold-change (ratio between both conditions) was calculated. To make possible the calculation of fold-change for proteins that were not detected in lactate (proteins specific for β-myrcene growth), a value of 1 was added to the relative abundance of all proteins in all conditions, allowing the fold-change calculation without affecting its ratios.

Table 18 - Number of proteins identified for each strain and condition. Each strain was grown in MM supplemented with β-myrcene or lactate and grown until an OD_{600nm} 0.5 or 0.8.

Strain	Condition	Number of proteins identified
Ancestor	β-myrcene OD_{600nm} 0.5	1937
	β-myrcene OD_{600nm} 0.8	2510
	lactate OD_{600nm} 0.5	2306
	lactate OD_{600nm} 0.8	2153
M2C19	β-myrcene OD_{600nm} 0.5	2209
	β-myrcene OD_{600nm} 0.8	2256
	lactate OD_{600nm} 0.5	1573
	lactate OD_{600nm} 0.8	2192
M3C22	β-myrcene OD_{600nm} 0.5	1795
	β-myrcene OD_{600nm} 0.8	2222
	lactate OD_{600nm} 0.5	2186
	lactate OD_{600nm} 0.8	2364
M1ΔFleQ	β-myrcene OD_{600nm} 0.5	580
	β-myrcene OD_{600nm} 0.8	1280
	lactate OD_{600nm} 0.5	942
	lactate OD_{600nm} 0.8	1347

The analysis focused on proteins with a fold-change greater than 1.5 or lower than -1.5. This criterion was chosen to find proteins with substantial differences in abundance across different growth conditions. All the fold-change values between β-myrcene and lactate growth for all strains (Ancestor, M2C19, M3C22, and M1ΔFleQ) at both growth stages (OD_{600nm}) along with their corresponding adjusted p-values is provided in supplementary material, Table S4. From the entire pool of identified proteins, only those

meeting the fold-change criteria >1.5-fold indicating higher abundance in β-myrcene conditions compared to lactate, or <-1.5-fold indicating lower abundance in β-myrcene versus lactate were considered. The proteins meeting the specified criteria were counted, as summarized in Table 19. This step provided an overview of proteins with significant changes in abundance under the stated conditions. A notable trend emerged where the number of proteins identified was greater when the fold-change was lower than -1.5-fold compared to when it was higher than 1.5-fold for the evolved isolates and for the ancestor strain in the OD_{600nm} 0.5. This suggests a widespread decrease in protein expression when cells are cultivated in β-myrcene, aligning with findings from earlier RNA-seq analyses for the *Pseudomonas* sp. M1 wild-type strain[128]. However, for the ancestor strain grown to an OD_{600nm} 0.8 and for the M1ΔFleQ mutant, this trend was not followed.

Table 19 - Number of proteins at least 5 times more abundant in growth with β-myrcene than in lactate (FC > 1.5) or 5 times less abundant in growth with β-myrcene than in lactate (FC < -1.5).

Strain	Ratio	Number proteins FC>1.5	Number proteins FC<-1.5
Ancestor	β-myrcene/Lactate OD_{600nm} 0.5	233	582
	β-myrcene/Lactate OD_{600nm} 0.8	496	442
M2C19	β-myrcene/Lactate OD_{600nm} 0.5	248	643
	β-myrcene/Lactate OD_{600nm} 0.8	262	896
M3C22	β-myrcene/Lactate OD_{600nm} 0.5	446	471
	β-myrcene/Lactate OD_{600nm} 0.8	248	1031
M1ΔFleQ	β-myrcene/Lactate OD_{600nm} 0.8	354	242

To ensure statistical significance, only proteins meeting the fold-change criteria alongside a p-value below 0.1 were considered. While the standard p-value of 0.05 is commonly employed, the p-value 0.1 was chosen in the context of this work, to capture potential statistical trends without sacrificing valuable data. Thus, the number of proteins following the fold-change and p-value criteria is depicted in Table 20.
Considering the significance threshold of p-value < 0.1, it was observed that the frequency of protein identification was higher when the fold-change was <-1.5, compared to when it exceeded 1.5 across all strains, except for the M1ΔFleQ mutant. In this case, this may be due to the low significance levels (higher p-values) experienced for the proteins with a FC<-1.5, since this sample only had 43 proteins with an FC<-1.5 and a p-value lower than 0.1 which is a low protein number in comparison with the other samples.

Table 20 - Number of proteins at least 5 times more abundant in growth with β-myrcene than in lactate (FC > 1.5) or 5 times less abundant in growth with β-myrcene than in lactate (FC < -1.5) and p-value lower than 0.1.

Strain	Ratio	Number proteins FC>1.5	Number proteins FC<-1.5
Ancestor	β-myrcene/Lactate OD_{600nm} 0.5	213	732
	β-myrcene/Lactate OD_{600nm} 0.8	87	174
M2C19	β-myrcene/Lactate OD_{600nm} 0.5	198	495
	β-myrcene/Lactate OD_{600nm} 0.8	161	517
M3C22	β-myrcene/Lactate OD_{600nm} 0.5	261	341
	β-myrcene/Lactate OD_{600nm} 0.8	60	524
M1ΔFleQ	β-myrcene/Lactate OD_{600nm} 0.8	150	43

For a deeper understanding of the distinct biological processes implicated in β-myrcene growth as opposed to lactate, an analysis of the overall Clusters of Orthologous Groups (COGs) for each strain was performed (Figure 27). In Figure 27A (FC > 1.5), the COGs J and C stand out consistently across all strains, holding a significant number of associated proteins. These COGs primarily comprise proteins linked to Translation (J) and Energy production and conversion (C). When *Pseudomonas* sp. M1 cells are challenged with β-myrcene, the cells must reconfigure their catabolic pathways to metabolize this hydrophobic compound. This shift needs a reconfiguration from the metabolic pathways typically used for hydrophilic substrates like lactate. Consequently, variations in proteins associated with energy production and conversion are expected. Interestingly, several proteins belonging to COG J were identified as RNA methyltransferases across all strains. This observation suggests a potentially important role for RNA modifications in the cellular response to β-myrcene. In Figure 27B (FC < -1.5), a distinct pattern emerges with COGs E, J, and M standing out. These COGs predominantly refer to proteins associated with Translation (J), Amino acid transport and metabolism (E), and Cell wall/membrane/envelope biogenesis (M). While the proteins of COG E likely play a crucial role in enabling the metabolic transition required for β-myrcene catabolism, and certain proteins within COG J may also contribute to this shift, the proteins within COG M are particularly interesting since these proteins are involved in processes such as motility and the transition between planktonic and biofilm states of the cells. As discussed earlier, the mutation or deletion of the *fleQ* gene has been shown to reduce the motility of *Pseudomonas* sp. M1 cells while upregulating the expression of genes essential for β-myrcene catabolism. This alteration is associated with an enhanced efficiency in metabolizing β-myrcene. Therefore, changes in the abundance of proteins categorized under COG M was an anticipated outcome.

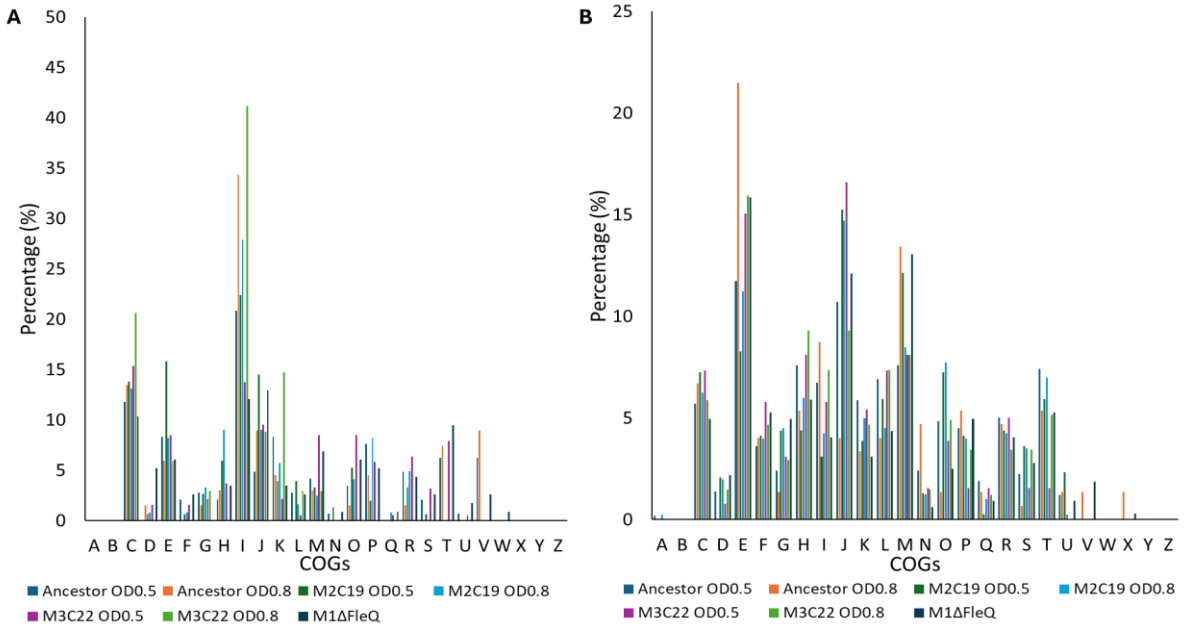


Figure 27 - Percentage of COGs function for proteins identified as at least 5 times more abundant after growth with β-myrcene in comparison with lactate (Fold-change > 1.5) – A – and for proteins identified as at least 5 times less abundant after growth with β-myrcene in comparison with lactate (Fold-change < -1.5) – B. Each letter stands for A - RNA processing and modification; B - Chromatin structure and dynamics; C - Energy production and conversion; D - Cell cycle control, cell division, chromosome partitioning; E - Amino acid transport and metabolism; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme transport and metabolism; I - Lipid transport and metabolism; J - translation, ribosomal structure and biogenesis; K – Transcription; L - Replication, recombination and repair; M - Cell wall/membrane/envelope biogenesis; N - Cell motility; O – Posttranslational modifications, protein turnover, chaperones; P - Inorganic ion transport and metabolism; Q - Secondary metabolites biosynthesis, transport and catabolism; R - General function prediction only; S - Function unknown; T - Signal transduction mechanisms; U - Intracellular trafficking, secretion, and vesicular transport; V - Defence mechanisms; W – Extracellular structures; X - Mobilome: prophages, transposons; Y - Nuclear structure; Z – Cytoskeleton.

To focus on specific proteins and get into more detail about the processes affected by β-myrcene catabolism, the 10 most abundant and the 10 less abundant proteins (FC>1.5 and FC<-1.5) for β-myrcene growth were retrieved, for both growth stages (OD_{600nm} 0.5 and 0.8) in the ancestor strain (Table 21). Notably, from the 10 proteins retrieved with FC>1.5, five proteins belong to the cluster necessary for β-myrcene catabolism in an OD_{600nm} 0.5, and 9 proteins belong to the same cluster in OD_{600nm} 0.8. The 6 proteins with increased abundance in β-myrcene and that do not belong to that specific cluster were the proteins WP_009622193.1, WP_009618817.1, WP_009623120.1, WP_009615985.1, WP_024128459.1 and WP_024128761.1. Watching closely the annotation of these proteins, the processes affected by its increased abundance are involved with the central metabolism, including the 2-methylcitrate cycle and the fatty-acids metabolism.

Table 21 - Top 10 proteins with at least 5 times more abundance in β-myrcene in comparison to lactate (FC > 1.5) or 5 times less abundance in in β-myrcene in comparison to lactate (FC < -1.5) for OD_{600nm} 0.5 and 0.8, and respective statistical significance, for the ancestor strain. In bold are the proteins belonging to the β-myrcene GI. * - p-value 0.05; ** - p-value between 0.05 and 0.1; - statistically non-significative.

	FC>1.5					FC<-1.5				
	Protein code	**Annotation**	**FC**	**COG**	**Significance**	**Protein code**	**Annotation**	**FC**	**COG**	**Significance**
OD600nm 0.5	WP_009622193.1	acyl-CoA synthetase	4.6	I	**	WP_009615600.1	ABC transporter permease	-4.8	E	**
	WP_009616263.1	LuxR C-terminal-related transcriptional regulator	4.4	-	**	WP_009615112.1	extracellular solute-binding protein	-4.6	G H	**
	WP_009618817.1	3-hydroxyacyl-CoA dehydrogenase NAD-binding domain-containing protein	4.4	I	**	WP_009615598.1	ABC transporter ATP-binding protein	-4.5	E	**
	WP_009623120.1	efflux RND transporter periplasmic adaptor subunit	4.3	V M	**	WP_009617445.1	tetratricopeptide repeat protein	-4.1	R	**
	WP_052058997.1	enoyl-CoA hydratase-related protein	4.3	I	**	WP_009613936.1	ribose-5-phosphate isomerase RpiA	-3.8	G	**
	WP_009616251.1	acetyl-CoA C-acetyltransferase	3.4	I	**	WP_009613589.1	bifunctional [glutamate–ammonia ligase]-adenylyl-L-tyrosine phosphorylase	-3.6	O	**
	WP_009615985.1	2-methylaconitate cis-trans isomerase PrpF	3.3	C	**	WP_009619418.1	HU family DNA-binding protein	-3.6	L	**
	WP_009616244.1	acyl-CoA dehydrogenase family protein	3.3	I	**	WP_009614957.1	two-component system response regulator OmpR	-3.4	K T	**
	WP_024128459.1	Fe/S-dependent 2-methylisocitrate dehydratase AcnD	3.2	C	**	WP_009613750.1	histidine utilization repressor	-3.4	-	**
	WP_009616271.1	FAD/NAD(P)-binding oxidoreductase	3.2	I	**	WP_009616649.1	RIP metalloprotease RseP	-3.2	K O	**
OD600nm 0.8	**WP_009616244.1**	acyl-CoA dehydrogenase family protein	8.7	I	*	WP_009615600.1	ABC transporter permease	-6.6	E	**
	WP_009616255.1	3-hydroxyacyl-CoA dehydrogenase NAD-binding domain-containing protein	7.6	I	**	WP_009615598.1	ABC transporter ATP-binding protein	-5.4	E	**
	WP_009616273.1	NAD(P)-dependent alcohol dehydrogenase	7.5	-	*	WP_009615599.1	ABC transporter substrate-binding protein	-4.4	E	**
	WP_024128489.1	acyl-CoA dehydrogenase family protein	7.4	I	**	WP_009613083.1	FAD-binding and (Fe-S)-binding domain-containing protein	-3.3	C	**
	WP_009616272.1	acyl-CoA synthetase	7.3	I	-	WP_009620864.1	flagellar filament capping protein FliD	-3.2	-	**
	WP_009616257.1	3-hydroxyacyl-CoA dehydrogenase	7.2	I	*	WP_009620637.1	high-affinity branched-chain amino acid ABC transporter ATP-binding protein LivG	-3.0	E	**
	WP_009616251.1	acetyl-CoA C-acetyltransferase	7.0	I	**	WP_009618604.1	cation-transporting P-type ATPase	-2.9	P	**
	WP_009616271.1	FAD/NAD(P)-binding oxidoreductase	6.6	I	**	WP_009614645.1	ABC transporter substrate-binding protein	-2.9	E	-
	WP_009616253.1	nitronate monooxygenase family protein	6.6	R	**	WP_009613081.1	lactate permease LctP family transporter	-2.8	C	**
	WP_024128761.1	3-hydroxyacyl-CoA dehydrogenase	6.2	I	**	WP_009616581.1	low molecular weight phosphotyrosine protein phosphatase	-2.6	T	-

For the proteins with decreased abundance in β-myrcene (FC<-1.5), most part of the proteins belong to COG E (Amino acid transport and metabolism). The annotation revealed that 7 proteins belong to the ABC transport family and the other 4 proteins are solute-binding proteins or permeases. In fact, the protein WP_009613081.1 is annotated as a "lactate permease LctP family transporter". Thus, the proteins with decreased abundance in β-myrcene, may be related with some processes not necessary during β-myrcene catabolism, including the transport of lactate into the cell. Following the same analysis for the M1-derived isolates adapted to β-myrcene, the isolate M2C19 had 7 proteins expressed by the β-myrcene genomic *locus* with increased expression at OD_{600nm} 0.5, comprising proteins belonging to TU1, 3 and 6 (Table 22), whilst the isolate M3C22 only had 5 proteins belonging to the referred *locus* with increased expression at OD_{600nm} 0.5, comprising proteins belonging to TU1 and 3 (Table 23). This data shows the different abilities of these isolates to catabolize β-myrcene. In fact, the isolate M2C19, indicated in chapter III as the one with higher specific growth rate using β-myrcene as sole carbon source, is also the one that has a higher number of proteins expressed by the β-myrcene genomic *locus* identified early in growth (OD_{600nm} 0.5) and with a higher relative abundance when compared with its lactate counterparts.

Among the proteins with increased abundance in M2C19 isolate, the protein WP_009615184.1 annotated as coniferyl aldehyde dehydrogenase was identified. In *Pseudomonas* sp. strain HR199 (DSM7063), this protein is induced during growth with eugenol[227]. Since this protein does not belong to the genomic *locus* necessary for the catabolism of β-myrcene, this might be a redundant process used to catabolize terpenoids. Looking at the protein annotations, all the other proteins that do not belong to the β-myrcene genomic *locus* and with an FC>1.5, are involved with central metabolism processes, including the 2-methylcitrate cycle and the fatty-acids metabolism, just like the ancestor strain. Focusing on the proteins with substantial decreased abundance in β-myrcene, for M2C19 isolate, several proteins are annotated as ABC transport family and solute-binding proteins or permeases. However, in this isolate, any protein related specifically with lactate was identified. Interestingly, the Rsd/AlgQ family anti-sigma factor was also found with a very decreased abundance. This anti-sigma factor has the capacity to inhibit the house keeping sigma 70 factor, which usually carries out the major transcriptional activity in *E. coli*[228,229]. Focusing on the isolate M3C22 (Table 23), besides the 5 proteins belonging to the β-myrcene catabolism already referred, the proteins with FC>1.5 are involved in central metabolism processes, much like what was observed for the ancestor strain.

Table 22 - Top 10 proteins with at least 5 times more abundance in β-myrcene (FC > 1.5) or 5 times less abundant (FC < -1.5) in β-myrcene in comparison to lactate for OD_{600nm} 0.5 and 0.8 for the isolate M2C19, and respective statistical significance. In bold are the proteins belonging to the β-myrcene GI. * - p-value 0.05; ** - p-value between 0.05 and 0.1; - statistically non-significative.

	FC>1.5					FC<-1.5				
OD600nm	**Protein code**	**FC**	**Annotation**	**COG**	**Significance**	**Protein code**	**Annotation**	**FC**	**COG**	**Significance**
0.5	**WP_024128489.1**	6.6	acyl-CoA dehydrogenase family protein	I	**	WP_009622838.1	dihydrolipoamide acetyltransferase family protein	-5.3	C	**
	WP_052058996.1	6.4	aldehyde dehydrogenase family protein	I	**	WP_009615600.1	ABC transporter permease	-4.8	E	**
	WP_009616244.1	5.8	acyl-CoA dehydrogenase family protein	I	-	WP_009622045.1	TonB-dependent siderophore receptor	-4.5	-	**
	WP_009616273.1	5.2	NAD(P)-dependent alcohol dehydrogenase	-	**	WP_009617996.1	inorganic phosphate transporter	-4.4	P	**
	WP_009616271.1	5.2	FAD/NAD(P)-binding oxidoreductase	I	**	WP_009614847.1	YbaL family putative K(+) efflux transporter	-4.2	P	**
	WP_009617873.1	5.1	acyl-CoA dehydrogenase family protein	I	**	WP_009617200.1	precorrin-8X methylmutase	-4.2	H	-
	WP_009616272.1	5.0	acyl-CoA synthetase	I	**	WP_009613929.1	HAD-IB family hydrolase	-4.2	E	**
	WP_009622919.1	4.9	aldehyde dehydrogenase	I	**	WP_009622335.1	I78 family peptidase inhibitor	-4.2	-	-
	WP_009618817.1	4.8	3-hydroxyacyl-CoA dehydrogenase NAD-binding domain-containing protein	I	**	WP_009612950.1	lipoprotein	-4.2	-	**
	WP_009616257.1	4.7	3-hydroxyacyl-CoA dehydrogenase	I	**	WP_009613516.1	ABC-type transport auxiliary lipoprotein family protein	-4.1	-	**
OD600nm	**WP_009616244.1**	9.1	acyl-CoA dehydrogenase family protein	I	**	WP_009615600.1	ABC transporter permease	-6.1	-	**
0.8	**WP_009616257.1**	8.3	3-hydroxyacyl-CoA dehydrogenase	I	**	WP_009615393.1	methyl-accepting chemotaxis protein	-4.3	T	**
	WP_009616251.1	8.3	acetyl-CoA C-acetyltransferase	I	**	WP_009615112.1	extracellular solute-binding protein	-4.3	G H	**
	WP_009616253.1	6.1	nitronate monooxygenase family protein	R	**	WP_009617902.1	16S rRNA (cytidine(1402)-2'-O)-methyltransferase	-4.2	J	-
	WP_009622919.1	5.7	aldehyde dehydrogenase	I	**	WP_009620914.1	phosphogluconate dehydratase	-4.2	G E	**
	WP_009622187.1	5.5	acyl-CoA dehydrogenase family protein	I	**	WP_009621155.1	MoxR family ATPase	-4.2	-	**
	WP_009615184.1	5.1	coniferyl aldehyde dehydrogenase	I	**	WP_009616508.1	tellurite resistance TerB family protein	-4.2	-	**
	WP_009622189.1	4.6	acetyl-CoA C-acyltransferase	I	**	WP_009616172.1	type IV pilus biogenesis/stability protein PilW	-4.1	-	**
	WP_009615985.1	4.1	2-methylaconitate cis-trans isomerase PrpF	C	**	WP_009616015.1	efflux RND transporter periplasmic adaptor subunit	-3.9	V M	**
	WP_009613109.1	3.9	OprD family porin	-	**	WP_009615050.1	Rsd/AlgQ family anti-sigma factor	-3.9	-	**

Table 23 - Top 10 proteins with at least 5 times more abundance in β-myrcene (FC > 1.5) or 5 times less abundant (FC < -1.5) in β-myrcene in comparison to lactate for OD_{600nm} 0.5 and 0.8 for the isolate M3C22, and respective statistical significance. In bold are the proteins belonging to the β-myrcene GI. * - p-value 0.05; ** - p-value between 0.05 and 0.1; - statistically non-significative.

	FC>1.5					FC<-1.5				
	Protein code	**Annotation**	**FC**	**COG**	**Significance**	**Protein code**	**Annotation**	**FC**	**COG**	**Significance**
OD600nm 0.5	**WP_009616255.1**	3-hydroxyacyl-CoA dehydrogenase NAD-binding domain-containing protein	9.2	I	**	WP_009615432.1	50S ribosomal protein L31	-4.2	J	**
	WP_009616253.1	nitronate monooxygenase family protein	7.3	R	**	WP_009620638.1	ATP-binding cassette domain-containing protein	-4.0	E	**
	WP_009616251.1	acetyl-CoA C-acetyltransferase	6.7	I	**	WP_009618376.1	nucleotide 5'-monophosphate nucleosidase PpnN	-3.8	F	**
	WP_009617873.1	acyl-CoA dehydrogenase family protein	6.2	I	**	WP_002553999.1	translation initiation factor IF-1	-3.5	J	**
	WP_009615576.1	PrkA family serine protein kinase	5.9	T	**	WP_009617419.1	carbonate dehydratase	-3.3	P	**
	WP_009616489.1	NADH:flavin oxidoreductase/NADH oxidase	5.8	C	**	WP_009613081.1	lactate permease LctP family transporter	-3.3	C	**
	WP_009616257.1	3-hydroxyacyl-CoA dehydrogenase	5.5	I	**	WP_009617170.1	30S ribosomal protein S8	-3.3	J	**
	WP_009615155.1	peptidoglycan-binding protein LysM	5.4	P	**	WP_009615598.1	ABC transporter ATP-binding protein	-3.3	P	**
	WP_009616246.1	acetyl-CoA C-acyltransferase	5.3	I	**	WP_009619439.1	YbaB/EbfC family nucleoid-associated protein	-3.3	K	**
	WP_009622811.1	aconitate hydratase AcnA	5.3	C	**	WP_009615071.1	4-hydroxy-3-polyprenylbenzoate decarboxylase	-3.2	H	-
OD600nm 0.8	**WP_009616257.1**	3-hydroxyacyl-CoA dehydrogenase	8.5	I	*	WP_009615112.1	extracellular solute-binding protein	-5.4	G H	**
	WP_009616253.1	nitronate monooxygenase family protein	8.4	R	**	WP_009619752.1	NADP-dependent succinate-semialdehyde dehydrogenase	-5.2	I	**
	WP_009617873.1	acyl-CoA dehydrogenase family protein	7.3	I	**	WP_009613054.1	23S rRNA (uridine(2552)-2'-O)-methyltransferase RlmE	-4.8	J	**
	WP_009622926.1	quinoprotein ethanol dehydrogenase	6.3	-	-	WP_232244603.1	polyphosphate kinase 2	-4.8	C	**
	WP_009616246.1	acetyl-CoA C-acyltransferase	6.3	I	**	WP_009615598.1	ABC transporter ATP-binding protein	-4.7	P	**
	WP_009618817.1	3-hydroxyacyl-CoA dehydrogenase NAD-binding domain-containing protein	6.2	I	**	WP_009615600.1	ABC transporter permease	-4.6	E	**
	WP_009622189.1	acetyl-CoA C-acyltransferase	5.5	I	**	WP_009615879.1	superoxide dismutase [Fe]	-4.6	P	-
	WP_024128459.1	Fe/S-dependent 2-methylisocitrate dehydratase AcnD	4.9	C	**	WP_009619439.1	YbaB/EbfC family nucleoid-associated protein	-4.5	K	-
	WP_009615987.1	2-methylcitrate synthase	4.8	C	-	WP_009621406.1	3-deoxy-manno-octulosonate cytidylyltransferase	-4.4	M	-
	WP_009616445.1	TonB-dependent receptor	4.5	P	-	WP_202883186.1	TIGR04255 family protein	-4.4	-	**

Among the proteins exhibiting substantially decreased fold-change in the isolate M3C22 (FC<-1.5), three proteins were identified as belonging to the ABC transport family and 2 proteins were identified as solute-binding proteins or permease. This included the protein WP_009613081.1 – "lactate permease LctP family transporter" – which was also found with FC<-1.5 in the ancestor strain. Other proteins identified in this category were WP_009621406.1 – "3-deoxy-manno-octulosonate cytidylyltransferase" – an enzyme involved in the incorporation of KDO (2-keto-3-deoxy-manno-octonic acid) into bacterial lipopolysaccharide in Gram-negative bacteria[230]. Moreover, 4 proteins related with translation processes, including ribosomal proteins and one translational initiation factor also had their fold-change decreased in β-myrcene for this isolate.

4.3.3. Overall fold change of protein content between evolved isolates and ancestor strain

Besides the comparison of proteins abundance between β-myrcene grown cells and Lactate, the importance of this work relies in the comparison between ancestor strain and evolved strains (M2C19 and M3C22). A difference in the identified proteins between the three strains was not observed, but differences in the abundance of several proteins were found. A table including the Fold-change values for all proteins identified, between the evolved isolates M2C19 and M3C22, the mutant M1ΔFleQ and the ancestor strain for both growth conditions, and respective adjusted p-value is depicted in supplementary material, Table S5. The number of proteins with considerable Fold-change differences in relation to the ancestor strain are depicted in Table 24.

Table 24 - Number of proteins at least 5 times more abundant (FC > 1.5) and 5 times less abundant (FC < -1.5) in the evolved isolates (M2C19 and M3C22) in comparison with the ancestor strain.

Condition	M2C19/Ancestor	M3C22/Ancestor
Fold change higher than 1.5		
β-myrcene OD_{600nm} 0.5	418	624
β-myrcene OD_{600nm} 0.8	373	260
Lactate OD_{600nm} 0.5	524	441
Lactate OD_{600nm} 0.8	798	545
Fold change lower than -1.5		
β-myrcene OD_{600nm} 0.5	339	282
β-myrcene OD_{600nm} 0.8	471	947
Lactate OD_{600nm} 0.5	388	386
Lactate OD_{600nm} 0.8	308	322

Firstly, the overall COGs for proteins with a fold-change higher than 1.5 (higher abundance of proteins in the evolved strains in comparison to the ancestor) a fold-change lower than -1.5 (lower abundance of proteins in the evolved strains in comparison to the ancestor) for β-myrcene grown-cells represented in Figure 28.

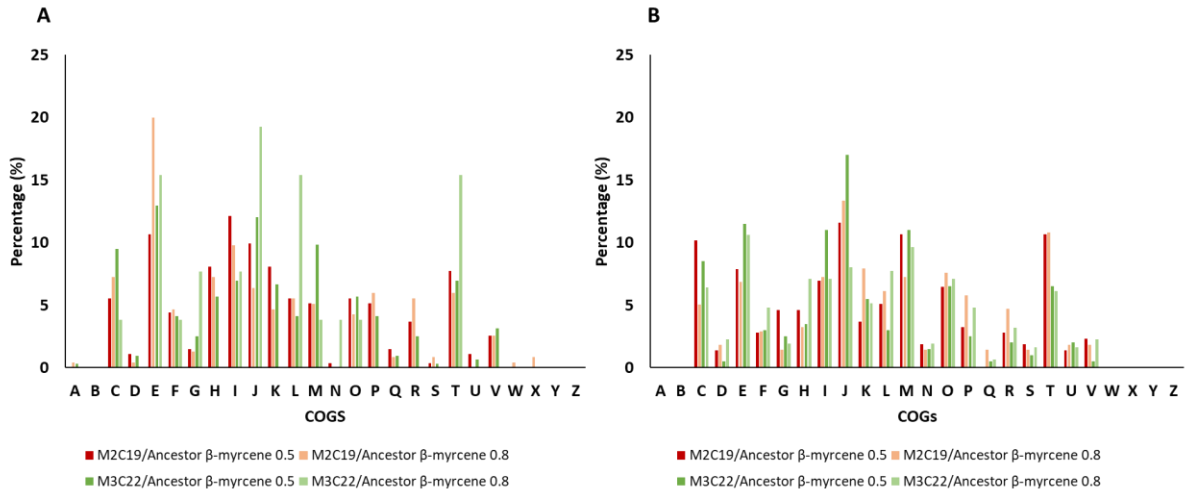


Figure 28 - Percentage of COGs function Number of proteins at least 5 times more abundant (FC > 1.5) – A – and 5 times less abundant (FC < -1.5) – B – in the evolved isolates (M2C19 and M3C22) in comparison with the ancestor strain. after growth with β-myrcene.–A - RNA processing and modification;–B - Chromatin structure and dynamics;–C - Energy production and conversion;–D - Cell cycle control, cell division, chromosome partitioning;–E - Amino acid transport and metabolism;–F - Nucleotide transport and metabolism;–G - Carbohydrate transport and metabolism;–H - Coenzyme transport and metabolism;–I - Lipid transport and metabolism;–J - translation, ribosomal structure and biogenesis; K – Transcription;–L - Replication, recombination and repair;–M - Cell wall/membrane/envelope biogenesis;–N - Cell motility;–O - Posttranslational modifications, protein turnover, chaperones;–P - Inorganic ion transport and metabolism;–Q - Secondary metabolites biosynthesis, transport and catabolism;–R - General function prediction only;–S - Function unknown;–T - Signal transduction mechanisms;–U - Intracellular trafficking, secretion, and vesicular transport;–V - Defence mechanisms;–W - Extracellular structures;–X - Mobilome: prophages, transposons;–Y - Nuclear structure; Z – Cytoskeleton.

Starting by analyzing the proteins with FC>1.5 in M2C19 isolate, when compared with the ancestor strain, the COGs I and J are the ones most well represented. As explained previously, these COGs comprise proteins related to lipid transport and metabolism (I) and Amino acid transport and metabolism (E). The proteins belonging to COG I and E may be related with the changes that the cells must undergo to metabolize β-myrcene. COG I includes the proteins specific for β-myrcene catabolism, thus, an increased fold-change of these proteins is expected in the evolved isolates. The same is true for the isolate M3C22, however for this isolate, a substantial increase in proteins belonging to the COGs L and T in the OD_{600nm} 0.8 is also observed. These COGs stand for– L - Replication, recombination and repair and– T - Signal transduction mechanisms. Both COGs can be related with stress management and

survival mechanisms necessary during stationary phase growth. In fact, in chapter III, an increased cellular mass was achieved by M3C22 isolate, in comparison to the ancestor strain and M2C19 isolate during stationary phase growth. For the proteins with decreased abundance in the isolates, when compared to the ancestor strain, most proteins belong to COGs J and M for both isolates. COG J stands for translation, ribosomal structure and biogenesis and COG M to Cell wall/membrane/envelope biogenesis.

Since motility/biofilm formation capacity seems to be the source of the increased efficiency towards β-myrcene catabolism in the evolved isolates, all the proteins grouped in the COGs M and N with substantial increased or decreased abundance in comparison to the ancestor were retrieved for both isolates and are depicted in supplementary material, Table S6 and Table S7. In both isolates, a decreased abundance of proteins WP_009621630.1 and WP_009620862.1, annotated as flagellin and flagellar motor switch protein FliN respectively, was observed, which is in accordance with the results observed until now. A decrease in the abundance of such proteins may lead to decreased motility. Moreover, in M2C19 isolate, some of the proteins with decreased abundance are directly related with peptidoglycan biosynthesis (WP_009618983.1 – D-alanine:D-alanine ligase) and its elongation (WP_052059035.1 – peptidoglycan D,D-metalloendopeptidase family protein) whereas in M3C22 isolate has decreased abundance of proteins involved with the synthesis of N-acetylmuramic acid (component of peptidoglycan). Moreover, several proteins related with polysaccharide biosynthesis, and other components of the cell envelope are also decreased in both isolates. However, also for both isolates, enzymes involved with synthesis of peptidoglycan showed an increased abundance in comparison with the ancestor strain, namely the proteins WP_009615975.1 – "UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase" and WP_009617909.1 – "UDP-N-acetylmuramoyl-tripeptide–D-alanyl-D-alanine ligase" for the isolate M2C19 and the proteins WP_009617318.1 – "lytic murein transglycosylase", WP_009615975.1 – "UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase" and WP_009621444.1 – "endolytic transglycosylase MltG" for the isolate M3C22. This behaviour was already described in the literature in *Pseudomonas aeruginosa* epidemic strains isolated from cystic fibrosis patients. In these strains, changes in the peptidoglycan composition were observed when the cystic fibrosis environment was mimicked. These changes included increased O-methylated stem peptide, decreased L,D-endopeptidase activity and increased abundance of de-N-acetylated sugars and L,D-transpeptidase activity. These changes seemed to lead to higher ability to grow biofilms, higher virulence, and antibiotic resistance[231,232].

4.3.4. The isolates M2C19 and M3C22, and the mutant M1ΔFleQ have major differences in abundance of proteins regulated by the FleQ regulator

The transcriptional regulator FleQ has been studied thoroughly in *Pseudomonas* for the past years. In *Pseudomonas aeruginosa*, *Pseudomonas fluorescens* and *Pseudomonas putida*, FleQ positively regulates the expression of genes involved with flagellar export (*flhA* and *fliLMNOPQ* operon), localization and regulation of the flagellar apparatus (*flhF* and *fleN*) and structural components of the flagellar basal body and motor switch complex (*fliEFG*). All these genes are also present and annotated in *Pseudomonas* sp. M1 genome. Upon extraction and identification of the total protein content of M1 cells ancestor, adapted isolates and M1ΔFleQ mutant, all the proteins corresponding to the expression of such genes were identified in all the variants. Moreover, the adapted isolates and M1ΔFleQ mutant showed differences in the relative abundance of the proteins FliL, FliN, FliO, FleN and FliF when compared with the ancestor strain either when grown with β-myrcene or with Lactate as carbon sources (Table 25). Moreover, the fold-change of the FleQ regulator is also depicted in Table 25, showing its decreased relative abundance in all isolates/mutant in comparison with the ancestor strain. However, despite the relative abundance of FleQ in the M2C19 isolate in comparison with the ancestor strain, its levels are still higher than the isolate M3C22 and the mutant M1ΔFleQ.

Table 25 - Fold-change of proteins which expression is reported as to be affected by the FleQ transcriptional regulator, between the adapted isolates/constructed mutant and the ancestor strain and respective statistical significance. The results shown are for the cells grown in β-myrcene, as only carbon source. * - p-value 0.05; ** - p-value between 0.05 and 0.1; - statistically non-significative.

Protein Accessions	Gene	FC M2C19/Ancestor	Significance	FC M3C22/Ancestor	Significance	FC M1ΔFleQ/Ancestor	Significance
WP_024128246.1	*fleQ*	-1.05	**	-2.18	**	-1.62	**
WP_009621628.1	*fliL*	**1.67**	**	-1.18	-	-0.62	-
WP_009621630.1	*fliN*	**1.02**	**	-1.16	**	0.18	-
WP_009621631.1	*fliO*	-0.33	-	0.82	-	0.00	**
WP_009621094.1	*fleN*	-1.05	**	-2.75	**	0.00	-
WP_009620872.1	*fliF*	0.00	-	-2.18	-	-1.62	-

Since FleQ is known as a positive regulator, its mutation that led to a decreased motility in both adapted isolates, would impact the abundance of the proteins dependent on this regulator negatively, leading to a decrease in relative abundance when compared with the ancestor strain. Moreover, since the gene coding for this regulator was deleted in the mutant constructed, the abundance of proteins dependent on this regulator would also have its abundance decreased in comparison to the *wt* strain. Despite this

assumption be relatively true for the M1ΔFleQ mutant and the isolate M3C22, when grown in β-myrcene, the same does not apply to the isolate M2C19. In this isolate, the proteins FliL and FliN showed an increased abundance in the M2C19 isolate, in comparison with the ancestor strain. As referred previously in chapter III, the mutations accumulated in the *fleQ* gene for both isolates were not the same, nor in the same location. Moreover, mutations in two different histidine kinases were reported, both seemingly related with motility and the switch between planktonic/biofilm lifestyle. Thus, the adapted isolates may not lose its motility ability in the presence of β-myrcene, but rather develop a mechanism that allows only a decrease in motility and biofilm formation when a hydrophobic particle of β-myrcene is found, allowing its efficient catabolism, and return to its usual motility capacity (or even higher motility) when in search of β-myrcene particles, based on the signal captured by the histidine kinases and the further levels of FleQ regulator in the cells and its interaction with the targets.

4.3.5. Proteome assessment of M2C19, M3C22 isolates and M1ΔFleQ showed increased expression of the 28-kb genomic *locus* proteins in comparison to the ancestor

The main purpose of this work was to adapt the strain *Pseudomonas* sp. M1 to increase its efficiency to catabolize monoterpenes, namely β-myrcene. The proteins responsible for this pathway were already known, as well as its coding genes which are clustered in a 28-kb genomic *locus*. Moreover, eight transcriptional units (TU) were also predicted for the expression of such genes. The increased efficiency to catabolize β-myrcene, will be dependent on the increased abundance of proteins necessary for its conversion. Thus, the evolved strains M2C19 and M3C22 should have a higher abundance of these proteins, when compared to the ancestor strain. Moreover, in chapter III, a mutant lacking the gene *fleQ* was constructed. This gene was the only one with detected mutations in all the independently evolving β-myrcene cultures, being also detected in both evolved isolates. Thus, the abundance of all the proteins coded by the 28-kb genomic *locus* in the evolved isolates and in the mutant M1ΔFleQ was compared with the abundance of the same proteins in the ancestor strain (FC). All the proteins with any increase in abundance in the isolates or mutant when compared with the ancestor strain were highlighted. These proteins are represented in Table 26 for the OD_{600nm} 0.5 and in Table 27 for the OD_{600nm} 0.8. For the exponential growth phase (OD_{600nm} 0.5) the isolate M2C19 had the highest number of proteins necessary for β-myrcene catabolism with a higher abundance in comparison with the ancestor strain. From the 13 proteins with higher fold-change, 7 proteins were statistically significative. The isolate M3C22 had only two proteins with higher abundance, both statistically significative. Moreover, the mutant M1ΔFleQ did not have any proteins related with β-myrcene catabolism, with higher abundance than the ancestor

strain. This is in accordance with the results obtained in chapter III where the mutant did not have higher activity of P5 promoter when compared with the ancestor strain during the exponential growth phase (OD_{600nm} 0.5). For the later growth phase (OD_{600nm} 0.8), the results are a bit different.

Table 26 - Essential proteins to catabolize β-myrcene with higher abundance in the isolates (M2C19, M3C22) and mutant M1ΔFleQ in comparison with the ancestor strain (Fold-change > 1) for OD_{600nm} 0.5. The mutant M1ΔFleQ did not have any protein identified with higher abundance than the ancestor strain. * - p-value 0.05; ** - p-value between 0.05 and 0.1; - statistically non-significative.

TU from β-myrcene genomic island	Protein code	Annotation	Statistical significance
		M2C19	
TU1	WP_024128489.1	acyl-CoA dehydrogenase family protein	**
TU1	WP_009616244.1	acyl-CoA dehydrogenase family protein	-
TU2	WP_009616247.1	methyl-accepting chemotaxis protein	-
TU2	WP_009616249.1	outer membrane beta-barrel protein	*
TU3	WP_009616257.1	3-hydroxyacyl-CoA dehydrogenase	*
TU3	WP_009616259.1	epoxide hydrolase	-
TU5	WP_238603225.1	fatty acid desaturase	**
TU6	WP_009616269.1	LuxR C-terminal-related transcriptional regulator	*
TU6	WP_009616271.1	FAD/NAD(P)-binding oxidoreductase	**
TU6	WP_009616272.1	acyl-CoA synthetase	*
TU6	WP_009616273.1	NAD(P)-dependent alcohol dehydrogenase	-
TU6	WP_052058996.1	aldehyde dehydrogenase family protein	-
TU7	WP_052058997.1	enoyl-CoA hydratase-related protein	-
		M3C22	
TU5	WP_238603225.1	fatty acid desaturase	**
TU8	WP_009616276.1	LysR substrate-binding domain-containing protein	**

The isolate M2C19 had fifteen proteins with increased fold-change, however only two proteins were statistically significative. The isolate M3C22 had fourteen proteins with increased fold-change, a fairly increased number when compared with the exponential growth phase. Moreover, the mutant M1ΔFleQ had four proteins with increased fold-change. Interestingly, the protein MyrH is one of the proteins with increased fold-change. Thus, the production of proteins related to the β-myrcene catabolism seem to happen in different growth phases for the different isolates/mutant. In fact, the mutant M1ΔFleQ showed a slightly increase in efficiency when compared with the ancestor strain, but not higher than the evolved isolates, and only after 24h induction. The results presented in this chapter are in accordance with the previous chapter III, corroborating the importance of motility genes and planktonic/sessile states switch for the efficient catabolism of β-myrcene in *Pseudomonas* sp. M1.

Table 27 - Essential proteins to catabolize β-myrcene with higher abundance in the isolates (M2C19, M3C22) and mutant M1ΔFleQ in comparison with the ancestor strain (Fold-change > 1) for OD_{600nm} 0.8. * - p-value 0.05; ** - p-value between 0.05 and 0.1; - statistically non-significative.

TU from β-myrcene genomic island	Protein code	Annotation	Statistical significance	TU from β-myrcene genomic island	Protein code	Annotation	Statistical significance	TU from β-myrcene genomic island	Protein code	Annotation	Statistical significance
M2C19				**M3C22**				**M1ΔFleQ**			
TU1	WP_024128489.1	acyl-CoA dehydrogenase family protein	-	**TU1**	WP_024128489.1	acyl-CoA dehydrogenase family protein	-	**TU3**	WP_009616255.1	3-hydroxyacyl-CoA dehydrogenase NAD-binding domain-containing protein	-
TU1	WP_009616244.1	acyl-CoA dehydrogenase family protein	-	**TU1**	WP_009616244.1	acyl-CoA dehydrogenase family protein	-	**TU5**	WP_238603225.1	fatty acid desaturase	**
TU2	WP_009616249.1	outer membrane beta-barrel protein	-	**TU2**	WP_009616249.1	outer membrane beta-barrel protein	-	**TU6**	WP_009616271.1	FAD/NAD(P)-binding oxidoreductase	*
TU3	WP_009616255.1	3-hydroxyacyl-CoA dehydrogenase NAD-binding domain-containing protein	-	**TU3**	WP_009616251.1	acetyl-CoA C-acetyltransferase	-	**TU8**	WP_009616277.1	EexN family lipoprotein	-
TU3	WP_009616257.1	3-hydroxyacyl-CoA dehydrogenase	-	**TU3**	WP_009616253.1	nitronate monooxygenase family protein	-				
TU3	WP_009616259.1	epoxide hydrolase	-	**TU3**	WP_009616255.1	3-hydroxyacyl-CoA dehydrogenase NAD-binding domain-containing protein	-				
TU3	WP_009616261.1	chalcone isomerase family protein	-	**TU3**	WP_009616259.1	epoxide hydrolase	-				
TU4	WP_009616263.1	LuxR C-terminal-related transcriptional regulator	-	**TU4**	WP_009616263.1	LuxR C-terminal-related transcriptional regulator	-				
TU5	WP_238603225.1	fatty acid desaturase	**	**TU5**	WP_238603225.1	fatty acid desaturase	**				
TU6	WP_009616269.1	LuxR C-terminal-related transcriptional regulator	-	**TU6**	WP_009616269.1	LuxR C-terminal-related transcriptional regulator	-				
TU6	WP_009616271.1	FAD/NAD(P)-binding oxidoreductase	*	**TU6**	WP_009616271.1	FAD/NAD(P)-binding oxidoreductase	*				
TU6	WP_009616272.1	acyl-CoA synthetase	-	**TU6**	WP_009616272.1	acyl-CoA synthetase	-				
TU6	WP_009616273.1	NAD(P)-dependent alcohol dehydrogenase	-	**TU6**	WP_009616273.1	NAD(P)-dependent alcohol dehydrogenase	-				
TU6	WP_052058996.1	aldehyde dehydrogenase family protein	-	**TU6**	WP_052058996.1	aldehyde dehydrogenase family protein	-				
TU7	WP_052058997.1	enoyl-CoA hydratase-related protein	-								

In summary, the extraction of the total protein content of *Pseudomonas* sp. M1 *wt* and derived strains – M2C19, M3C22 and M1ΔFleQ mutant – allowed the further comprehension of the adaptation processes of this bacterium towards β-myrcene. Firstly, the fold-change (FC) of the proteins identified during the β-myrcene growth in comparison with lactate growth was calculated. As expected, proteins crucial for β-myrcene catabolism showed significantly increased abundance (FC>1.5) across all strains. Additionally, proteins involved in central metabolism, including the 2-methylcitrate cycle and fatty acid metabolism, were also abundant (FC>1.5) in all strains. Interestingly, M2C19 exhibited increased abundance of a protein annotated as coniferyl aldehyde dehydrogenase, suggesting a redundant process for terpenoid catabolism. Conversely, proteins with markedly decreased abundance (FC<-1.5) during β-myrcene growth were predominantly solute-binding proteins or permeases across all strains. In the subsequent analysis, the FC of proteins identified in the evolved isolates (M2C19 and M3C22) compared to the ancestral strain was calculated. Notably, several proteins related to cell motility and the transition between sessile and motile states showed substantial FC increased or decreased in both evolved isolates. Intriguingly, these proteins were previously identified in *Pseudomonas aeruginosa* epidemic strains isolated from cystic fibrosis patients, associated with enhanced biofilm growth, virulence, and antibiotic resistance. Given the apparent loss of function in the *fleQ* gene in the evolved isolates, the FC of the FleQ transcriptional regulator was accessed, showing a decreased abundance in both evolved isolates and the M1ΔFleQ mutant compared to the ancestral strain. However, proteins encoded by genes regulated by FleQ showed increased FC abundance compared to the ancestral strain, particularly FliL and FliN proteins, in the evolved isolate M2C19. Moreover, consistent with Chapter III results, proteins specific to β-myrcene catabolism showed increased abundance in the evolved isolates and the M1ΔFleQ mutant compared to the ancestral strain, although at different growth stages.

Chapter V – First insights on the methylome dynamics of *Pseudomonas* sp. M1

5.1. Introduction

Epigenetics is the study of changes in an organism phenotype caused by modification in gene expression associated to the methylation of the DNA. In bacteria, despite the methylation patterns can be passed on to daughter cells during cell division – epigenetic adaptation – these changes are reversible allowing a versatile and dynamic response of bacteria to environmental changes[233,234]. DNA methylation refers to the addition of methyl groups usually related with specific DNA motifs recognized by methyltransferases enzymes. This process can control gene expression by affecting how DNA-binding proteins interact with the DNA[235] or by changing the shape of the DNA itself[236]. DNA methylation can occur in intergenic regions (like promoters) or within genes themselves. The methylation of promoter regions or protein binding-sites modulate the affinity of transcriptional regulators and RNA polymerases, influencing the transcription of genes. On the other hand, when certain regions of the DNA are heavily methylated, the DNA becomes more tightly packed, making it harder for proteins to access it, also influencing the transcription of genes. Moreover, methylation of histone-like proteins can also alter how these proteins interact with DNA, which may change the DNA topology[237,238]. Methylation in coding regions may inhibit gene expression. On one hand, methylation inside the gene may block transcription elongation, changing the RNA polymerase affinity for the DNA or blocking the enzyme. On the other hand, methylation of genes may affect the DNA structure thus impacting transcription[234,239]. There are 2 types of methylation that may happen in the bacterial genome: DNA adenine methylation (Dam) which is the predominant methylation found in bacteria produced by Dam methyltransferases that may recognize GATC sequences; and DNA cytosine methylation (Dcm) which remains relatively understudied in bacteria, being the sequence C^{m}CWGG usually recognized by Dcm methyltransferases. In bacteria, methylation occurs predominantly at N6-methyladenine (6mA methylations by Dam methyltransferases), being the methylation C5-methylcytosine (5mC methylations by Dcm methyltransferases) less frequent. A third methylation type – N4-methylcytosine (4mC) can also be found in bacteria. Although found in 1983[253], the current knowledge gathered about this methylation is still scarce. The only functions known to be somewhat related with 4mC methylation are the control of DNA replication and protection of DNA against degradation, as part of the restriction modification system. The study of this methylation fell behind due to the lack of effective and economical techniques for its identification[254].

5.1.1. DNA adenine methylation (Dam)

Despite most methyltransferases of this group globally recognize the motif GATC, sometimes species or strain specific sites that are also recognized[240]. These enzymes methylate the N at position 6 of Adenine

purine. The heritable control of gene expression influenced by Dam methylation pattern was first shown in the pyelonephritis-associated pilus (*pap*) operon in uropathogenic *E. coli*. Thus, the daughter cell inherits the expression phase of the parent cell, but it can be reversible between generations. In 1990, Blyn and collaborators stated that depending on the pilus expression state of the cell, the *pap* operon was heritably nonmethylated[241]. When the cell was expressing the *pap* pili (ON-phase), the GATC site proximal to its promoter (located 53 bp from the *papBA* transcription start site) was methylated whilst the promoter distal GATC site (located 102 bp upstream of GATC proximal) was non-methylated. However, when the cell was not expressing the *pap* pili (OFF-phase), the proximal GATC site was not methylated and the distal GATC site was methylated. These GATC sites control the binding of the global regulator Lrp and the coregulatory protein PapI to *pap* operon promoter region. However, GATC sites are also recognized and methylated by Dam methyltransferases. Thus, when the parent cell is expressing the *pap* operon, PapI-Lrp binds to the *pap* DNA sites distal to *papBA*, protecting the distal GATC site from methylation, whilst when the parent cell is not expressing the operon, Lrp binds to the *pap* DNA sites proximal to *papBA*, protecting the proximal GATC site from methylation. Upon cell division, the state ON or OFF phase can be reverted. During cell division, the newly synthesized DNA strand is non-methylated, whilst the template strand has its own methylations. This DNA state where 1 strand is methylated and the other not, is called hemimethylated. In this new, non-methylated strand, the Lrp regulator, may transiently bind all its binding sites. If the protein PapI is being expressed, it will bind to the Lrp in the distal site and stabilize in this position. If not, the Lrp regulator will stabilize in the proximal position, reverting the methylation pattern inherited by the parent cell[241–243]. Moreover, the expression of PapI protein is environmentally controlled by carbon source via the cAMP level. In a similar process, methylation also controls the expression of other genes, either related with pili (e.g. *fim* regulation controlling type 1 pili)[244,245], flagellar synthesis (non-methylated sites in *flh* promoter) and other functions (iron transport, stress response, among others) [246,247].

5.1.2. DNA cytosine methylation (Dcm)

Cytosine modifications have been most studied in eukaryotic organisms rather than prokaryotic organisms. Although cytosine methylations happen in bacteria, these are not the most common modifications, being the Dam methylation most frequent. However, the role of Dcm methylation is as important for the cells. The sequence recognized by Dcm methyltransferases in C^{m}CWGG where W can be any amino acid, which is the same recognition sequence for the restriction enzyme EcoRII[248]. Thus, it is hypothesized that this methylation system arose to protect the DNA from invasion of restriction-

modification complexes (bacteriophages). These systems may be fatal for the cells upon double strand break. In 2002 Takashi and collaborators demonstrated that the methylation of CCWGG site increased cell viability when in the presence of the parasitic RM system containing the EcoRII gene complex[249]. However, the methylation of cytosine, may lead to its conversion to the DNA base thymine upon deamination, during replication[250]. To correct replicating errors, the mismatch repair system must be able to distinguish between the template strand and the newly synthesized strand. This distinction is made through the recognition of the methylated strand, since only the template will be methylated. Thus, when the methylated cytosine is deaminated and transforms into a thymine during replication, the mismatch repair system is able to recognize the newly synthesized strand and correct the error. Firstly, the enzyme MutS binds to the mismatched base pairs, MutL interacts physically with MutS, allowing the assembly of all the components of MMR system, including the enzyme MutH. MutH recognizes Dam hemimethylated DNA sequences and nicks the DNA in the mismatched position, allowing for its correction. In the absence of Dam methylation, MutH may cleave the daughter strand, the parental strand, or both DNA strands. If the cell survives double-strand DNA breakage, 50% of the time the mutant daughter strand is used as a template to "repair" the parental strand. Thus, methylation also is important for the DNA repair system in the cells[251,252].

5.1.4. Role of methylation in transposition

Transposable elements are accumulated overtime in the DNA of all organisms – prokaryotic and eukaryotic. These elements are able to move within the genome in a process called transposition. The activity of transposable elements influences the genome through gene inactivation, modulation of gene expression and illegitimate recombination[253,254]. A GATC site (recognized by Dam methyltransferases) is present in the -10 region of the transposase promoter pIN (which regulates the transcription of the transposase enzyme), and several insertion sequencies such as IS10 and IS50, contain a GATC site near their inner terminus. Fully methylation of either site will i) inhibit RNA polymerase binding and ii) inhibit transposase binding. In the case of transposition of transposons (e.g.Tn5), Dam methylation represses the *tnp* promoter activity, avoiding the transcription of genes within the transposon (all the machinery associated with the transposition event), impairing the transposition event[255–257]. Thus, methylation may have an important role in the regulation of some transposition events.

5.1.5. Methylation detection with Oxford-Nanopore technology

Advances in third generation long-read sequencing platforms enabled the direct detection of methylation without the need of pre-sequencing treatment. Most previous approaches as the traditional bisulphite sequencing (BS-seq) requires specialized treatment before sequencing. In this technology, unmethylated cytosine residues are deaminated and converted to uracil. However, the methylated cytosines remain as cytosines. After treatment, the DNA can be sequenced, for instance, using Illumina technology, and the methylated cytosines detected[258]. BS-seq can also detect other methylations, such as 6mA and 4mC, but less effectively[259]. The main throwback between third generation sequencing and technologies such as BS-seq is the high probability of DNA degradation during the DNA pre-treatment and the loss of information regarding the organization of methylation patterns in all the genome and in different genomic contexts[260]. In the context of this work, this is a very important advantage as it allows the comparison of changes in the methylation patterns of different growth conditions, thus providing a means for comparative epigenomics. Between the third generation sequencing tools – PacBio and Oxford-Nanopore – the second one allows the sequencing *in situ*, thanks to the easiness of library preparation and cost effectiveness for small scale experiments, whilst PacBio needs specialized personal and sequencers[17,18]. However, for Nanopore technology, the limiting step resides in the tools developed to analyze and detect such modifications, depending on the algorithm used and its training. As explained previously, Nanopore generates *Fast5* files containing ionic current data store, that differs from base to base. These files are afterwards base called using deep learning models contained for instance in *Guppy* and converted in *FastQ* files. However, the methylated bases have a different signal in comparison with non-methylated bases. Thus, after base calling, the reads are mapped against a reference genome and the electric signal is compared again to identify the bases with the alternative signal (modified bases). Several tools were designed to identify modified bases, including methylated bases, for instance *Megalodon*[19,261]. Thus, the Nanopore detection of modified bases is dependent on the correct base calling step for these bases.

In this chapter, Oxford-Nanopore technology was used to uncover the methylome dynamics of *Pseudomonas* sp. M1 *wt*. The relative abundance of the methyltransferases identified in *Pseudomonas* sp. M1 was accessed using the data acquired during Chapter IV of this thesis. In this chapter, the correlation between the relative abundance of methyltransferases and the different methylome profiles on different growth conditions was studied.

5.2. Material and Methods

5.2.1. High molecular weight genomic DNA extraction

The DNA of *Pseudomonas* sp. M1 *wt* was extracted in duplicate for each condition. The cells were grown overnight in 50 mL Mineral Media containing only β-myrcene or lactate as carbon source (2 cultures of each condition) with agitation at 30°C. In the morning, the cells were refreshed and grown until the early exponential phase (OD_{600nm} between 0.2 and 0.3) in the same conditions and media as previously. After growth, the cultures were centrifuged in falcon at 2500 *xg* for 5 minutes. The gDNA was extracted following the kit Wizard® HMW DNA Extraction Kit from Promega, using wide-bored tips in order to decrease DNA fragmentation. After extraction, the gDNA was purified using Mag-Bind® Total Pure NGS beads. The beads were firstly vortexed and added to the gDNA in the same volume as the volume of sample obtained and pipeted down with wide-bored tips 5 to 10 times. Afterwards the samples were placed on a magnetic plate and let it sit for 2 minutes before aspirating the supernatant. The beads were washed with 200 µL of fresh ethanol 70% three times. After washing, the traces of ethanol were dried in an incubator at 37°C for 30 minutes. The DNA was eluted from the beads using Tris-HCl 10 mM, pH 6.8. DNA integrity was accessed by electrophoresis gel 0.8%, run at 60 V and the DNA was quantified using the qubit.

5.2.2. gDNA sequencing with MinION from Oxford-Nanopore technology

The gDNA extracted was first treated with AMPure XP beads (magnetic beads). Briefly, the beads were vortexed, pipetted up and down 5 to 10 times and let sit for 10 minutes. The beads were added to the sample and incubated for 5 minutes at room temperature. The Eppendorf was then placed into a magnet plate for at least 2 minutes to allow the complete separation of the beads from the solution. The supernatant was aspirated and discarded without disturbing the beads and the beads were cleaned with 200 µL of 70 % (v/v) ethanol. The cleaning process was repeated, and the beads were air dried. The DNA was eluted with 40 µL of elution buffer, mixed well and incubated for 5 minutes at room temperature. The Eppendorfs were left in the magnetic separation until the beads were completely cleared from the solution. The cleared supernatant containing the purified DNA was then transferred to a clean tube. Purified DNA was used to prepare the library to be sequenced on MinION. To barcode the samples, one microgram of each DNA sample was mixed with water until 7.5 µL and 2.5 µL of Fragmentation Mix RB01-12 (one for each sample) was added. The samples were incubated at 30 °C for 1 minute and at 80 °C for 1 minute. The barcoded samples were pooled in equal quantities into a microtube, and 1 µL of a rapid adapter was added. The solution was incubated for 5 minutes at RT.

Afterwards, the barcoded samples were mixed in the same proportion in a new tube, 1 μL of RAP was added to the pooled barcoded samples and incubated for 5 minutes at room temperature. After incubation, 11 μL of the barcoded DNA pool was mixed with 34 μl Sequencing Buffer (SQB), 25.5 μl Loading Beads (LB) and 4.5 μl Nuclease-free water, being ready to be sequenced (library). The priming of the MinION flow cell was performed using the Flow Cell Priming Kit (EXP-FLP002). The flow priming was prepared by mixing 30 μL of Flush Tether (FLT) directly to the tube of the Flush Buffer (FLB) and mixing by vortexing at room temperature. Opening the MinION device, the priming port was checked for bubbles, according to the flow cell manual and 800 μL of the priming mix was inserted via the priming port and let incubate for 5 minutes. Afterwards, the remaining 200 μL of the mixture were added to the flow cell via SpotON port with the priming port open. Afterwards, 75 μL of library were added to the flow cell via SpotON, all the ports were closed, and the device was ready to sequence. The samples were sequenced for 12h.

5.2.3. Onward analysis of MinION data

The data obtained with MinION was firstly basecalled using *Guppy* [18]. This data processing toolkit contains basecalling algorithms of Oxford-Nanopore Technologies and several bioinformatic post-processing features. The reads were transformed into basecalls using the dna_r9.4.1_450bps_sup.cfg mathematical model. The sequences retrieved were filtered by quality using *FitLong*. This tool filters by read length and read identity. To assess the methylome profiling, Megalodon was used to perform a modified base calling where the methylated bases were identified. The trained models used were dna_r9.4.1_450bps_modbases_dam-dcm-cpg_hac.cfg to identify 6mA modified bases and the model dna_r9.4.1_450bps_modbases_5mc_hac.cfg to identify 5mC modified bases. The use of both models was necessary due to sensitivity to both modifications.

5.3. Results and discussion

In the previous chapters, different mutations and mutation rates were detected at 600 generations during the adaptation of *Pseudomonas* sp. M1 to the different carbon sources. When looking into the mutations detected in β-myrcene cultures evolved until 600 generations, mutations in the 28-kb genomic *locus* (specific for β-myrcene catabolism) were not detected. This led to a question: may the genomic island be protected in some way from mutational events? Could this cluster be hypermethylated, thus being protected from mutation, namely when cells grew in β-myrcene? Moreover, for the two isolates seemingly

more adapted to β-myrcene, only two mutations were detected in the genome of each isolate. Nonetheless, different proteomic profiles were revealed for each isolate, including different abundance of proteins belonging to the β-myrcene 28-kb genomic *locus* in the different growth phases. In fact, M2C19 isolate showed higher abundance of several proteins necessary for β-myrcene catabolism, earlier in growth when compared with the isolate M3C22 and the mutant M1ΔFleQ. Genome methylation is related to the regulation of gene expression, influencing transcription and translation, and interplaying with DNA repair mechanisms. Thus, to understand the implications of DNA methylation in M1 cells, namely when grown in β-myrcene, all the methyltransferases identified in the proteome of the ancestor strain after growth in lactate or β-myrcene were retrieved. Analyzing the annotation of *Pseudomonas* sp. M1 genome, three proteins were annotated as DNA methyltransferases – WP_009617102.1, WP_009618168.1 and WP_043299885.1. For the 3 DNA methyltransferases identified in the proteomics data, the Fold-change in abundance between cells grown in β-myrcene or lactate and in different OD_{600nm} were represented in Table 28.

Table 28 - DNA methyltransferases identified for *Pseudomonas* sp. M1 ancestor strain by mass spectrometry and respective fold-change based in the relative abundance of proteins in β-myrcene vs lactate growing cells.

OD_{600nm} 0.5	**FC β-myrcene/Lac**	**OD_{600nm} 0.8**	**FC β-myrcene/Lac**
WP_009618168.1	0.91	**WP_009618168.1**	0.60
WP_043299885.1	0.90	**WP_043299885.1**	0.61
WP_009617102.1	0.82	**WP_009617102.1**	1.45

According to Table 28, at an OD_{600nm} 0.5 all DNA methyltransferases are less abundant in β-myrcene than in lactate. At an OD_{600nm} 0.8, the methyltransferases WP_009618168.1 and WP_043299885.1 are still more abundant in lactate. However, the methyltransferase WP_009617102.1 is 1.45 times more abundant in β-myrcene than lactate. Taking into consideration the relative abundance of these proteins, the methyltransferases may be more abundant in β-myrcene compared to lactate, later in growth, which may influence the methylation patterns predicted for these bacteria.

To understand if, in fact, there were predicted differences in the methylome dynamics of *Pseudomonas* sp. M1 *wt* (ancestor strain) when grown in lactate or β-myrcene, the DNA of cells with the same OD_{600nm} was extracted and sequenced in Oxford-Nanopore MinION. The number of methylations in Cytosine (5mC) and Adenine (6mA) were provided showing that overall, the cells have its DNA more methylated after growth in β-myrcene than in lactate. Moreover, the overall number of predicted methylated sites and genes methylated for 6mA was higher than for 5mC for both conditions (Figure 29).

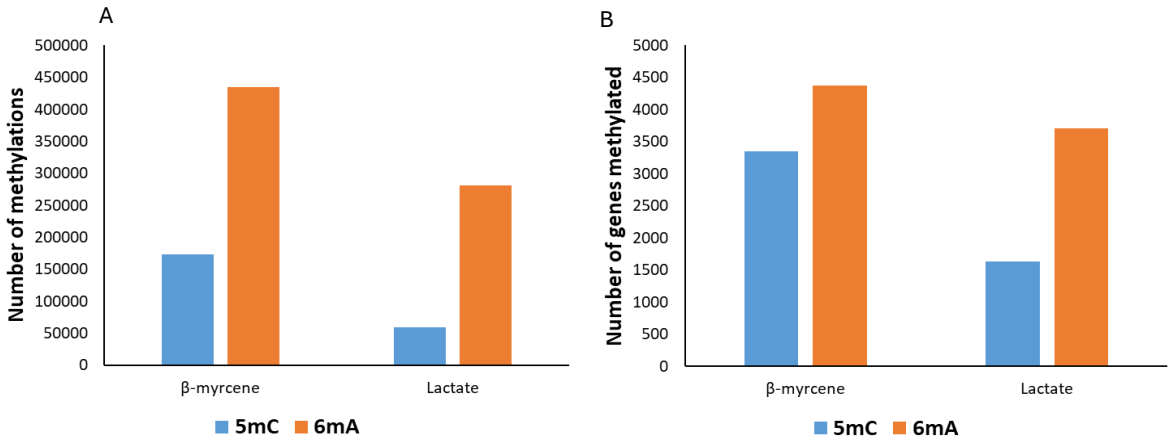


Figure 29 - Predicted number of methylations (A) and predicted number of methylated genes (B) for *Pseudomonas* sp. M1 cells growing in β-myrcene or lactate. In blue are the predicted numbers for 5mC methylation and in orange the predicted numbers for 6mA methylation.

As referred previously, the DNA mismatch repair system is dependent on DAM methylation. In fact, M. Carraway and collaborators showed that *E. coli* K12 strain with dam- phenotype had an increased spontaneous mutation rate in comparison with its dam+ counterparts[266]. Moreover, dam methylation is essential to regulate transposons activity, as mentioned above. Thus, the differences in mutation rates observed previously in chapter III during evolution for cells growing in β-myrcene or lactate may be related with genome methylation, mainly Dam methylation. However, as explained before, the models to predict the methylations in this technique may have failures. Thus, to understand if this correlation extends in other regions of the genome, the top 25 most expressed regions of the genome (for β-myrcene growth), based in a correlation with protein abundance were retrieved, as well as its methylation levels for β-myrcene vs lactate conditions (Table 29). This data was generated in Chapter IV of this thesis. Moreover, the fold-change levels in methylation between β-myrcene and lactate growth for all genes and intergenic regions of *Pseudomonas* sp. M1 genome are available in supplementary material Table S8 and Table S9. As expected, proteins specific to β-myrcene catabolism (belonging to the 28-kb genomic *locus*) showed the highest abundance among all the proteins identified in proteomic data for β-myrcene growth in contrast to lactate growth (data obtained in Chapter IV of this thesis). Several of these genes exhibited statistically significant methylation patterns for both 6mA and 5mC methylation (Table 29), with 5mC methylation ranging from 1 to 2.9 and 6mA methylation ranging between 1 and 3.3. Upon analyzing 6mA methylation events within genes encoding abundant proteins beyond the 28-kb genomic *locus* dedicated to β-myrcene catabolism, methylation levels exhibited a fold-change spanning from 1 to 2.2. Despite this range, these observations did not achieve statistical significance. However, analyzing the

5mC methylation type for these genes, the fold-change spanned from 1 to 5.8, with statistical significance observed in 4 genes – *MRY70_14540*, *MRY70_04880*, *MRY70_04875* and *MRY70_23555*. Remarkably, during the adaptation and evolution experiment (ALE) explained in Chapter III of this thesis, the genes *MRY70_14540*, *MRY70_04880*, and *MRY70_04875* exhibited an absence of detected mutations in the β-myrcene cultures and only 10% of the Myr 2 culture population displayed a detected mutation in the *MRY70_23555* gene. Conversely, within the lactate-evolving cultures, the gene *MRY70_04880* exhibited multiple mutations across Lac2 and Lac3 cultures, ranging from 5% to 25% of the population. Notably, all the other genes remained without detected mutations in lactate conditions. These genes code for proteins involved in the central metabolism, including the 2-methylcitrate cycle and the fatty-acids metabolism, already referred to in Chapter IV of this thesis. Among the genes highly methylated in β-myrcene and belonging to the 28-kb genomic locus, three lactate cultures showed mutations in MRY70_06120 (coding for the 3-hydroxyacyl-CoA dehydrogenase belonging to TU3) after 600 generations, in percentages ranging from 18% to 66% of the population. Notably, this gene represents the sole mutation within the genomic island. Considering the predictions generated by the models used in this study, when the gene MRY70_06120 exhibits enhanced methylation during β-myrcene growth mutations do not occur. Conversely, when methylation of this gene is diminished during lactate growth, mutation occurs. This observation aligns with the findings regarding gene MRY70_04880.

Table 29 – Correlation between dispersion and gene methylation of the 25 proteins most abundant in β-myrcene growing cells vs lactate growing cells (FC abundance from Chapter IV) and respective fold-change for 5mC and 6mA methylation. Significance was calculated based in the p-value. *- p-value between 0.05 and 0.1; **- p-value lower than 0.05.

	Relative abundance				Methylation dynamics 5mC		Methylation dynamics 6mA	
Protein code	**FC**	**Significance**	**Annotation**	**Locus-tag**	**FC**	**Significance**	**FC**	**Significance**
WP_009616244.1	8.7	*	acyl-CoA dehydrogenase family protein	MRY70_06085 (GI)	2.3	**	3	**
WP_009616255.1	7.6	**	3-hydroxyacyl-CoA dehydrogenase NAD-binding domain-containing protein	MRY70_06115 (GI)	2.1	**	1	*
WP_009616273.1	7.5	*	NAD(P)-dependent alcohol dehydrogenase	MRY70_06165 (GI)	1.3	-	1	-
WP_024128489.1	7.4	**	acyl-CoA dehydrogenase family protein	MRY70_06080 (GI)	1	-	1	-
WP_009616272.1	7.3	-	acyl-CoA synthetase	MRY70_06160 (GI)	2.2	**	5	**
WP_009616257.1	7.2	*	3-hydroxyacyl-CoA dehydrogenase	MRY70_06120 (GI)	1.4	-	2.8	-
WP_009616251.1	6.9	**	acetyl-CoA C-acetyltransferase	MRY70_06105 (GI)	2.9	**	1	-
WP_009616271.1	6.6	**	FAD/NAD(P)-binding oxidoreductase	MRY70_06155 (GI)	1.5	-	3.3	*
WP_009616253.1	6.6	**	nitronate monooxygenase family protein	MRY70_06110 (GI)	2.2	-	1	-
WP_024128761.1	6.2	**	3-hydroxyacyl-CoA dehydrogenase	MRY70_14540	2	**	1	-
WP_009618816.1	4.6	**	acetyl-CoA C-acetyltransferase	MRY70_11060	1	-	1.4	-
WP_009617184.1	4.5	-	bacterioferritin	MRY70_26820	1.4	-	1	-
WP_009622187.1	4.5	**	acyl-CoA dehydrogenase family protein	MRY70_14550	2	-	1.2	-
WP_009615988.1	4.4	**	methylisocitrate lyase	MRY70_04885	2.3	-	2.2	-
WP_009615987.1	4.4	*	2-methylcitrate synthase	MRY70_04880	3.0	*	1.5	-
WP_024128459.1	4.2	**	Fe/S-dependent 2-methylisocitrate dehydratase AcnD	MRY70_04875	5.7	**	1.6	-
WP_009617873.1	4.1	**	acyl-CoA dehydrogenase family protein	MRY70_24840	5.8	-	1	-
WP_009619571.1	4.1	**	5-methyltetrahydropteroyltriglutamate–homocysteine S-methyltransferase	MRY70_18045	2	-	1.5	-
WP_009613109.1	4.1	**	OprD family porin	MRY70_03300	1	-	1.4	-
WP_009613177.1	4	**	MetQ/NlpA family ABC transporter substrate-binding protein	MRY70_03090	1	-	1	-
WP_009615391.1	3.8	-	methyl-accepting chemotaxis protein	MRY70_27910	1.4	-	1	-
WP_009615985.1	3.7	**	2-methylaconitate cis-trans isomerase PrpF	MRY70_04870	1	-	1	-
WP_009616044.1	3.5	-	DUF1244 domain-containing protein	MRY70_05155	2	-	1	-
WP_009622919.1	3.5	*	aldehyde dehydrogenase	MRY70_12545	1.6	-	1.5	-
WP_009618290.1	3.4	-	type 4b pilus Flp biogenesis ATPase TadA	MRY70_23555	3.9	*	1.4	-

Comparing the methylation between genes and intergenic regions, only methylations 5mC were observed for intergenic regions. For this parameter, 30952 methylations were identified for lactate growth, corresponding to almost 1.8 times less methylations in intergenic regions than in coding regions and 55842 methylations were identified for β-myrcene growth, corresponding to 3 times less methylations in intergenic regions than in coding regions (Figure 30).

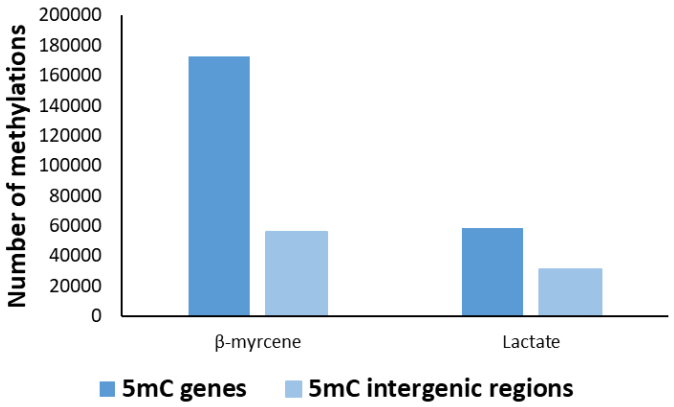


Figure 30 - Predicted number of 5mC methylations in genes (dark blue) or intergenic regions (light blue) in β-myrcene and lactate growing cells.

When analyzing methylation specifically in the genomic island related with β-myrcene catabolism, several genes were highly methylated after growth with β-myrcene, in comparison to lactate growth both in 6mA and 5mC methylation types (Table 30).

Besides genes, methylation in intergenic regions of the genomic island was also accessed. However, intergenic regions of the island were not much methylated, being only observed methylation of type 5mC upstream the locus-tag PM1_01191, coding for the protein WP_009616249.1 belonging to TU2, and upstream the locus-tag PM1_01207 coding for the protein WP_009616276.1 belonging to TU8. However, the methylation level observed for β-myrcene, or lactate growth was the same, showing no differences between both growth conditions.

Table 30 - Fold-change of 5mC and 6mA methylation predicted for proteins specific for β-myrcene catabolism, belonging to the genomic island, respective transcriptional unit and statistical significance. *- p-value between 0.05 and 0.1; **- p-value lower than 0.05.

Protein ID	Transcriptional unit	Annotation	6mA FC	Significance	5mC FC	Significance
WP_024128489.1	TU1	acyl-CoA dehydrogenase family protein	1.0	-	1.0	-
WP_009616244.1	TU1	acyl-CoA dehydrogenase family protein	2.3	**	3.0	**
WP_009616246.1	TU1	acetyl-CoA C-acyltransferase	2.1	*	1.0	-
WP_009616247.1	TU2	methyl-accepting chemotaxis protein	1.2	-	1.8	**
WP_009616249.1	TU2	outer membrane beta-barrel protein	1.0	-	1.0	-
WP_009616251.1	TU3	acetyl-CoA C-acetyltransferase	2.9	**	1.0	-
WP_009616253.1	TU3	nitronate monooxygenase family protein	2.2	-	1.0	-
WP_009616255.1	TU3	3-hydroxyacyl-CoA dehydrogenase NAD-binding domain-containing protein	2.1	**	1.0	*
WP_009616257.1	TU3	3-hydroxyacyl-CoA dehydrogenase	1.4	-	2.8	-
WP_009616259.1	TU3	epoxide hydrolase	4.4	-	1.0	-
WP_009616261.1	TU3	chalcone isomerase family protein	1.0	-	1.2	-
WP_009616263.1	TU4	LuxR C-terminal-related transcriptional regulator	1.9	-	2.7	-
WP_009616265.1	TU5	rubredoxin	1.0	-	1.0	-
WP_238603225.1	TU5)	fatty acid desaturase	1.4	-	1.0	-
WP_009616269.1	TU6	LuxR C-terminal-related transcriptional regulator	1.8	**	8.6	**
WP_009616271.1	TU6	FAD/NAD(P)-binding oxidoreductase	1.5	-	3.3	*
WP_009616272.1	TU6	acyl-CoA synthetase	2.2	**	5.0	**
WP_009616273.1	TU6	NAD(P)-dependent alcohol dehydrogenase	1.3	-	1.0	-
WP_052058996.1	TU6	aldehyde dehydrogenase family protein	1.3	-	4.4	-
WP_052058997.1	TU7	enoyl-CoA hydratase-related protein	1.1	-	1.0	-
WP_009616276.1	TU8	LysR substrate-binding domain-containing protein	2.0	-	5.8	**
WP_009616277.1	TU8	EexN family lipoprotein	1.0	-	1.0	-

In summary, the gDNA extracted from *Pseudomonas* sp. M1 grown in β-myrcene had a higher number of methylations predicted than the gDNA extracted from M1 cells grown in lactate. Moreover, the number of methylations predicted for Adenine (6mA) - Dam methylations - was higher than the number of methylations predicted for cytosine (5mC) - Dcm methylations - in both cases. Differences in mutation rates observed previously in Chapter III during the evolution of M1 cells adapting to β-myrcene or lactate may be related with genome methylation, mainly Dam methylation since M. Carraway and collaborators showed that E. coli K12 strain with dam- phenotype had an increased spontaneous mutation rate in

comparison with its dam+ counterparts. Thus, the higher number of mutations in M1 cells adapting to lactate may be correlated with the lower number of methylations predicted for this condition, when compared to β-myrcene. Regarding the methylation of genes belonging to the 28-kb *locus* related with the catabolism of β-myrcene, several genes showed increased methylation, with statistical significance, in M1 cells grown in β-myrcene when compared to lactate. This may justify the absence of mutations in this cluster, in M1 cells adapting to β-myrcene, since hypermethylation may protect these genes from mutation. The changes reported at the methylome level were supported by the identification and fold-change calculation of methyltransferases in the proteome of M1 *wt* strain, where one methyltransferase showed 1.45 times more abundance in β-myrcene than in lactate.

Chapter VI – Final remarks and Future Perspectives

6.1. Confirmation that *Pseudomonas* sp. M1 belongs to *Pseudomonas citronellolis* species

In Chapter II of this thesis, the taxonomic profile of *Pseudomonas* sp. M1 was reevaluated, corroborating that taxonomically this bacterium belongs to the *P. citronellolis* species. Specifically, this reevaluation involved 16S rRNA phylogeny studies and MLST, as well as whole-genome comparison approaches (pangenome, ANI and DDH) and metabolic fingerprinting with Biologs. In previous studies, several *Pseudomonas* sp. M1-like strains were isolated from the rhizosphere of Portuguese cultivars located in different geographic regions. All these isolates harbored the 28kb-GI associated with the catabolism of β-myrcene (as *Pseudomonas* sp. M1). Resorting to ANI comparison, these isolates showed between 99.8% and 100% identity with *Pseudomonas* sp. M1. However, despite the high similarity in the regions of the genome that overlap with M1, a percentage of the genome of the soil isolates did not overlap with M1 genome (UME83 – 1.3%, UMC76 – 1.8%, UMC3129 – 1.9%, UMA603 – 2.1%, UMA601 – 2.1%, UMC3103 – 2.6%, UMC631 – 2.7%, UMA643 – 2.8%, UMC3106 – 4%). However, similarities at the phylogenetic and phenotypic levels were significant. Resorting to phylogenetic analysis using the 16S rRNA gene, the isolates cluster together with *Pseudomonas* sp. M1 in the same clade. Through metabolic fingerprinting using Biologs, these isolates were very similar to *Pseudomonas* sp. M1 being grouped together upon PLS-DA analysis. Thus, *Pseudomonas* sp. M1 may be considered a new strain within the *P. citronellolis* species. A strain is defined as a genome or a set of genomes with no significant single-nucleotide or gene content differences that affect important phenotypic traits and can include multiple clones. It is worth noting that wild-type strains isolated from nature and their laboratory-adapted counterparts are often regarded as the same strain, despite potential differences in their phenotypes. To address this issue, some authors defend that a strain can be defined more precisely using a threshold of 99.99% average nucleotide identity (ANI), which ensures high gene content similarity (>99% of total genes) and thus, phenotypic similarity[267,268]. This is in accordance with the possibility raised.

Although *Pseudomonas* sp. M1 was considered as belonging to *P. citronellolis* species, the taxonomic classification of *Pseudomonas* species has been reviewed and updated several times, showing the challenge that this genus poses towards the classification at the species level. In fact, *Pseudomonas putida* KT2440, one of the most well-characterized model strains belonging to the *Pseudomonas* genus, was proposed for re-identification as *Pseudomonas alloputida* showing the difficulties of taxonomic classification of *Pseudomonas* species[269].

6.2. ALE increased the efficiency of *Pseudomonas* sp. M1 to mineralize β-myrcene

The main purpose of this thesis was to enhance the efficiency of *Pseudomonas* sp. M1 in metabolizing β-myrcene. The isolates derived from the Adaptive Laboratory Evolution (ALE) experiment, namely M2C19 and M3C22 isolates, exhibited significantly higher activation of the P5 promoter. This promoter regulates the transcription of genes coding for the first step of β-myrcene catabolism (oxidation), being a limiting and essential step for this pathway, which in turn led to an increased relative abundance of proteins encoded by genes within this cluster. These changes were linked to mutations in the *fleQ* gene, which encodes a transcriptional regulator that affects the transcription of motility-related genes. This hypothesis was verified with the construction of a mutant lacking the gene *fleQ* (M1ΔFleQ). This mutant showed reduced motility and enhanced aggregation, resulting in the increase of transcriptional activity of genes within the 28-kb genomic *locus* associated with β-myrcene metabolism and a higher abundance of the corresponding proteins. In fact, the ability of microorganisms to aggregate appears to be linked to enhanced degradation of hydrophobic compounds. Aggregation promotes the contact between bacteria and hydrophobic compounds, including hydrocarbons, enhancing substrate uptake and biodegradation rates. Experimental evidence in the literature shows that the breakdown of polycyclic aromatic hydrocarbons relies on bioavailability-enhancing mechanisms of bacteria, such as adhesion to the substrate. This supports the hypothesis that increased microbial aggregation contributes to improved degradation rates[270]. This was described for *Pseudomonas*, *Burkholderia* and *Comamonas* species towards the biodegradation of aromatic compounds[271].

Fluorescence and contrast microscopy confirmed the increased aggregation phenotype in both M2C19 and M3C22 isolates, supporting these findings. To better prove that the loss of motility and increased transcription of promoter P5 was indeed related with mutations in the gene *fleQ*, the ALE-derived isolates should be transformed with a vector expressing this gene non mutated. With this approach, the phenotype shown for the ALE-derived isolates should be once again similar to the ancestor strain, if the gene *fleQ* was indeed the only contributor for this phenotype (genotype complementation).

The increased efficiency of these mutants in metabolizing β-myrcene is theoretically correlated to the enhanced production of terpenoid metabolites via this metabolic pathway. This research is significant due to the industrial applications of terpenoids produced by *Pseudomonas* sp. M1 – for instance, myrcen-8-ol is widely used as a fragrance. Therefore, it is crucial to measure the production yields of these compounds, comparing the yields from the ancestor strain, the ALE-derived isolates, and the mutant M1ΔFleQ. This may be achieved through gas chromatography coupled with mass spectrometry (GC-MS) since these are volatile compounds. The procedure used for GC-MS measurements may be the same

used in previous works of this research group[127,167]. After verifying the increased production of terpenoids, it is necessary to interrupt the β-myrcene catabolic pathway to allow the accumulation of the desired terpenoid. This mutation may be performed in the ALE derived isolates to allow the accumulation of myrcen-8-ol.

6.3. Evaluation of the impact of *fleQ*-deletion in the biotransformation of other terpenes

Although the aim of this thesis was directed towards the compound β-myrcene, the mutations accumulated, namely the mutation in the *fleQ* gene, may be important for the growth of *Pseudomonas* sp. M1 using other terpenes. In fact, M1 showed the production of terpenoids with industrial interest using other terpenes, for instance limonene, as substrate, however with a lower yield[129]. Preliminary tests, growing the ALE-derived isolates M2C19 and M3C22 in MM supplemented only with limonene, showed a higher ability for the isolates to grow on this compound, in comparison to the ancestor strain, reaching a higher OD_{600nm} after growth for 24h in Limonene (data not shown). Thus, mutation in *fleQ* gene may be also important for the biotransformation of other terpenes.

6.4. Isolate and sequence β-myrcene adapted individuals with decreased susceptibility to antibiotics

As reported in Chapter III of this thesis, all the adapted cultures showed decreased susceptibility to antibiotics. Due to the challenges posed by working with such isolates, only those without decreased susceptibility were further analyzed. However, this decreased susceptibility seemed to be part of the evolution process of *Pseudomonas* sp. M1, since all cultures exhibited this trait even without the presence of antibiotics in the media. This behavior has been documented in the literature for *E. coli* and *Salmonella enterica* after 500 to 1000 generations in minimal media. Interestingly, no differences in susceptibility were found in lineages evolved in LB[272]. Therefore, the isolates with decreased antibiotic susceptibility may have developed adaptive mutations beneficial for the catabolism of β-myrcene, which were not present in the susceptible isolates. Future research should involve the extraction and analysis of the gDNA from these isolates.

6.5. Genome methylation levels may impact the evolution course as well as the efficiency of β-myrcene catabolism in *Pseudomonas* sp. M1

During the course of the ALE experiment, no mutations within the 28-kb *locus* necessary for the catabolism of β-myrcene were reported in cells adapting to this compound. The dynamics of methylome profile predicted for *Pseudomonas* sp. M1 *wt*, hinted hypermethylation on this cluster, when cells were grown in β-myrcene, thus protecting the cluster from mutations. However, mutations in this cluster may be essential for the increased efficiency towards the catabolism of β-myrcene. Thus, the deletion of DNA methyltransferases enzymes may decrease the methylations in this cluster and thus allow mutations in these genes. In fact, the targeted deletion of methyltransferases genes from *Pseudomonas* sp. M1 genome is already ongoing in our research group.

6.6. Further adaptive laboratorial evolution of *Pseudomonas* sp. M1 mutants towards terpenes

The evolution of Pseudomonas sp. M1 may be continued either in β-myrcene or in other terpenes to increase the yield of biotransformation of terpenes. However, taking into consideration the results of this thesis, ALE of mutants with multiple mutations could benefit the adaptation of this bacterium towards terpenes. In this way, the ALE-derived isolates M2C19, M3C22 and the mutant with *fleQ* gene deleted may be further evolved towards β-myrcene, after deletion of DNA methyltransferases. Theoretically, the deletion of methyltransferases may allow the mutation of the 28-kb GI indispensable for the catabolism of the terpene during the adaptation process.

6.7. Long-term perspectives

The results of this work allowed not only to construct evolved isolates with potentially higher efficiency to catabolize β-myrcene, but also to understand the mechanisms underlying this phenotype, as well as the mechanisms impacting the adaptation of this organism and the mutations accumulated in its genome. However, further steps must be taken to explore the biotechnological potential of *Pseudomonas* sp. M1 and the β-myrcene catabolic pathway. *Pseudomonas* sp. M1 is able to co-metabolize carbon sources. The most studied example is the co-metabolism of β-myrcene and lactate[125,194]. This allows the interruption of the β-myrcene pathway, enabling the accumulation of the terpenoid of interest while the cells keep growing thanks to the consumption of lactate. However, the patterns of mutation rates during co-metabolism are still unknown. On one hand, in the presence of β-myrcene only, M1 cells showed a higher stability of the genome as well as an increased methylation of its genome. However, using lactate

as main carbon source, its genome stability may decrease, as observed during the evolution in lactate only. In order to avoid decreased genome stability, the bacteria genome must be altered to render a lower mutation number during growth and avoid hypermutants. To accomplish this purpose, mobile elements (e.g. integrases, transposases, phage-related loci) must be deleted from M2C19 and M3C22 isolates genomes since these elements are known mobile elements, decreasing the genome stability[208,273]. In fact, mutations in phage-related genes were detected in both Lac and Myr cultures (e.g. gene *MRY70_27190*), as well as mutations in intergenic regions upstream genes coding for transposases in Myr cultures (e.g. gene *MRY70_21280*). The mobile elements can be predicted *in-silico* for instance using the *Prophinder*[274] and *PHASTER*[275] phage element databases, and its deletion may be performed using homologous recombination coupled with counter-selection with CRISPR-Cas9, the same procedure used in this thesis to create the M1ΔFleQ mutant.

After successfully increasing the yield of *Pseudomonas* sp. M1 for terpenoid production, such as myrcenol, and enhancing its genome stability, the next step should be to develop efficient purification methods for the terpenoids. Although no publications specifically detail the extraction of myrcen-8-ol, some research studies describe methods for purifying microbially produced hydrophobic terpenes. To allow the continuous extraction of monoterpenes, namely carvone, Janoschek and collaborators used membrane-assisted extraction that allowed the extraction of the terpenoid without damaging the producing host organism due to solvent contact, which is important in the biotech field. In this setup, a microporous membrane in a hollow-fibre contactor is placed between the solvent phase and the aqueous phase, allowing dispersion-free contact of the two phases and the purification of the terpene. The hollow-fibre adsorbs the terpene from the aqueous phase allowing its extraction with the solvent (n-hexane) and avoiding contact with the producing cells[276]. Another approach described in the literature is the use of adsorption-based chromatography. This approach was demonstrated for the terpene β-caryophyllene. In this setup, the supernatant where the producing host grew was passed through a column containing an adsorbent, in this case the polymeric material Rensa® RP, and eluted with ethanol, reaching a purity of >99% and a yield >80%[277].

Bibliography

1. Masyita, A., Sari, R. M., Astuti, A. D., Yasir, B., Rumata, N. R., Emran, T. B. & Simal-Gandara, J. Terpenes and terpenoids as main bioactive compounds of essential oils, their roles in human health and potential application as natural food preservatives. *Food Chem. X* **13**, (2022).
2. de Sousa, D. P., Damasceno, R. O. S., Amorati, R., Elshabrawy, H. A., de Castro, R. D., Bezerra, D. P. & Lima, T. C. Essential Oils: Chemistry and Pharmacological Activities. *Biomolecules* **13**, (2023).
3. Singh, B. & Sharma, R. A. Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. *3 Biotech* **5**, (2015).
4. Ninkuu, V. *et al.* Biochemistry of Terpenes and Recent Advances in Plant Protection. *Int. J. Mol. Sci.* **22**, (2021).
5. Tsolakis, N., Bam, W., Srai, J. S. & Kumar, M. Renewable chemical feedstock supply network design: The case of terpenes. *J. Clean. Prod.* **222**, (2019).
6. Corzo, D., Rosas-Villalva, D., Tostado-Blázquez, G., Alexandre, E. B., Hernandez, L. H., Han, J. & Baran, D. High-performing organic electronics using terpene green solvents from renewable feedstocks. *Nat. Energy* **8**, (2023).
7. Behr, A. & Johnen, L. Myrcene as a natural base chemical in sustainable chemistry: A critical review. *ChemSusChem* **2**, (2009).
8. Soares-Castro, P., Soares, F. & Santos, P. M. Current Advances in the Bacterial Toolbox for the Biotechnological Production of Monoterpene-Based Aroma Compounds. *Molecules* **26**, (2020).
9. Hornung, C. H., Álvarez-Diéguez, M. Á., Kohl, T. M. & Tsanaktsidis, J. Diels–Alder reactions of myrcene using intensified continuous-flow reactors. *Beilstein J. Org. Chem.* **13**, (2017).
10. Behr, A. & Johnen, L. Myrcene as a natural base chemical in sustainable chemistry: a critical review. *ChemSusChem* **2**, (2009).
11. Ivarsson, P. & Birgersson, G. Regulation and biosynthesis of pheromone components in the double spined bark beetle Ips duplicatus. *J. Insect Physiol.* **41**, (1995).
12. Bull, A. T., Ward, A. C. & Goodfellow, M. Search and discovery strategies for biotechnology: the paradigm shift. *Microbiol. Mol. Biol. Rev.* **64**, (2000).
13. Mardis, E. The Impact of Next-Generation Sequencing Technology on Genetics. *Trends Genet.* **24**, (2008).
14. Buermans, H. P. J. & den Dunnen, J. T. Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842**, (2014).
15. Pervez, M. T., Abbas, S. H., Moustafa, M. F., Aslam, N., & Shah, S. S. M. A Comprehensive

Review of Performance of Next-Generation Sequencing Platforms. *Biomed Res. Int.* **2022**, (2022).

16. Liu, Y., Rosikiewicz, W., Pan, Z., Jillette, N., Wang, P., Taghbalout, A. & Li, S. DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol.* **22**, (2021).
17. Athanasopoulou, K., Boti, M. A., Adamopoulos, P. G., Skourou, P. C. & Scorilas, A. Third-Generation Sequencing: The Spearhead towards the Radical Transformation of Modern Genomics. *Life (Basel, Switzerland)* **12**, (2021).
18. MacKenzie, M. & Argyropoulos, C. An Introduction to Nanopore Sequencing: Past, Present, and Future Considerations. *Micromachines* **14**, (2023).
19. Lin, B., Hui, J. & Mao, H. Nanopore Technology and Its Applications in Gene Sequencing. *Biosensors* **11**, (2021).
20. Jagodnik, J. & Gourse, R. L. Deciphering the RNA capping process in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **117**, (2020).
21. Hocine, S., Singer, R. H. & Grünwald, D. RNA processing and export. *Cold Spring Harb. Perspect. Biol.* **2**, (2010).
22. Luciano, D. J., Levenson-Palmer, R. & Belasco, J. G. Stresses that Raise Np(4)A Levels Induce Protective Nucleoside Tetraphosphate Capping of Bacterial RNA. *Mol. Cell* **75**, (2019).
23. Trinquier, A., Durand, S., Braun, F. & Condon, C. Regulation of RNA processing and degradation in bacteria. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1863**, (2020).
24. Conrad, T. M., Frazier, M., Joyce, A. R., Cho, B. K., Knight, E. M., Lewis, N. E. & Palsson, B. Ø. RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. *Proc. Natl. Acad. Sci. U. S. A.* **107**, (2010).
25. Sarkar, N. Polyadenylation of mRNA in prokaryotes. *Annu. Rev. Biochem.* **66**, (1997).
26. Hershey, J. W. B., Sonenberg, N. & Mathews, M. B. Principles of Translational Control. *Cold Spring Harb. Perspect. Biol.* **11**, (2019).
27. Smith, R. W. Mass Spectrometry. in *Encyclopedia of Forensic Sciences (Second Edition)* (eds. Siegel, J. A., Saukko, P. J. & Houck, M. M.) 603–608 (Academic Press, 2013).
28. Pitt, J. J. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin. Biochem. Rev.* **30**, (2009).
29. Aydoğan, C. Chapter 12 - Liquid chromatography-high resolution mass spectrometry for the analysis of bioactive natural products. in *Bioactive Natural Products* (ed. Atta-ur-Rahman) vol. 66

331–353 (Elsevier, 2020).

30. Sarker, S. D. & Nahar, L. Chapter 19 - Applications of High Performance Liquid Chromatography in the Analysis of Herbal Products. in *Evidence-Based Validation of Herbal Medicine* (ed. Mukherjee, P. K.) 405–425 (Elsevier, 2015).
31. Fernández-Costa, C., Martínez-Bartolomé, S., McClatchy, D. B., Saviola, A. J., Yu, N. K., & Yates III, J. R. Impact of the Identification Strategy on the Reproducibility of the DDA and DIA Results. *J. Proteome Res.* **19**, (2020).
32. Gregory, T. R. Understanding Natural Selection: Essential Concepts and Common Misconceptions. *Evol. Educ. Outreach* **2**, (2009).
33. Bradley, B. Natural selection according to Darwin: cause or effect? *Hist. Philos. Life Sci.* **44**, (2022).
34. Dimijian, G. G. Darwinian natural selection: its enduring explanatory power. *Proc. (Bayl. Univ. Med. Cent).* **25**, (2012).
35. Hershberg, R. Mutation–The Engine of Evolution: Studying Mutation and Its Role in the Evolution of Bacteria. *Cold Spring Harb. Perspect. Biol.* **7**, (2015).
36. Arenas, M., Araujo, N. M., Branco, C., Castelhano, N., Castro-Nallar, E., & Pérez-Losada, M. Mutation and recombination in pathogen evolution: Relevance, methods and controversies. *Infect. Genet. Evol.* **63**, (2018).
37. Bao, K., Melde, R. H. & Sharp, N. P. Are mutations usually deleterious? A perspective on the fitness effects of mutation accumulation. *Evol. Ecol.* **36**, (2022).
38. Loewe, L. & Hill, W. G. The population genetics of mutations: good, bad and indifferent. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, (2010).
39. De Visser, J. A. G. M. & Rozen, D. E. Clonal interference and the periodic selection of new beneficial mutations in *Escherichia coli*. *Genetics* **172**, (2006).
40. Raynes, Y., Halstead, A. L. & Sniegowski, P. D. The effect of population bottlenecks on mutation rate evolution in asexual populations. *J. Evol. Biol.* **27**, (2014).
41. Fox, R. J., Donelson, J. M., Schunter, C., Ravasi, T. & Gaitán-Espitia, J. D. Beyond buying time: the role of plasticity in phenotypic adaptation to rapid environmental change. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* , (2019).
42. Rolland, J. *et al.* Conceptual and empirical bridges between micro- and macroevolution. *Nat. Ecol. Evol.* **7**, (2023).
43. Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D. & Kim, J. F. Genome

evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, (2009).

44. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* **4**, (2003).

45. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* **4**, (2003).

46. Schlötterer, C., Kofler, R., Versace, E., Tobler, R. & Franssen, S. U. Combining experimental evolution with next-generation sequencing: A powerful tool to study adaptation from standing genetic variation. *Heredity (Edinb).* **114**, (2015).

47. Bennett, A. F., Lenski, R. E. & Mittler, J. E. Evolutionary Adaptation to Temperature. I. Fitness Responses of *Escherichia coli* to Changes in its Thermal Environment. *Evolution (N. Y).* **46**, (1992).

48. Corl, A., Bi, K., Luke, C., Challa, A. S., Stern, A. J., Sinervo, B., & Nielsen, R. The Genetic Basis of Adaptation following Plastic Changes in Coloration in a Novel Environment. *Curr. Biol.* **28**, (2018).

49. Cooper, T. F., Rozen, D. E. & Lenski, R. E. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **100**, (2003).

50. Schluter, D., Clifford, E. A., Nemethy, M. & McKinnon, J. S. Parallel evolution and inheritance of quantitative traits. *Am. Nat.* **163**, (2004).

51. Microbial Evolution: This report is based on a colloquium convened by the American Academy of Microbiology on August 28-30, 2009, in San Cristobal, Ecuador. Washington (DC): American Society for Microbiology; *EMBO Rep.* **20**, (2019).

53. Papadopoulos, D., Schneider, D., Meier-Eiss, J., Arber, W., Lenski, R. E., & Blot, M. Genomic evolution during a 10,000-generation experiment with bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **96**, (1999).

54. Bachmann, H., Pronk, J. T., Kleerebezem, M. & Teusink, B. Evolutionary engineering to enhance starter culture performance in food fermentations. *Curr. Opin. Biotechnol.* **32**, (2015).

55. LaCroix, R. A., Sandberg, T. E., O'Brien, E. J., Utrilla, J., Ebrahim, A., Guzman, G. I. & Feist, A. M. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of *Escherichia coli* K-12 MG1655 on glucose minimal medium. *Appl. Environ. Microbiol.* **81**, (2015).

56. Lee, D. H. & Palsson, B. O. Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a nonnative carbon source, L-l,2-propanediol. *Appl. Environ. Microbiol.* **76**, (2010).

57. Barrett, R. D. H., MacLean, R. C. & Bell, G. Experimental evolution of *Pseudomonas fluorescens*

in simple and complex environments. *Am. Nat.* **166**, (2005).

58. Sardessai, Y. & Bhosle, S. Tolerance of bacteria to organic solvents. *Res. Microbiol.* **153**, (2002).
59. Kuepper, J., Otto, M., Dickler, J., Behnken, S., Magnus, J., Jäger, G., & Wierckx, N. Adaptive laboratory evolution of *Pseudomonas putida* and *Corynebacterium* glutamicum to enhance anthranilate tolerance. *Microbiology* **166**, 1025–1037 (2020).
60. Shibai, A. Mutation accumulation under UV radiation in *Escherichia coli*. *Sci. Rep.* **7**, 14531 (2017).
61. Pfeifer, G. P. Mechanisms of UV-induced mutations and skin cancer. *Genome Instab. Dis.* **1**, (2020).
62. Dong, Y., Lin, H., Wang, H., Mo, X., Fu, K., & Wen, H. Effects of ultraviolet irradiation on bacteria mutation and bioleaching of low-grade copper tailings. *Miner. Eng.* **24**, (2011).
63. Kodym, A. & Afza, R. Physical and chemical mutagenesis. *Methods Mol. Biol.* **236**, (2003).
64. Sikora, P., Chawade, A., Larsson, M., Olsson, J. & Olsson, O. Mutagenesis as a tool in plant genetics, functional genomics, and breeding. *Int. J. Plant Genomics* **2011**, (2011).
65. Li, N. Tn5 Transposase Applied in Genomics Research. *Int. J. Mol. Sci.* **21**, (2020).
66. Xu, T., Bharucha, N. & Kumar, A. Genome-wide transposon mutagenesis in Saccharomyces cerevisiae and Candida albicans. *Methods Mol. Biol.* **765**, (2011).
67. Muñoz-López, M. & García-Pérez, J. L. DNA transposons: nature and applications in genomics. *Curr. Genomics* **11**, (2010).
68. Redman, M., King, A., Watson, C. & King, D. What is CRISPR/Cas9? *Arch. Dis. Child. Educ. Pract. Ed.* **101**, (2016).
69. Csörgő, B., Nyerges, A. & Pál, C. Targeted mutagenesis of multiple chromosomal regions in microbes. *Curr. Opin. Microbiol.* **57**, (2020).
70. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, (2013).
71. Aparicio, T., de Lorenzo, V. & Martínez-García, E. CRISPR/Cas9-enhanced ssDNA recombineering for *Pseudomonas putida*. *Microb. Biotechnol.* **12**, (2019).
72. Steele, H. L. & Streit, W. R. Metagenomics: Advances in ecology and biotechnology. *FEMS Microbiol. Lett.* **247**, (2005).
73. Coughlan, L. M., Cotter, P. D., Hill, C. & Alvarez-Ordóñez, A. Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. *Front. Microbiol.* **6**, (2015).
74. Schoch, C. L. NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* **2020**, (2020).

75. Kumar Awasthi, M. Metagenomics for taxonomy profiling: tools and approaches. *Bioengineered* **11**, (2020).

76. Kim, S. & Goodfellow, M. Bergey's Manual of Systematics of Archaea and Bacteria. (2015).

77. Vaneechoutte, M. & Heyndrickx, M. 9 - Application and Analysis of ARDRA Patterns in Bacterial Identification, Taxonomy and Phylogeny. in (eds. Dijkshoorn, L., Towner, K. J. & Struelens, M. B. T.-N. A. for the G. and A. of M. T. D.) (Elsevier Science B.V., 2001).

78. Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C. & Baird, D. J. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One* **6**, (2011).

79. Kress, W. J. & Erickson, D. L. DNA barcodes: genes, genomics, and bioinformatics. *Proc. Natl. Acad. Sci. U. S. A.* **105**, (2008).

80. Ibarz Pavón, A. B. & Maiden, M. C. J. Multilocus sequence typing. *Methods Mol. Biol.* **551**, (2009).

81. Maiden, M. C. J. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature reviews. Microbiology* vol. (2013).

82. Yoon, S.-H., Ha, S.-M., Lim, J., Kwon, S. & Chun, J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* **110**, (2017).

83. Arahal, D. R. Chapter 6 - Whole-Genome Analyses: Average Nucleotide Identity. in *New Approaches to Prokaryotic Systematics* (eds. Goodfellow, M., Sutcliffe, I. & Chun, J. B. T.-M. in M.) vol. 41 103–122 (Academic Press, 2014).

84. Auch, A. F., Klenk, H.-P. & Göker, M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand. Genomic Sci.* **2**, (2010).

85. Peix, A., Ramírez-Bahena, M.-H. & Velázquez, E. Historical evolution and current status of the taxonomy of genus *Pseudomonas*. *Infect. Genet. Evol.* **9**, (2009).

86. Peix, A., Ramírez-Bahena, M.-H. & Velázquez, E. The current status on the taxonomy of Pseudomonas revisited: An update. *Infect. Genet. Evol.* **57**, (2018).

87. Rouli, L., Merhej, V., Fournier, P.-E. & Raoult, D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New microbes new Infect.* **7**, (2015).

88. Rosconi, F. A bacterial pan-genome makes gene essentiality strain-dependent and evolvable. *Nat. Microbiol.* **7**, (2022).

89. Yi, B. & Dalpke, A. H. Revisiting the intrageneric structure of the genus *Pseudomonas* with complete whole genome sequence information: Insights into diversity and pathogen-related

genetic determinants. *Infect. Genet. Evol.* **97**, (2022).

90. Kunitsky, C., Osterhout, G. & Sasser, M. Identification of microorganisms using fatty acid methyl ester (FAME) analysis and the MIDI Sherlock Microbial Identification System. *Encycl. Rapid Microbiol. Methods* **3**, (2006).
91. El-Liethy, M. BIOLOG GEN III for Identification of microbial isolates. (2017).
92. Medić, A. B. & Karadžić, I. M. *Pseudomonas* in environmental bioremediation of hydrocarbons and phenolic compounds- key catabolic degradation enzymes and new analytical platforms for comprehensive investigation. *World J. Microbiol. Biotechnol.* **38**, (2022).
93. Vélez, J. M. B., Martínez, J. G., Ospina, J. T. & Agudelo, S. O. Bioremediation potential of *Pseudomonas* genus isolates from residual water, capable of tolerating lead through mechanisms of exopolysaccharide production and biosorption. *Biotechnol. reports (Amsterdam, Netherlands)* **32**, (2021).
94. Raghavan, P. U. M. & Vivekanandan, M. Bioremediation of oil-spilled sites through seeding of naturally adapted *Pseudomonas putida*. *Int. Biodeterior. Biodegradation* **44**, (1999).
95. Ambust, S., Das, A. J. & Kumar, R. Bioremediation of petroleum contaminated soil through biosurfactant and *Pseudomonas* sp. SA3 amended design treatments. *Curr. Res. Microb. Sci.* **2**, (2021).
96. Ferschl, A., Loidl, M., Ditzelmüller, G., Hinteregger, C. & Streichsbier, F. Continuous degradation of 3-chloroaniline by calcium-alginate-entrapped cells of *Pseudomonas acidovorans* CA28: influence of additional substrates. *Appl. Microbiol. Biotechnol.* **35**, (1991).
97. Lee, S. W., Glickmann, E. & Cooksey, D. A. Chromosomal locus for cadmium resistance in *Pseudomonas putida* consisting of a cadmium-transporting ATPase and a MerR family response regulator. *Appl. Environ. Microbiol.* **67**, (2001).
98. Saxena, D., Gowri, P. M., Mago, R. & Srivastav, S. Removal of copper by *Pseudomonas putida* strain S4 isolated from copper mines. *Indian J. Exp. Biol.* **39**, (2001).
99. Bettmann, H. & Rehm, H. J. Continuous degradation of phenol(s) by *Pseudomonas putida* P8 entrapped in polyacrylamidehydrazide. *Appl. Microbiol. Biotechnol.* **22**, (1985).
100. Wang, C. L. Cadmium removal by a new strain of *Pseudomonas aeruginosa* in aerobic culture. *Appl. Environ. Microbiol.* **63**, 4075–4078 (1997).
101. Wang, X. A Biocontrol Strain of Pseudomonas aeruginosa CQ-40 Promote Growth and Control *Botrytis cinerea* in Tomato. *Pathog. (Basel, Switzerland)* **10**, (2020).
102. Ganeshan, G. & Kumar, A. M. *Pseudomonas fluorescens*, a potential bacterial antagonist to

control plant diseases. *J. Plant Interact.* **1**, (2005).

103. Abo-Elyousr, K. A. M., Abdel-Rahim, I. R., Almasoudi, N. M. & Alghamdi, S. A. Native Endophytic *Pseudomonas putida* as a Biocontrol Agent against Common Bean Rust Caused by Uromyces appendiculatus. *J. fungi (Basel, Switzerland)* **7**, (2021).
104. Mikiciński, A., Puławska, J., Molzhigitova, A. & Sobiczewski, P. Bacterial species recognized for the first time for its biocontrol activity against fire blight (Erwinia amylovora). *Eur. J. Plant Pathol.* **156**, (2020).
105. De Vleesschauwer, D., Djavaheri, M., Bakker, P. A. H. M. & Höfte, M. *Pseudomonas fluorescens* WCS374r-induced systemic resistance in rice against Magnaporthe oryzae is based on pseudobactin-mediated priming for a salicylic acid-repressible multifaceted defense response. *Plant Physiol.* **148**, (2008).
106. Wang, D., Xu, A., Elmerich, C. & Ma, L. Z. Biofilm formation enables free-living nitrogen-fixing rhizobacteria to fix nitrogen under aerobic conditions. *ISME J.* **11**, (2017).
107. He, D. Distribution of Culturable Phosphate-Solubilizing Bacteria in. **10**, (2000).
108. Patten, C. L. & Glick, B. R. Role of Pseudomonas putida indoleacetic acid in development of the host plant root system. *Appl. Environ. Microbiol.* **68**, (2002).
109. Craig, K., Johnson, B. R. & Grunden, A. Leveraging *Pseudomonas* Stress Response Mechanisms for Industrial Applications. *Front. Microbiol.* **12**, (2021).
110. Kawahara, H., Li, J., Griffith, M. & Glick, B. Relationship Between Antifreeze Protein and Freezing Resistance in *Pseudomonas putida* GR12-2. *Curr. Microbiol.* **43**, (2001).
111. Mehmood, N. Multifaceted Impacts of Plant-Beneficial *Pseudomonas* spp. in Managing Various Plant Diseases and Crop Yield Improvement. *ACS omega* **8**, (2023).
112. Wang, S. *Pseudomonas* spp. as cell factories (MCFs) for value-added products: from rational design to industrial applications. *Crit. Rev. Biotechnol.* **40**, (2020).
113. Mozejko-Ciesielska, J. Chapter 10 - *Pseudomonas putida*–based cell factories. in (ed. Singh, V. B. T.-M. C. F. E. for P. of B.) 165–181 (Academic Press, 2021).
114. Ruchi, G., Anshu, G. & Khare, S. K. Lipase from solvent tolerant *Pseudomonas aeruginosa* strain: Production optimization by response surface methodology and application. *Bioresour. Technol.* **99**, (2008).
115. Bahia, F. M. Rhamnolipids production from sucrose by engineered *Saccharomyces cerevisiae*. *Sci. Rep.* **8**, (2018).
116. Tienda, S. Soil Application of a Formulated Biocontrol Rhizobacterium, P*seudomonas*

chlororaphis PCL1606, Induces Soil Suppressiveness by Impacting Specific Microbial Communities. *Front. Microbiol.* **11**, (2020).

117. Anderson, J. A., Staley, J., Challender, M. & Heuton, J. Safety of *Pseudomonas chlororaphis* as a gene source for genetically modified crops. *Transgenic Res.* **27**, (2018).
118. Meehan, S., Shannon, A., Gruber, B., Rackl, S. M. & Lucy, F. E. Ecotoxicological impact of Zequanox®, a novel biocide, on selected non-target Irish aquatic species. *Ecotoxicol. Environ. Saf.* **107**, (2014).
119. Dias, F. M. S., Gomez, J. G. C. & Silva, L. F. da. Exploring the microbial production of aromatic fine chemicals to overcome the barriers of traditional methods. *Adv. Appl. Sci. Res* **8**, (2017).
120. EPA. MUNOX SR®.
121. Santos, P. M. & Sá-Correia, I. Characterization of the unique organization and co-regulation of a gene cluster required for phenol and benzene catabolism in *Pseudomonas* sp. M1. *J. Biotechnol.* **131**, (2007).
122. Santos, P. M., Mignogna, G., Heipieper, H. J. & Zennaro, E. Occurrence and properties of glutathione S-transferases in phenol-degrading *Pseudomonas* strains. *Res. Microbiol.* **153**, (2002).
123. Santos, P. M. *et al.* Mechanistic insights into the global response to phenol in the phenol-biodegrading strain *Pseudomonas* sp. M1 revealed by quantitative proteomics. *OMICS* **11**, (2007).
124. Iurescia, S. *et al.* Identification and sequencing of beta-myrcene catabolism genes from *Pseudomonas* sp. strain M1. *Appl. Environ. Microbiol.* **65**, (1999).
125. Santos, P. M. & Sá-Correia, I. Adaptation to beta-myrcene catabolism in *Pseudomonas* sp. M1: an expression proteomics analysis. *Proteomics* **9**, (2009).
126. Soares-Castro, P. & Santos, P. M. Deciphering the genome repertoire of *Pseudomonas* sp. M1 toward β-Myrcene biotransformation. *Genome Biol. Evol.* **7**, (2014).
127. Strain, M. & Santos, P. M. Functional Characterization of a 28-Kilobase Catabolic Island from Related Plant-Derived Volatiles. *Appl. Environ. Microbiol.* **83**, (2017).
128. Soares-Castro, P. & Santos, P. M. Deciphering the genome repertoire of *Pseudomonas* sp. M1 toward β-myrcene biotransformation. *Genome Biol. Evol.* **7**, (2014).
129. Soares-Castro, P., Montenegro-Silva, P., Heipieper, H. J. & Santos, P. M. Functional characterization of a 28-kilobase catabolic island from *Pseudomonas* sp. strain M1 involved in biotransformation of β-myrcene and related plant-derived volatiles. *Appl. Environ. Microbiol.* **83**,

(2017).

130. Soares-Castro, P., Montenegro-Silva, P., Heipieper, H. J. & Santos, P. M. Functional Characterization of a 28-Kilobase Catabolic Island from *Pseudomonas* sp. Strain M1 Involved in Biotransformation of β-Myrcene and Related Plant-Derived Volatiles. *Appl. Environ. Microbiol.* **83**, (2017).

131. Santos, P. M. & Sá-Correia, I. Characterization of the unique organization and co-regulation of a gene cluster required for phenol and benzene catabolism in *Pseudomonas* sp. M1. *J. Biotechnol.* **131**, (2007).

132. Santos, P. M. & Sá-Correia, I. Adaptation to β-myrcene catabolism in *Pseudomonas* sp. M1: An expression proteomics analysis. *Proteomics* **9**, (2009).

133. Tayeb, L., Elisabeth, A., Grimont, F. & Grimont, P. Molecular phylogeny of the genus *Pseudomonas* based on *rpoB* sequences and application for identification of isolates. *Res. Microbiol.* **156**, (2005).

134. Andreani, N. A. Tracking the blue: a MLST approach to characterise the *Pseudomonas fluorescens* group. *Food Microbiol.* **39**, (2014).

135. Ogura, K., Shimada, K. & Miyoshi-Akiyama, T. A multilocus sequence typing scheme of *Pseudomonas putida* for clinical and environmental isolates. *Sci. Rep.* **9**, (2019).

136. Curran, B., Jonas, D., Grundmann, H., Pitt, T. & Dowson, C. G. Development of a multilocus sequence typing scheme for the opportunistic pathogen *Pseudomonas aeruginos*a. *J. Clin. Microbiol.* **42**, (2004).

137. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, (2013).

138. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **38**, (2021).

139. Ojha, K. K., Mishra, S. & Singh, V. K. Chapter 5 - Computational molecular phylogeny: concepts and applications. in (eds. Singh, D. B. & Pathak, R. K. B. T.-B.) 67–89 (Academic Press, 2022).

140. Kimura, M. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci.* **78**, (1981).

141. Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**, (1981).

142. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**,

(2018).

143. Jain, C., Koren, S., Dilthey, A., Phillippy, A. M. & Aluru, S. A fast adaptive algorithm for computing whole-genome homology maps. *Bioinformatics* **34**, (2018).

144. Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L. & Göker, M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res.* **50**, (2022).

145. Gomila, M., Peña, A., Mulet, M., Lalucat, J. & García-Valdés, E. Phylogenomics and systematics in Pseudomonas. *Front. Microbiol.* **6**,(2015).

146. Shaiber, A. *et al.* Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. *Genome Biol.* **21**, (2020).

147. Shea, A., Wolcott, M., Daefler, S. & Rozak, D. A. Biolog phenotype microarrays. *Methods Mol. Biol.* **881**, (2012).

148. Hartmans, S., Smits, J. P., van der Werf, M. J., Volkering, F. & de Bont, J. A. Metabolism of Styrene Oxide and 2-Phenylethanol in the Styrene-Degrading *Xanthobacter* Strain 124X. *Appl. Environ. Microbiol.* **55**, (1989).

149. Team, R. C. R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna, Austria.* (2021).

150. Kucheryavskiy, S. mdatools – R package for chemometrics. *Chemom. Intell. Lab. Syst.* **198**, (2020).

151. Villanueva, R. A. M. & Chen, Z. J. ggplot2: Elegant Graphics for Data Analysis (2nd ed.). *Meas. Interdiscip. Res. Perspect.* **17**, (2019).

152. Gupta, S. K., Kumari, R., Prakash, O. & Lal, R. *Pseudomonas panipatensis* sp. nov., isolated from an oil-contaminated site. *Int. J. Syst. Evol. Microbiol.* **58**, (2008).

153. Iurescia, S. *et al.* Identification and sequencing of beta-myrcene catabolism genes from *Pseudomonas* sp. strain M1. *Appl. Environ. Microbiol.* **65**, (1999).

154. Moore, E. R. B. Nonmedical: P*seudomonas* BT - The Prokaryotes: A Handbook on the Biology of Bacteria Volume 6: Proteobacteria: Gamma Subclass. in (eds. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H. & Stackebrandt, E.) 646–703 (Springer New York, 2006).

155. Ciufo, S. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *Int. J. Syst. Evol. Microbiol.* **68**, (2018).

156. Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A.* **106**, (2009).

157. Shahid, F. Chapter 13 - Applications of microbial omics in healthcare. in (ed. Barh, D. B. T.-B. in H.) 227–248 (Academic Press, 2022).

158. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, (2021).

159. Van Dongen, S. Graph Clustering Via a Discrete Uncoupling Process. *SIAM J. Matrix Anal. Appl.* **30**, (2008).

160. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, (2004).

161. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, (2010).

162. Bhattacharya, D., Sarma, P. M., Krishnan, S., Mishra, S. & Lal, B. Evaluation of genetic diversity among *Pseudomonas citronellolis* strains isolated from oily sludge-contaminated sites. *Appl. Environ. Microbiol.* **69**, (2003).

163. Avise, J. C. & Ayala, F. J. In the light of evolution III: two centuries of Darwin. *Proc. Natl. Acad. Sci. U. S. A.* **106 Suppl**, (2009).

164. Förster-Fromme, K. & Jendrossek, D. Identification and characterization of the acyclic terpene utilization gene cluster of *Pseudomonas citronellolis*. *FEMS Microbiol. Lett.* **264**, (2006).

165. Harder, J. & Probian, C. Microbial degradation of monoterpenes in the absence of molecular oxygen. *Appl. Environ. Microbiol.* **61**, (1995).

166. Koutinas, M. *et al.* Enhanced biodegradation and valorization of drilling wastewater via simultaneous production of biosurfactants and polyhydroxyalkanoates by *Pseudomonas citronellolis* SJTE-3. *Bioresour. Technol.* **340**, 125679 (2021).

167. Soares-Castro, P., Soares, F., Reis, F., Lino-Neto, T. & Santos, P. M. Bioprospection of the bacterial β-myrcene-biotransforming trait in the rhizosphere. *Appl. Microbiol. Biotechnol.* **107**, (2023).

168. Blount, Z. D., Borland, C. Z. & Lenski, R. E. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **105**, (2008).

169. Ramasamy, K. P., Brugel, S., Eriksson, K. I. A. & Andersson, A. *Pseudomonas* ability to utilize different carbon substrates and adaptation influenced by protozoan grazing. *Environ. Res.* **232**, (2023).

170. Barrick, J. E. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*.

Nature **461**, (2009).

171. Mathivanan, K. Bacterial adaptive strategies to cope with metal toxicity in the contaminated environment – A review. *Ecotoxicol. Environ. Saf.* **226**, (2021).
172. Ahmad, S., Ahmad, H. W. & Bhatt, P. Microbial adaptation and impact into the pesticide's degradation. *Arch. Microbiol.* **204**, (2022).
173. Cisneros-Mayoral, S., Graña-Miraglia, L., Pérez-Morales, D., Peña-Miller, R. & Fuentes-Hernández, A. Evolutionary History and Strength of Selection Determine the Rate of Antibiotic Resistance Adaptation. *Mol. Biol. Evol.* **39**, (2022).
174. López-Maury, L., Marguerat, S. & Bähler, J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat. Rev. Genet.* **9**, (2008).
175. Huang, W. *et al.* Spontaneous mutations and the origin and maintenance of quantitative genetic variation. *Elife* **5**, (2016).
176. Fisher, R. A. XXI.—On the dominance ratio. *Proc. R. Soc. Edinburgh* **42**, (1923).
177. Fisher, R. A. XVII.—The distribution of gene ratios for rare mutations. *Proc. R. Soc. Edinburgh* **50**, (1931).
178. Mutation Rates and Gene Location: Some Like It Hot. *PLoS Biology* **2**, (2004).
179. Desai, M. M. & Fisher, D. S. Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* **176**, (2007).
180. Kim, Y. & Maruki, T. Hitchhiking effect of a beneficial mutation spreading in a subdivided population. *Genetics* **189**, (2011).
181. Lenski, R. E. Revisiting the Design of the Long-Term Evolution Experiment with *Escherichia coli*. *J. Mol. Evol.* **91**, (2023).
182. Dragosits, M. & Mattanovich, D. Adaptive laboratory evolution – principles and applications for biotechnology. *Microb. Cell Fact.* **12**, (2013).
183. Hoang, K. L., Morran, L. T. & Gerardo, N. M. Experimental Evolution as an Underutilized Tool for Studying Beneficial Animal-Microbe Interactions. *Front. Microbiol.* **7**, (2016).
184. Kang, M. *et al.* Inactivation of a Mismatch-Repair System Diversifies Genotypic Landscape of Escherichia coli During Adaptive Laboratory Evolution. *Front. Microbiol.* **10**, (2019).
185. de Visser, J. A. G. M. & Lenski, R. E. Long-term experimental evolution in *Escherichia coli*. XI. Rejection of non-transitive interactions as cause of declining rate of adaptation. *BMC Evol. Biol.* **2**, (2002).
186. Bailey, S. F., Blanquart, F., Bataillon, T. & Kassen, R. What drives parallel evolution?: How

population size and mutational variation contribute to repeated evolution. *Bioessays* **39**, (2017).

187. Herron, M. D. & Doebeli, M. Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. *PLoS Biol.* **11**, (2013).

188. Seong, W. *et al.* Adaptive laboratory evolution of *Escherichia coli* lacking cellular byproduct formation for enhanced acetate utilization through compensatory ATP consumption. *Metab. Eng.* **62**, (2020).

189. Li, W.-J. Laboratory evolution reveals the metabolic and regulatory basis of ethylene glycol metabolism by *Pseudomonas putida* KT2440. *Environ. Microbiol.* **21**, (2019).

190. Lu, Q., Zhou, X.-L. & Liu, J.-Z. Adaptive laboratory evolution and shuffling of *Escherichia coli* to enhance its tolerance and production of astaxanthin. *Biotechnol. Biofuels Bioprod.* **15**, (2022).

191. Yu, Z. & Morrison, M. Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* **36**, (2004).

192. Chen, W. *et al.* CRISPR/Cas9-based Genome Editing in *Pseudomonas aeruginosa* and Cytidine Deaminase-Mediated Base Editing in *Pseudomona*s Species. *iScience* **6**, (2018).

193. Ha, D.-G., Kuchma, S. L. & O'Toole, G. A. Plate-based assay for swimming motility in *Pseudomonas aeruginosa*. *Methods Mol. Biol.* **1149**, (2014).

194. Soares-Castro, P. & Santos, P. M. Deciphering the genome repertoire of *Pseudomonas* sp. M1 toward β-Myrcene biotransformation. *Genome Biol. Evol.* **7**, (2014).

195. Notley-McRobb, L., King, T. & Ferenci, T. *rpoS* mutations and loss of general stress resistance in Escherichia coli populations as a consequence of conflict between competing stress responses. *J. Bacteriol.* **184**, (2002).

196. Shimizu, K. 3 - Metabolic regulation by global regulators in response to culture environment. in *Woodhead Publishing Series in Biomedicine* (ed. Shimizu, K. B. T.-B. C. M. S.) 95–213 (Woodhead Publishing, 2013).

197. Vijayakumar, S. R. V, Kirchhof, M. G., Patten, C. L. & Schellhorn, H. E. RpoS-regulated genes of *Escherichia coli* identified by random lacZ fusion mutagenesis. *J. Bacteriol.* **186**, (2004).

198. Maslowska, K. H., Makiela-Dzbenska, K. & Fijalkowska, I. J. The SOS system: A complex and tightly regulated response to DNA damage. *Environ. Mol. Mutagen.* **60**, (2019).

199. Sprouffske, K., Aguilar-Rodríguez, J., Sniegowski, P. & Wagner, A. High mutation rates limit evolutionary adaptation in *Escherichia coli*. *PLoS Genet.* **14**, (2018).

200. Denamur, E. & Matic, I. Evolution of mutation rates in bacteria. *Mol. Microbiol.* **60**, (2006).

201. Ramiro, R. S., Durão, P., Bank, C. & Gordo, I. Low mutational load and high mutation rate variation

in gut commensal bacteria. *PLoS Biol.* **18**, (2020).

202. Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **109**, (2012).
203. Lampe, N. Reducing the ionizing radiation background does not significantly affect the evolution of Escherichia coli populations over 500 generations. *Sci. Rep.* **9**, (2019).
204. Vinchhi, R., Jena, C. & Matange, N. Adaptive laboratory evolution of antimicrobial resistance in bacteria for genetic and phenotypic analyses. *STAR Protoc.* **4**, (2023).
205. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol. Biol.* **1151**, (2014).
206. Deatherage, D. E., Traverse, C. C., Wolf, L. N. & Barrick, J. E. Detecting rare structural variation in evolving microbial populations from new sequence junctions using breseq. *Front. Genet.* **5**, (2014).
207. Barrick, J. E. Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. *BMC Genomics* **15**, (2014).
208. Darmon, E. & Leach, D. R. F. Bacterial genome instability. *Microbiol. Mol. Biol. Rev.* **78**, (2014).
209. Rodrigues-Prause, A. *et al.* A Case Study of Genomic Instability in an Industrial Strain of *Saccharomyces cerevisiae*. *G3 (Bethesda).* **8**, (2018).
210. Kadisch, M., Willrodt, C., Hillen, M., Bühler, B. & Schmid, A. Maximizing the stability of metabolic engineering-derived whole-cell biocatalysts. *Biotechnol. J.* **12**, (2017).
211. Saxer, G. Mutations in global regulators lead to metabolic selection during adaptation to complex environments. *PLoS Genet.* **10**, (2014).
212. Jaishankar, J. & Srivastava, P. Molecular Basis of Stationary Phase Survival and Applications. *Front. Microbiol.* **8**, (2017).
213. Kinnersley, M., Schwartz, K., Yang, D.-D., Sherlock, G. & Rosenzweig, F. Evolutionary dynamics and structural consequences of de novo beneficial mutations and mutant lineages arising in a constant environment. *BMC Biol.* **19**, (2021).
214. Lang, G. I. & Desai, M. M. The spectrum of adaptive mutations in experimental evolution. *Genomics* **104**, (2014).
215. Park, M. K. Enhancing recombinant protein production with an *Escherichia coli* host strain lacking insertion sequences. *Appl. Microbiol. Biotechnol.* **98**, (2014).
216. Marisch, K., Bayer, K., Cserjan-Puschmann, M., Luchner, M. & Striedner, G. Evaluation of three

industrial *Escherichia coli* strains in fed-batch cultivations during high-level SOD protein production. *Microb. Cell Fact.* **12**, (2013).

217. Lee, J. H. Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. *Microb. Cell Fact.* **8**, (2009).
218. Jennie R. Lill William R. Mathews, C. M. R. & Schirle, M. Proteomics in the pharmaceutical and biotechnology industry: a look to the next decade. *Expert Rev. Proteomics* **18**, (2021).
219. Al-Amrani, S., Al-Jabri, Z., Al-Zaabi, A., Alshekaili, J. & Al-Khabori, M. Proteomics: Concepts and applications in human medicine. *World J. Biol. Chem.* **12**, (2021).
220. Diz, A. P., Martínez-Fernández, M. & Rolán-Alvarez, E. Proteomics in evolutionary ecology: linking the genotype with the phenotype. *Mol. Ecol.* **21**, (2012).
221. Johnson, D., Boyes, B., Fields, T., Kopkin, R. & Orlando, R. Optimization of data-dependent acquisition parameters for coupling high-speed separations with LC-MS/MS for protein identifications. *J. Biomol. Tech.* **24**, (2013).
222. Hu, A., Noble, W. S. & Wolf-Yadlin, A. Technical advances in proteomics: new developments in data-independent acquisition. *F1000Research* **5**, (2016).
223. Laemmli, U. K. SDS gel electrophoresis. *Nat.* **227**, (1970).
224. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, (1999).
225. Muth, T. The MetaProteomeAnalyzer: a powerful open-source software suite for metaproteomics data analysis and interpretation. *J. Proteome Res.* **14**, (2015).
226. Choi, M. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 6 (2014).
227. Achterholt, S., Priefert, H. & Steinbüchel, A. Purification and characterization of the coniferyl aldehyde dehydrogenase from *Pseudomonas* sp. Strain HR199 and molecular characterization of the gene. *J. Bacteriol.* **180**, (1998).
228. Sharma, U. K. & Chatterji, D. Transcriptional switching in *Escherichia coli* during stress and starvation by modulation of σ70 activity. *FEMS Microbiol. Rev.* **34**, (2010).
229. Hofmann, N., Wurm, R. & Wagner, R. The *E. coli* anti-sigma factor Rsd: studies on the specificity and regulation of its expression. *PLoS One* **6**, (2011).
230. Jelakovic, S., Jann, K. & Schulz, G. E. The three-dimensional structure of capsule-specific CMP: 2-keto-3-deoxy-manno-octonic acid synthetase from *Escherichia coli*. *FEBS Lett.* **391**, (1996).
231. Anderson, E. M., Shaji Saji, N., Anderson, A. C., Brewer, D., Clarke, A. J., & Khursigara, C. M. (

Pseudomonas aeruginosa Alters Peptidoglycan Composition under Nutrient Conditions Resembling Cystic Fibrosis Lung Infections. *mSystems* **7**, (2022).

232. Bucher, T., Oppenheimer-Shaanan, Y., Savidor, A., Bloom-Ackermann, Z. & Kolodkin-Gal, I. Disturbance of the bacterial cell wall specifically interferes with biofilm formation. *Environ. Microbiol. Rep.* **7**, (2015).
233. Riber, L. & Hansen, L. H. Epigenetic Memories: The Hidden Drivers of Bacterial Persistence? *Trends Microbiol.* **29**, (2021).
234. Hamilton, J. P. Epigenetics: principles and practice. *Dig. Dis.* **29**, (2011).
235. Tate, P. H. & Bird, A. P. Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr. Opin. Genet. Dev.* **3**, (1993).
236. Buitrago, D. Impact of DNA methylation on 3D genome structure. *Nat. Commun.* **12**, (2021).
237. Malik, M., Bensaid, A., Rouviere-Yaniv, J. & Drlica, K. Histone-like Protein HU and Bacterial DNA Topology: Suppression of an HU Deficiency by Gyrase Mutations. *J. Mol. Biol.* **256**, (1996).
238. Pérez, A. Impact of methylation on the physical properties of DNA. *Biophys. J.* **102**, (2012).
239. Chen, P., Bandoy, D. J. D. & Weimer, B. C. Bacterial epigenomics: Epigenetics in the age of population genomics. *Pangenome Divers. Dyn. Evol. Genomes*, (2020)
240. Geier, G. E. & Modrich, P. Recognition sequence of the dam methylase of *Escherichia coli* K12 and mode of cleavage of Dpn I endonuclease. *J. Biol. Chem.* **254**, (1979).
241. Blyn, L. B., Braaten, B. A. & Low, D. A. Regulation of pap pilin phase variation by a mechanism involving differential dam methylation states. *EMBO J.* **9**, (1990).
242. Zamora, M., Ziegler, C. A., Freddolino, P. L. & Wolfe, A. J. A Thermosensitive, Phase-Variable Epigenetic Switch: pap Revisited. *Microbiol. Mol. Biol. Rev.* **84**, (2020).
243. van der Woude, M. W. & Bäumler, A. J. Phase and antigenic variation in bacteria. *Clin. Microbiol. Rev.* **17**, (2004).
244. Sohanpal, B. K., El-Labany, S., Lahooti, M., Plumbridge, J. A. & Blomfield, I. C. Integrated regulatory responses of *fimB* to N-acetylneuraminic (sialic) acid and GlcNAc in Escherichia coli K-12. *Proc. Natl. Acad. Sci. U. S. A.* **101**, (2004).
245. El-Labany, S., Sohanpal, B. K., Lahooti, M., Akerman, R. & Blomfield, I. C. Distant cis-active sequences and sialic acid control the expression of *fimB* in *Escherichia coli* K-12. *Mol. Microbiol.* **49**, 1109–1118 (2003).
246. Hale, W. B., van der Woude, M. W. & Low, D. A. Analysis of nonmethylated GATC sites in the *Escherichia coli* chromosome and identification of sites that are differentially methylated in

response to environmental stimuli. *J. Bacteriol.* **176**, 3438–3441 (1994).

247. Wang, M. X. & Church, G. M. A whole genome approach to in vivo DNA-protein interactions in *E. coli*. *Nature* **360**, (1992).

248. Sohail, A., Lieb, M., Dar, M. & Bhagwat, A. S. A gene required for very short patch repair in *Escherichia coli* is adjacent to the DNA cytosine methylase gene. *J. Bacteriol.* **172**, (1990).

249. Takahashi, N., Naito, Y., Handa, N. & Kobayashi, I. A DNA methyltransferase can protect the genome from postdisturbance attack by a restriction-modification gene complex. *J. Bacteriol.* **184**, (2002).

250. Lu, A.-L. B. T.-R. M. in L. S. Mismatch Repair, Long/Short Patch. in (Elsevier, 2017).

251. Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Res.* **18**, (2008).

252. Lieb, M. Bacterial genes *mutL*, *mutS*, and dcm participate in repair of mismatches at 5-methylcytosine sites. *J. Bacteriol.* **169**, (1987).

253. Kidwell, M. G. & Lisch, D. R. Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* **55**, (2001).

254. Vandecraen, J., Chandler, M., Aertsen, A. & Van Houdt, R. The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit. Rev. Microbiol.* **43**, (2017).

255. Weinreich, M. D. & Reznikoff, W. S. Fis plays a role in Tn5 and IS50 transposition. *J. Bacteriol.* **174**, (1992).

256. McCommas, S. A. & Syvanen, M. Temporal control of transposition in Tn5. *J. Bacteriol.* **170**, (1988).

257. Roberts, D., Hoopes, B. C., McClure, W. R. & Kleckner, N. IS10 transposition is regulated by DNA adenine methylation. *Cell* **43**, (1985).

258. Laird, P. W. The power and the promise of DNA methylation markers. *Nat. Rev. Cancer* **3**, (2003).

259. Yu, M. Base-resolution detection of N4-methylcytosine in genomic DNA using 4mC-Tet-assisted-bisulfite- sequencing. *Nucleic Acids Res.* **43**, (2015).

260. Dai, Q. Ultrafast bisulfite sequencing detection of 5-methylcytosine in DNA and RNA. *Nat. Biotechnol.* (2024)

261. Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **39**, (2021).

262. Malone, T., Blumenthal, R. M. & Cheng, X. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* **253**, (1995).

263. Cheng, X. & Roberts, R. J. AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res.* **29**, (2001).

264. Schluckebier, G., O'Gara, M., Saenger, W. & Cheng, X. Universal catalytic domain structure of AdoMet-dependent methyltransferases. *Journal of molecular biology* , (1995).

265. Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W. & Roberts, R. J. Crystal structure of the *HhaI* DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell* **74**, (1993).

266. Carraway, M., Youderian, P. & Marinus, M. G. Spontaneous mutations occur near dam recognition sites in a dam- *Escherichia coli* host. *Genetics* **116**, (1987).

267. Rodriguez-R, L. M. An ANI gap within bacterial species that advances the definitions of intra-species units. *MBio* **15**, (2024).

268. Konstantinidis, K. Sequence-discrete species for prokaryotes and other microbes: A historical perspective and pending issues. *mLife* **2**, (2023).

269. Morimoto, Y. Re-identification of strains deposited as *Pseudomonas aeruginosa*, *Pseudomonas fluorescens* and *Pseudomonas putida* in GenBank based on whole genome sequences. *Int. J. Syst. Evol. Microbiol.* **70**, (2020).

270. Harms, H., Smith, K.E.C., Wick, L.Y. Microorganism-Hydrophobic Compound Interactions. In: Timmis, K.N. (eds) Handbook of Hydrocarbon and Lipid Microbiology. *Springer, Berlin, Heidelberg*, (2010).

271. Ahmad, F., Zhu, D. & Sun, J. Bacterial chemotaxis: a way forward to aromatic compounds biodegradation. *Environ Sci Eur* **32**, (2020).

272. Knöppel, A., Näsvall, J. & Andersson, D. I. Evolution of Antibiotic Resistance without Antibiotic Exposure. *Antimicrob. Agents Chemother.* **61**, (2017).

273. LeBlanc, N. & Charles, T. C. Bacterial genome reductions: Tools, applications, and challenges. *Front. genome Ed.* **4**, (2022).

274. Lima-Mendez, G., Van Helden, J., Toussaint, A. & Leplae, R. Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics* **24**, (2008).

275. Arndt, D. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16-21 (2016).

276. Janoschek, L., Grozdev, L. & Berensmeier, S. Membrane-assisted extraction of monoterpenes: from in silico solvent screening towards biotechnological process application. *R. Soc. open Sci.* **5**, (2018).

277. Grozdev, L., Kaiser, J. & Berensmeier, S. One-Step Purification of Microbially Produced

Hydrophobic Terpenes via Process Chromatography. *Front. Bioeng. Biotechnol.* **7**, (2019).

Supplementary material

The supplementary material of this thesis can be found in the following link:

https://drive.google.com/file/d/1OLEr-sw1XJ5BFHGEbng5skRl2T7FNYQw/view?usp=sharing