



**T.C.  
İSTANBUL ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**



**M. Sc. THESIS**

**IN SILICO CHARACTERIZATION OF MATE AND IAAL  
PROTEINS IN PSEUDOMONAS SYRINGAE AND  
PSEUDOMONAS SAVASTANOI**

**Fatima ALIHODŽIĆ**

**Moleküler Biyoloji ve Genetik Anabilim Dalı**

**Moleküler Biyoteknoloji ve Genetik Programı**

**Supervisor**

**Dr.Öğr.Üyesi Andrés Octavio ARAVENA DUARTE**

**Haziran, 2022**

**İSTANBUL**

Bu çalışma 27.06.2022 tarihinde ařağıdaki jüri tarafından Moleküler Biyoloji ve Genetik Anabilim Dalı Moleküler Biyoteknoloji ve Genetik Programında Yüksek Lisans Tezi olarak kabul edilmiştir.

### **Tez Jürisi**

Dr.Öğr.Üyesi Andrés Octavio ARAVENA DUARTE (Danışman)  
İstanbul Üniversitesi  
Fen Fakültesi

Doç.Dr. ÖZGÜR ÇAKIR  
İstanbul Üniversitesi  
Fen Fakültesi

Dr.Öğr.Üyesi PINAR PİR  
Gebze Teknik Üniversitesi  
Mühendislik Fakültesi



- **İntihal Programı Beyanı**

20.04.2016 tarihli resmi gazetede yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, İstanbul Üniversitesi'nin abonesi olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü'nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

- **Tezden Üretilmiş Yayınların Künye Bilgileri**

Alihodzic F., Aravena A. "How does MatE expression affect the toxicity of *Pseudomonas syringae* and *Pseudomonas savastanoi* in tomato and kiwi?" Proceedings of the 8th International Congress of the Molecular Biology Association of Turkey, MolBiyoKon2022 (Poster Presentation)

## PREFACE

*“Verily with hardship comes ease.” (94:5)*

Every journey into science is an adventure, embroidery of successes and failures. We finalize it with knowledge, experience, memories, and support from the dearest people. My gratitude and deepest respect are towards my parents for their endless support, my siblings, and R.

My special thanks and respect is to my mentor, Prof. Aravena, who helped me to find the way and end this adventure.

*“Zaista, s mukom je i olakšanje.” (94:5)*

Nijedan put kroz nauku nije lahak, ali uz podršku najbližih, krunišemo ga znanjem, iskustvom i uspomenama. Moja zahvalnost i najdublje poštovanje pripadaju roditeljima, brati, Tuljumbici, i R. Hvala vam!

Posebnu zahvalnost i poštovanje izražavam svom mentoru, prof. Araveni, koji mi je pomogao da pronađem svoj put i završim ovu avanturu.

Haziran, 2022

Fatima ALIHODŽIĆ

# CONTENTS

|                                                                                  | Page        |
|----------------------------------------------------------------------------------|-------------|
| <b>PREFACE</b> .....                                                             | <b>iv</b>   |
| <b>CONTENTS</b> .....                                                            | <b>viii</b> |
| <b>LIST OF FIGURES</b> .....                                                     | <b>xi</b>   |
| <b>LIST OF TABLES</b> .....                                                      | <b>xii</b>  |
| <b>LIST OF ABBREVIATIONS</b> .....                                               | <b>xiii</b> |
| <b>ÖZET</b> .....                                                                | <b>xvii</b> |
| <b>SUMMARY</b> .....                                                             | <b>xix</b>  |
| <b>1. INTRODUCTION</b> .....                                                     | <b>1</b>    |
| 1.1. AUXINS .....                                                                | 1           |
| 1.2. <i>PSEUDOMONAS</i> GENUS .....                                              | 2           |
| 1.3. <i>PSEUDOMONAS SYRINGAE</i> .....                                           | 2           |
| 1.4. INFECTION MECHANISM .....                                                   | 3           |
| 1.5. SPECIFIC PATHOVARS .....                                                    | 7           |
| 1.6. <i>PSEUDOMONAS SYRINGAE</i> PV. <i>ACTINIDIAE</i> .....                     | 8           |
| 1.7. INFECTION .....                                                             | 8           |
| 1.8. <i>PSEUDOMONAS SYRINGAE</i> PV. <i>TOMATO</i> DC3000 .....                  | 10          |
| 1.9. <i>PSEUDOMONAS SAVASTANOI</i> PV. <i>SAVASTANOI</i> .....                   | 12          |
| 1.10. IAA : INDOLE-3-ACETIC ACID .....                                           | 14          |
| 1.11. INTERACTION OF PLANT AND BACTERIA .....                                    | 18          |
| 1.12. TYPE 3 SECRETION SYSTEM - T3SS .....                                       | 19          |
| 1.13. MATE PROTEIN FAMILY, <i>MATE</i> AND <i>IAAL</i> GENES, MATE PROTEIN ..... | 21          |
| <b>2. METHODS</b> .....                                                          | <b>29</b>   |

|                                                                                                       |           |
|-------------------------------------------------------------------------------------------------------|-----------|
| 2.1. RETRIEVING THE <i>MATE</i> AND <i>IAAL</i> GENE AND CORRESPONDING PROTEIN SEQUENCES .....        | 29        |
| 2.2. MULTIPLE SEQUENCE ALIGNMENT (MSA) .....                                                          | 31        |
| 2.3. PHYLOGENETIC TREE CONSTRUCTION .....                                                             | 32        |
| 2.4. 3D STRUCTURE PROTEIN PREDICTION AND VALIDATION .....                                             | 33        |
| 2.5. IDENTIFICATION OF PROTEIN DOMAINS .....                                                          | 35        |
| 2.6. PREDICTION OF THE INTERACTOME .....                                                              | 35        |
| 2.7. SUBCELLULAR LOCALIZATION OF <i>MATE</i> AND <i>IAAL</i> PROTEINS .....                           | 36        |
| 2.8. PREDICTION OF DOCKING SITES .....                                                                | 36        |
| 2.9. RAMACHANDRAN PLOT AND QMEAN QUALITY ASSESSMENT .....                                             | 37        |
| 2.10.MSA AND PHYLOGENETIC TREE COMPARISON BETWEEN MORE DIFFERENT BACTERIAL STRAINS .....              | 38        |
| 2.11.CONSERVATION ANALYSIS OF THE <i>IAAL</i> AND <i>MATE</i> INTERGENIC AND PROMOTER SEQUENCES ..... | 39        |
| <b>3. RESULTS .....</b>                                                                               | <b>41</b> |
| 3.1. COMPARISON BETWEEN PROTEINS .....                                                                | 41        |
| 3.2. <i>IAAL</i> PROTEINS .....                                                                       | 43        |
| 3.2.1. Full Sequences of the Proteins .....                                                           | 43        |
| 3.2.2. Multiple Sequence Alignment .....                                                              | 45        |
| 3.2.3. MSA Results of <i>IaaL</i> Proteins .....                                                      | 46        |
| 3.2.4. Identity Matrix .....                                                                          | 53        |
| 3.2.5. Phylogenetic Tree .....                                                                        | 54        |
| 3.2.6. 3D Structure Prediction .....                                                                  | 57        |
| 3.3. <i>MATE</i> .....                                                                                | 60        |
| 3.3.1. Full Sequences of the Proteins .....                                                           | 60        |
| 3.3.2. MSA Results for <i>matE</i> .....                                                              | 62        |
| 3.3.3. Identity Matrix .....                                                                          | 66        |
| 3.3.4. Phylogenetic Tree .....                                                                        | 69        |

|                                                                                    |            |
|------------------------------------------------------------------------------------|------------|
| 3.3.5. 3D Structure Protein Prediction and Validation, SWISS-MODEL and Phyre2      | 71         |
| 3.3.6. Predicted Protein Domains                                                   | 81         |
| 3.3.7. Prediction of Docking Sites                                                 | 84         |
| 3.4. PREDICTION OF THE INTERACTOME                                                 | 84         |
| 3.5. SUBCELLULAR LOCALIZATION OF MATE AND IAAL PROTEINS                            | 87         |
| 3.6. MSA AND PHYLOGENETIC TREE COMPARISON BETWEEN MORE DIFFERENT BACTERIAL STRAINS | 88         |
| 3.7. INTERGENIC AND PROMOTER SEQUENCES CONSERVATION ANALYSIS                       | 88         |
| <b>4. DISCUSSION</b>                                                               | <b>93</b>  |
| 4.1. BLAST TOOL SEARCH                                                             | 93         |
| 4.2. MULTIPLE SEQUENCE ALIGNMENT                                                   | 96         |
| 4.3. PERCENT IDENTITY MATRIX                                                       | 100        |
| 4.4. PHYLOGENETIC TREE                                                             | 101        |
| 4.5. 3D PROTEIN PREDICTION                                                         | 103        |
| 4.6. PROTEIN DOMAINS IDENTIFICATION                                                | 111        |
| 4.7. INTERACTOME ANALYSIS                                                          | 112        |
| 4.8. SUBCELLULAR LOCALIZATION ANALYSIS                                             | 113        |
| 4.9. DOCKING SITE PREDICTION                                                       | 114        |
| 4.10.MSA BETWEEN MORE DIFFERENT SPECIES                                            | 115        |
| 4.11.INTERGENIC AND PROMOTER SEQUENCES CONSERVATION ANALYSIS                       | 115        |
| 4.12.CONCLUSION                                                                    | 116        |
| <b>REFERENCES</b>                                                                  | <b>123</b> |
| <b>APENDICES</b>                                                                   | <b>130</b> |
| APPENDIX 1 Multiple Sequence Alignment for IaaL Proteins                           | 131        |
| APPENDIX 2 Multiple Sequence Alignment for MatE Proteins                           | 133        |
| APPENDIX 3 T-coffee Multiple Sequence Alignment for IaaL Proteins                  | 137        |
| APPENDIX 4 T-coffee MSA for MatE Proteins                                          | 141        |

|                                                                                                                   |            |
|-------------------------------------------------------------------------------------------------------------------|------------|
| APPENDIX 5 COBALT Analysis for IaaL Proteins .....                                                                | 144        |
| APPENDIX 6 COBALT Analysis for MatE Proteins .....                                                                | 147        |
| APPENDIX 7 SMART Table for MatE Domains .....                                                                     | 149        |
| APPENDIX 8 Intergenic sequence of the MatE and IaaL protein .....                                                 | 150        |
| APPENDIX 9 Promoter sequence of MatE and IaaL .....                                                               | 163        |
| APPENDIX 10 Hit table best scoring BLAST results for IaaL proteins with accession<br>numbers and code names ..... | 176        |
| APPENDIX 11 Hit table BLASTP for MatE and IaaL proteins .....                                                     | 176        |
| APPENDIX 12 Phylogenetic tree for IaaL proteins accession numbers and descriptions                                | 177        |
| APPENDIX 13 Phylogenetic tree for the MatE proteins accession numbers and de-<br>scriptions: .....                | 178        |
| APPENDIX 14 Phylogenetic tree for the MatE proteins accession numbers and de-<br>scriptions extended tree: .....  | 179        |
| <b>CURRICULUM VITAE .....</b>                                                                                     | <b>180</b> |

## LIST OF FIGURES

|                                                                                                                                                                                                                                                                                           | Page |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| <b>Fig. 1.1:</b> <i>Pseudomonas syringae</i> shown using scanning electron microscopy. ....                                                                                                                                                                                               | 3    |
| <b>Fig. 1.2:</b> Immune recognition and immune response to T3SS-expressing bacteria in susceptible and resistant plants. ....                                                                                                                                                             | 5    |
| <b>Fig. 1.3:</b> Distribution map of <i>Pseudomonas syringae</i> pv. <i>actinidiae</i> globally.....                                                                                                                                                                                      | 9    |
| <b>Fig. 1.4:</b> <i>Pseudomonas syringae</i> pv. <i>actinidiae</i> symptoms on a plant: lesions and yellow spots. The photo was taken from ( <i>Pseudomonas syringae</i> pv. <i>Tomato DC3000 bacterial spotting on a plant leaf.</i> , 2021). ....                                       | 10   |
| <b>Fig. 1.5:</b> <i>Pseudomonas syringae</i> pv. <i>tomato DC3000</i> bacterial spotting on a plant leaf. ....                                                                                                                                                                            | 12   |
| <b>Fig. 1.6:</b> The Shikimic Acid Pathway in plants. ....                                                                                                                                                                                                                                | 16   |
| <b>Fig. 1.7:</b> Indoleacetic acid (Auxin) biosynthesis: a simple two-step pathway of the tryptophan conversion into the indole-3-acetic acid in bacteria.....                                                                                                                            | 17   |
| <b>Fig. 1.8:</b> Schematic representation of the T3SS from the plant (A) and animal (B) pathogenic bacteria.....                                                                                                                                                                          | 22   |
| <b>Fig. 1.9:</b> Physical relationships between IAA metabolism, T3SS, and drug efflux in <i>P. savastanoi</i> pv. <i>nerii</i> strain Psn23. ....                                                                                                                                         | 27   |
| <b>Fig. 1.10:</b> Physical and functional relationships between IAA metabolism, T3SS, and drug efflux in <i>P. savastanoi</i> pv. <i>nerii</i> strain Psn23. ....                                                                                                                         | 28   |
| <b>Fig. 3.1:</b> ClustalW multiple sequence alignment of IaaL proteins from <i>P. syringae</i> pv. <i>actinidiae</i> , <i>P. syringae</i> pv. <i>tomato DC3000</i> , <i>P. savastanoi</i> pv. <i>savastanoi</i> , and additional sequences of interest. ....                              | 47   |
| <b>Fig. 3.2:</b> ClustalW analysis of the IaaL proteins from <i>Pseudomonas</i> species without outgroup organisms. ....                                                                                                                                                                  | 48   |
| <b>Fig. 3.3:</b> COBALT analysis of multiple sequence alignment of IaaL proteins of interest from <i>P. syringae</i> pv. <i>actinidiae</i> , <i>P. syringae</i> pv. <i>tomato DC3000</i> , <i>P. savastanoi</i> pv. <i>savastanoi</i> , with additional proteins from other species. .... | 49   |
| <b>Fig. 3.4:</b> MSA by T-Coffee of <i>iaaL</i> gene-encoded IaaL proteins from <i>Pseudomonas</i> , <i>Erwinia</i> , <i>Arsenophonus</i> , and <i>Escherichia</i> species. ....                                                                                                          | 51   |
| <b>Fig. 3.5:</b> Phylogenetic tree for the IaaL proteins of <i>Pseudomonas</i> family including outgroups of <i>Arsenophonus</i> and <i>Erwinia</i> genus. ....                                                                                                                           | 55   |

|                                                                                                                                                                                                                                                                                                                                                    |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Fig. 3.6:</b> Phylogenetic tree of IaaL proteins from <i>Pseudomonas</i> family, including <i>Erwinia mallo-</i><br><i>tivora</i> as only outgroup.....                                                                                                                                                                                         | 56 |
| <b>Fig. 3.7:</b> Swiss-model and Ramachandran plot of the Indoleacetic acid lysine ligase 3D protein<br>from <i>Pseudomonas syringae</i> pv. <i>actinidiae</i> .....                                                                                                                                                                               | 59 |
| <b>Fig. 3.8:</b> ClustalW multiple protein sequence alignments for the MatE protein from <i>P. syringae</i><br>pv. <i>actinidiae</i> , <i>P. syringae</i> pv. <i>tomato</i> DC3000, <i>P. savastanoi</i> pv. <i>savastanoi</i> and <i>Glycine max</i><br>as a focus with additional <i>Pseudomonas</i> species. ....                               | 63 |
| <b>Fig. 3.9:</b> COBALT analysis for multiple sequence alignment of the proteins of interest for MatE<br>proteins from <i>P. syringae</i> pv. <i>actinidiae</i> , <i>P. syringae</i> pv. <i>tomato</i> DC3000, <i>P. savastanoi</i> pv.<br><i>savastanoi</i> as a focus protein, and additional proteins from the <i>Pseudomonas</i> species. .... | 64 |
| <b>Fig. 3.10:</b> MSA by T-Coffee of matE gene-encoded proteins from four species of interest, <i>P.</i><br><i>syringae</i> pv. <i>actinidiae</i> , <i>P. syringae</i> pv. <i>tomato</i> DC3000, <i>P. savastanoi</i> pv. <i>savastanoi</i> , and<br>additional proteins from the <i>Pseudomonas</i> species. ....                                 | 65 |
| <b>Fig. 3.11:</b> Phylogenetic tree of MatE proteins from <i>Pseudomonas</i> family including <i>Glycine max</i> as<br>an outgroup organism.....                                                                                                                                                                                                   | 69 |
| <b>Fig. 3.12:</b> Phylogenetic tree of MatE proteins in <i>Pseudomonas</i> family.....                                                                                                                                                                                                                                                             | 70 |
| <b>Fig. 3.13:</b> SWISS-MODEL results for the MatE protein of <i>Pseudomonas syringae</i> pv. <i>actinidiae</i><br>including query sequence identity with the template sequence and coverage of the alignment. 72                                                                                                                                  |    |
| <b>Fig. 3.14:</b> SWISS-MODEL of the MatE protein model-template model of <i>Pseudomonas syringae</i> . 73                                                                                                                                                                                                                                         |    |
| <b>Fig. 3.15:</b> MatE 3D protein shape and its alpha-helical structure of <i>Pseudomonas syringae</i> pv.<br><i>actinidiae</i> .....                                                                                                                                                                                                              | 73 |
| <b>Fig. 3.16:</b> PHYRE2 3D model of the MatE protein sequence WP_003380687.1 of <i>Pseudomonas</i><br><i>syringae</i> pv. <i>actinidiae</i> . ....                                                                                                                                                                                                | 74 |
| <b>Fig. 3.17:</b> Aligned secondary structure of MatE protein sequence WP_003380687.1 of <i>Pseu-</i><br><i>domonas syringae</i> pv. <i>actinidiae</i> , consisting of the alpha helices, and the primary protein<br>structure. ....                                                                                                               | 75 |
| <b>Fig. 3.18:</b> Aligned secondary structure of the MatE protein WP_005782319.1 of <i>Pseudomonas</i><br><i>savastanoi</i> pv. <i>savastanoi</i> with red encircled residues different from two other proteins aligned<br>from the same species. ....                                                                                             | 76 |
| <b>Fig. 3.19:</b> Aligned secondary structure of MatE protein WP_054995848.1 of <i>Pseudomonas sy-</i><br><i>ringae</i> pv. <i>tomato</i> DC3000 with red encircled residues different from two other proteins<br>aligned from the same species. ....                                                                                              | 77 |
| <b>Fig. 3.20:</b> Aligned secondary structure of MatE protein WP_003380687.1 of <i>Pseudomonas sy-</i><br><i>ringae</i> pv. <i>actinidiae</i> with red encircled residues different from two other proteins aligned<br>from the same species. ....                                                                                                 | 78 |

|                                                                                                                                                                                   |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Fig. 3.21:</b> PHYRE2 3D model and Ramachandran plot of the MatE protein of <i>Pseudomonas syringae</i> pv. <i>actinidiae</i> based on the template 6idp.1.A. ....             | 79 |
| <b>Fig. 3.22:</b> PHYRE2 3D model of the MatE protein of <i>Pseudomonas syringae</i> pv. <i>actinidiae</i> with emphasis on the conservation level of the sequence. ....          | 79 |
| <b>Fig. 3.23:</b> Ramachandran plot and 3D protein predictions of MatE protein of <i>Pseudomonas syringae</i> pv. <i>actinidiae</i> . ....                                        | 80 |
| <b>Fig. 3.24:</b> Ramachandran plot and 3D protein predictions of MatE protein of <i>Pseudomonas syringae</i> pv. <i>tomato</i> DC3000. ....                                      | 80 |
| <b>Fig. 3.25:</b> Ramachandran plots and 3D protein predictions of MatE protein of <i>Pseudomonas savastanoi</i> pv. <i>savastanoi</i> . ....                                     | 81 |
| <b>Fig. 3.26:</b> Pfam analysis of MatE domains of <i>Pseudomonas syringae</i> pv. <i>actinidiae</i> . ....                                                                       | 82 |
| <b>Fig. 3.27:</b> Pfam results of protein domain identification of MatE protein sequence of <i>Pseudomonas syringae</i> pv. <i>actinidiae</i> . ....                              | 82 |
| <b>Fig. 3.28:</b> SMART analysis of protein domain identification of MatE protein of <i>Pseudomonas syringae</i> pv. <i>actinidiae</i> . ....                                     | 82 |
| <b>Fig. 3.29:</b> SMART analysis of protein domains of 3 MatE query proteins of interest with a graphical representation of different domains and regions they are found at. .... | 83 |
| <b>Fig. 3.30:</b> Interaction of the ligand “OLC” and the MatE protein of <i>P. syringae</i> pv. <i>actinidiae</i> . ....                                                         | 84 |
| <b>Fig. 3.31:</b> Closer look at the protein and ligand interaction of the MatE protein of the <i>Pseudomonas syringae</i> pv. <i>actinidiae</i> with the ligand “OLC”. ....      | 85 |
| <b>Fig. 3.32:</b> Tabular representation of the ligand details which interact with the MatE protein of the <i>P. syringae</i> pv. <i>actinidiae</i> . ....                        | 85 |
| <b>Fig. 3.33:</b> Interactome analysis of the MatE protein (PSPTO_0370) of <i>P. syringae</i> pv. <i>actinidiae</i> by STRING. ....                                               | 86 |
| <b>Fig. 3.34:</b> Wider interactome analysis of the MATE efflux family protein (PSPTO_0370) of <i>P. syringae</i> pv. <i>actinidiae</i> by STRING. ....                           | 86 |
| <b>Fig. 3.35:</b> MSA of the various MatE efflux proteins and Indoleacetic-lysine ligase proteins by ClustalW server. ....                                                        | 89 |
| <b>Fig. 3.36:</b> ClustalW analysis of conservation level of the intergenic sequences between <i>matE</i> and <i>iaaL</i> genes in different organisms. ....                      | 90 |
| <b>Fig. 3.37:</b> ClustalW conservation analysis of the promoter sequence of <i>matE</i> and <i>iaaL</i> genes in different organisms. ....                                       | 90 |

## LIST OF TABLES

|                                                                                                                                                                                                                                                      | Page |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| <b>Table 2.1:</b> MatE and IaaL protein accession numbers from three <i>Pseudomonas</i> species of interest.                                                                                                                                         | 39   |
| <b>Table 3.1:</b> Results of comparing protein sequences of IaaL as found in <i>Pseudomonas syringae</i> pv. <i>actinidiae</i> versus <i>Pseudomonas syringae</i> pv. <i>tomato</i> DC3000 and <i>Pseudomonas savastanoi</i> pv. <i>savastanoi</i> . | 42   |
| <b>Table 3.2:</b> Results of comparing protein sequences of MatE as found in <i>Pseudomonas syringae</i> pv. <i>actinidiae</i> versus <i>Pseudomonas syringae</i> pv. <i>tomato</i> DC3000 and <i>Pseudomonas savastanoi</i> pv. <i>savastanoi</i> . | 42   |
| <b>Table 3.3:</b> Protein homology of partial and full sequences of IaaL proteins.                                                                                                                                                                   | 44   |
| <b>Table 3.4:</b> Accessions for the MSA and T-coffee for IaaL and MatE proteins.                                                                                                                                                                    | 52   |
| <b>Table 3.5:</b> Percent Identity Matrix table for proteins encoded by <i>iaaL</i> genes by ClustalW.                                                                                                                                               | 53   |
| <b>Table 3.6:</b> Extended Identity Matrix table for IaaL proteins by ClustalW.                                                                                                                                                                      | 54   |
| <b>Table 3.7:</b> Code names of the IaaL proteins for the T-coffee and ClustalW analysis.                                                                                                                                                            | 57   |
| <b>Table 3.8:</b> Protein homology of partial and full sequences of MatE proteins.                                                                                                                                                                   | 61   |
| <b>Table 3.9:</b> Identity Matrix table for proteins encoded by <i>matE</i> genes by ClustalW.                                                                                                                                                       | 67   |
| <b>Table 3.10:</b> Extended Identity Matrix table for MatE proteins by ClustalW.                                                                                                                                                                     | 68   |
| <b>Table 3.11:</b> Code names of the MatE proteins for the T-coffee and ClustalW analysis.                                                                                                                                                           | 71   |
| <b>Table 3.12:</b> Subcellular localization of the proteins of interest based on two-server analysis and available scores and accuracy for better comparison and analysis.                                                                           | 87   |
| <b>Table 7.1:</b> <i>P. syringae</i> pv. <i>actinidiae</i>                                                                                                                                                                                           | 149  |
| <b>Table 7.2:</b> <i>P. syringae</i> pv. <i>tomato</i>                                                                                                                                                                                               | 149  |
| <b>Table 7.3:</b> <i>P. savastanoi</i> pv. <i>savastanoi</i>                                                                                                                                                                                         | 149  |

## LIST OF ABBREVIATIONS

| Abbreviations    | Description                                             |
|------------------|---------------------------------------------------------|
| <b>Psa</b>       | : <i>Pseudomonas syringae</i> pv. <i>actinidiae</i>     |
| <b>Psav</b>      | : <i>Pseudomonas syringae</i> pv. <i>savastanoi</i>     |
| <b>PsPto</b>     | : <i>Pseudomonas syringae</i> pv. <i>tomato</i> DC3000  |
| <b>MatE</b>      | : Multidrug and toxic compound extrusion protein        |
| <b>IaaL</b>      | : Indoleacetic acid lysine ligase                       |
| <b>IAA</b>       | : Indole-3-acetic acid                                  |
| <b>T3SS/TTSS</b> | : Type 3 Secretion System                               |
| <b>T3SS5</b>     | : Type 3 Secretion System 5                             |
| <b>hrp</b>       | : Hypersensitive response and pathogenicity             |
| <b>hrc</b>       | : Hypersensitive response and conserved                 |
| <b>PAMPs</b>     | : Pathogen-associated molecular patterns                |
| <b>MAMPs</b>     | : Microbe-associated molecular patterns                 |
| <b>PTI</b>       | : Pattern-triggered immunity                            |
| <b>MDR</b>       | : Multidrug resistance                                  |
| <b>Trp</b>       | : Tryptophan                                            |
| <b>IAM</b>       | : Indole-3-acetamide                                    |
| <b>IaaH</b>      | : Tryptophan-2-hydrolase                                |
| <b>IaaM</b>      | : Tryptophan-2-monooxygenase                            |
| <b>IPyA</b>      | : Indole-3-pyruvic acid                                 |
| <b>PAIs</b>      | : Pathogenicity islands                                 |
| <b>MATE</b>      | : Multidrug and toxic compound extrusion protein family |
| <b>MSA</b>       | : Multiple Sequence Alignment                           |
| <b>PDB</b>       | : Protein Data Bank                                     |
| <b>DIP</b>       | : Database of Interacting Proteins                      |

## ÖZET

### YÜKSEK LİSANS TEZİ

#### PSEUDOMONAS SYRINGAE VE PSEUDOMONAS SAVASTANOI'DEKİ MATE VE IAAAL PROTEİNLERİNİN SİLİKO KARAKTERİZASYONU

Fatima ALIHODŽIĆ

İstanbul Üniversitesi

Fen Bilimleri Enstitüsü

Moleküler Biyoloji ve Genetik Anabilim Dalı

Danışman: Dr.Öğr.Üyesi Andrés Octavio ARAVENA DUARTE

*Pseudomonas* cinsinden (genusten) üç bakteri türü vardır: *Pseudomonas syringae* pv. *actinodiae*, *Pseudomonas syringae* pv. *tomato* DC3000 ve *Pseudomonas savastanoi* pv. *savastanoi* hastalığa neden olan ajanlar olarak tanınmıştır.

Son analizler, *matE* ve *iaaL* genlerinin ve karşılık gelen MatE ve IaaL proteinlerinin virülans ve hastalık oluşumu üzerinde yüksek bir korelasyon ve etkisini göstermektedir. Bu genler ve proteinler, bakterilerin virülans faktörlerinin bitki hücrelerine iletilmesinden sorumlu olan Tip 3 Salgı Sistemi bakteri makinesinin bir parçasıdır ve bitkideki hastalık sürecinde rol oynayan oksin hormonu üretimi ile bağlantılıdır.

Oksin hormonu, hücresel düzeyde organ gelişiminde ve tüm bitki işleyişinden bitki büyüme ve gelişmesine kadar rol oynar. Gövdelerde hücre büyümesi ve uzamasının düzenlenmesinde, hücre bölünmesinde ve farklılaşmasında, meyve gelişiminde, kök oluşumunda ve yaprak dökümünde (dökülmede) görev alır (Kunkel and Harper 2017). Oksin olmadan bitkiler büyüyemez ve gelişemez. Aynı şekilde yüksek miktarda oksin, bitki için ölümcül olabilir.

Bakteriler de hayatta kalabilmek için gerekli olan oksinleri üretirler. Birkaç farklı oksin türü vardır. Bitki ve bakteriler tarafından üretilen en yaygın tür indol-3-asetik asittir (IAA). IAA, bakteriyel toksisite ve ilaç savunması ile ilgilidir. IAA ayrıca toksinlerin bakteriyel atılımını destekler, çoklu ilaç ihracatçı genlerin ekspresyonunu tetikler ve plazmit kopyalarının korunmasına katkıda bulunur. Ayrıca IAA, bakterilerde metabolik kontrolde aktif bir sinyal görevi görür. IAA sentezinin öncüsü olan indol, amino asit metabolizmasında yer alan genleri aktive edebilir, ancak bunun tam etki mekanizması bilinmemektedir (Hu et al. 2010).

Bitkideki oksin miktarı yükseldiğinde oksin, bitkinin hastalık gelişiminde bir virülans faktörü olarak karakterize edilir. Bakteriler bitkilerle etkileşime girerek ve onların oksin miktarını değiştirerek herhangi bir normal oksin sürecine müdahale etme potansiyeline sahiptir. Eksojen oksinin bitki gelişimi üzerindeki etkisi, pozitif etkiden negatife değişir. Bitki için sonuç genellikle (1) bitki için mevcut üretilen IAA miktarının ve (2) bitki dokusunun IAA konsantrasyonundaki değişikliklere duyarlılığının bir fonksiyonudur (Spaepen, Vanderleyden, and Remans 2007). Bu nedenle, enfeksiyon sırasında bitkide tüm farklı süreçler meydana gelir ve kaçınılmaz olarak bitki IAA biyosentez mekanizmasının gerçekleşme şeklini değiştirir. Bakteriler tarafından üretilen oksinin bitki hücrelerine mi girdiği yoksa bakterilerin bitkideki oksin üretimini toksik düzeye kadar uyarıp uyarmadığı hususu henüz bilinmemektedir (Zhao 2010).

Bir bitkiyi kolonize etmek için bakteriler, bitki hücre zarı boyunca efektör moleküller ve virülans faktörleri iletmek için bir taşıma sistemi kullanır. Taşıma sistemi bakteri kullanımı Tip 3 Salgı Sistemidir (T3SS), *matE* ve *iaaL* genlerinin bulunduğu aynı operon olan HrpL operonundan kodlanır. *matE* geni, T3SS mekanizmasının bir parçası olan MatE proteinini kodlar ve *iaaL* geni, IAA'nın sentezinde rol oynar. Bu şekilde, IAA, T3SS, MatE proteinleri ve genler arasındaki ilişki kurulabilir ve genel olarak bakterilerin domates ve kivi üzerindeki toksisitesini etkiler.

MatE proteini, efektör moleküllerin bakterinin iç zarından geçiş mekanizmasının iğne ve pilusuna geçişini sağlayan membran taşıma mekanizmasının ana parçalarından biridir. *Pseudomonas syringae* pv. *actinidiae*'nin MatE proteininin karakterizasyonu, alanları, etkileşim bölgesi, öngörülen protein 3D modeli ve diğer analizlerin yanı sıra promotör ve intergenik dizilerin analizi, proteinin toksisite indüksiyonundaki olası rolünü belirlememize yardımcı olmuştur.

BLAST aracıyla, ilgili üç MatE ve üç IaaL proteininin tam dizilerini ve bu proteinlerle diğer organizmalardan tüm homolog dizileri elde ettik. Tam protein dizileri, bulduğumuz ilk diziler eksik ve dolayısıyla alakasız olduğundan araştırmamızın anahtarıydı. MatE proteininin *Archaea* ve *Eukarya* gibi farklı türlerde de bulunduğunu görmek ilginçti. IaaL proteini için aynı arama sonuçlarını aldık.

Çoklu dizi hizalaması (MSA), analiz ettiğimiz proteinler arasındaki kalıntı polimorfizm bölgelerini ve mutasyonları gösterdi. Sonuçlar bize proteinler arasındaki benzersiz kalıntı farklarını gösterdi, bu da aynı olmayan dizilerle çalıştığımızın kanıtıydı. Organizmalar ve türler arasındaki koruma düzeyi yüksekti, ancak %100 değildi. MSA, üç farklı sunucu tarafından yapıldı ve elde ettiğimiz sonuçlar aynıydı. Filogenetik ağaç, MSA sonuçlarını gösteren, analiz edilen proteinlerin mesafesini gösterir.

Filogenetik ağaç, MatE proteinlerini evrimsel mesafeye göre üç gruba ayırdı. Odak noktamız *P. syringae* pv. *actinidiae*'nin MatE proteiniydi ve farklı organizmalardan üç MatE proteininin 3D protein modellerini yapmaya karar verdik. Dizileri karşılaştırmak için filogenetik ağaç tarafından ayrılan gruplardan dallanma mesafelerine göre üç protein seçtik.

*Pseudomonas syringae* pv. *actinidiae*'nin MatE proteininin 3 boyutlu yapısı, protein model yapısının en olası şeklini sergiledi. İki farklı sunucu tarafından bir 3D model yaptık ve her iki sunucuda mevcut en iyi şablon *Vibrio cholerae*'dandı. Bu şablon, proteinimizle yalnızca %21 özdeşlik eşleşmesine sahiptir, ancak veri bankasındaki mevcut en iyi şablondur. Bu analiz sayesinde, ek araştırmalara ve daha iyi şablon bankalarına büyük ihtiyaç olduğunu gördük. Elde ettiğimiz sonuçlar %100 doğru modelin temsili olamaz, ancak bu araştırmada

elde edebileceğimiz en yakın model yapılarıdır. Yaptığımız 3D model, döngülerle birbirine bağlanan alfa sarmallara ve tüp benzeri bir yapıya sahiptir. MatE proteininde beta zinciri yoktur.

QMEAN değeri, değerli veri aralığında ve kapsama alanı yüksek, %95'ten fazla. *P. syringae* pv. *tomato* DC3000 ve *P. savastanoi* pv. *savastanoi*'den elde edilen MatE proteinlerinin 3D modelleri, tüp benzeri yapıdaki halkalarla birbirine bağlanan sarmallardan yapılmış, görsel olarak benzer bir protein şekline sahipti. *P. savastanoi* modeli tahmin edildiğinden ve AlphaFold Protein Veritabanında mevcut olduğundan, bize değerli bilgiler ve ilgilendiğimiz proteini onunla ilişkilendirme şansı verir. Yapılar oldukça benzerdir, bu nedenle genel işlev ve etki arasındaki paralellik de çizilebilir.

Döngü bölgelerinde bu üç MatE proteininin ikincil yapılarında ilginç değişiklikler bulduk. Bu üç protein karşılaştırıldığında ilmek bölgelerinde polimorfizmler, mutasyonlar ve delesyonlar mevcuttur. Bu önemlidir çünkü ilmek bölgeleri sarmallar kadar sabit değildir, bu nedenle ilmek kalıntı yapısındaki değişiklikler protein katlanma modelini, protein etkileşimini ve genel davranışı etkileyebilir. Proteinin katlanma ve etkileşime girme yeteneği veya komple taşıma makinesinin faaliyetini artırma yeteneği bu şekilde etkilenebilir. Bu nedenle, bulduğumuz değişiklikler şekil 3.18, 3.19 ve 3.20'da kırmızı ile çevrelenmiştir.

Ayrıca, protein alanlarını belirlemek önemliydi. Bu işlevsel birimler, proteinin genel işlevini belirleyebilir ve organizmalar arasında korunduğu tespit edilirse, ilgilendiğimiz proteinin hastalık gelişimindeki tam işlevini ve etkisini gösterebilir. İki farklı araç tarafından tanımlanan iki "MatE" alanı bulduk. Etki alanları 200 kalıntıdan daha kısaydı. Her üç *Pseudomonas* türünde de aynı alanlar bulundu ve şekil 3.29'de grafiksel olarak sunuldu.

Proteinlerin etkileşen bölgeleri, proteini aktive edebilen/deaktif edebilen veya onu birçok başka şekilde etkileyebilen ligandlarla protein etkileşiminin açık bölgeleridir. İlgilendiğimiz MatE proteinimiz ile etkileşime girdiğini tespit ettiğimiz "OLC" ligandı, birkaç zincire ve basit bir yapıya sahiptir. Ek ligand bulunamadı. Bu ligand, MatE proteininin tüp benzeri yapı oluşumunun iç kısmına bağlanır.

MatE proteininin yalnızca bir etkileşimli ligandı olmasına rağmen, genel protein etkileşimlerini kontrol etmemiz gerekiyordu. STRING etkileşim veritabanlarına göre, *P. syringae* pv. *actinidiae*'nin MatE proteini, yalnızca iki protein, IaaL ve *P. syringae* pv. *tomato* DC3000'den dış zar porin proteini ile etkileşime girer. Aynı zamanda, IaaL proteini, geniş bir etkileşimli protein ağına sahiptir. Nedeni bu iki proteinin pozisyonu olabilir. MatE proteini, hücre zarına tamamen gömülü olan iç zar proteini olarak öne çıkarken, IaaL sitoplazmik bir proteindir. Bu nedenle, IaaL proteini çok çeşitli çevreleyen proteinlere ve diğer proteinlerle etkileşime girebileceği hücreler arası boşluğa sahiptir. Hücre altı yerleştirme analizi iki farklı sonucu tarafından gerçekleştirilir.

Daha farklı türler arasındaki ve MatE ve IaaL proteini için olan filogenetik ağaç, analiz ettiğimiz proteinlerin benzersiz yapısına ve aralarındaki büyük farklılıklara ilişkin bulgularımızı tasvir etti. Analizde ek kontroller olarak ve MatE proteininin türler arasında korunup korunmadığını görmek için dış grup organizmaları kullanıldı. MatE proteini diğer türlerde de bulunabilse de, hastalığa neden olan ana ajan olduğu için asıl odak noktamız bakterilerdi.

Son analizimiz proteinler değil, *matE* ve *iaaL* gen dizileri üzerindiydi. Protein analizi ile tanımlayamadığımız nükleotid seviyesinde bazı ek farklılıklar olup olmadığını kontrol etmemiz gerekti. İlk olarak, *Pseudomonas* türlerinin çoklu organizmaları için intergenik bölgeyi

kontrol ettik. Sonuçlar, bu bölgenin polimorfizmler, boşluklar veya mutasyonlar olmadan yüksek oranda korunduğunu gösterdi. Dizinin intergenik kısmı stabildir ve farklı organizmalarda aynıdır. Öte yandan *matE* ve *iaaL* genlerinin promotör bölgesinde polimorfizmler bulduk. Bazı organizmalarda promotör bölgesinin nükleotid yapısındaki değişiklikler, arttırılmış veya baskılanmış promotör aktivitesinin nedeni olabilir. Bu özellikle önemli olabilir, çünkü *iaaL* ve *matE* genlerinin aynı promotörden kopyalandığını ve bunlardan biri baskılandığında diğerinin de etkin olmadığını biliyoruz.

Elde ettiğimiz sonuçları düzenleyerek ve MatE proteininin karakterizasyonuna ve *Pseudomonas syringae* pv. *actinidiae*'nin domates ve kivi üzerindeki toksisitesindeki olası rolüne ilişkin araştırmanın amacını yansıtarak, aşağıdaki sonuç açıklamalarına sahibiz. İlk varsayımımız, MatE proteinlerinin farklı olacağı ve toksisitedeki rollerinin deşifre edilmesinin tam protein yapısına dayanabileceği idi. Bu varsayım, aldığımız sonuçlardan dolayı eksik ve yanlış gösterildi. Türler arasındaki MatE proteinlerinin dizileri oldukça benzerdir ve belirli bir dereceye kadar korunur, ancak dizideki polimorfizmler, türler arasındaki benzersiz özelliklerin ve filogenetik çeşitliliğin açık işaretleridir.

Ayrıca, promotör bölgesindeki ve özellikle *P. syringae* pv. *actinidiae*'nin MatE proteininin halka bölgelerindeki farklılıklar, protein işleyişindeki olası varyansı gösterir ve enfeksiyon sürecinde virülans faktörlerinin taşınmasını etkiler.

MatE proteini hücrenin iç zarına gömülü olduğundan, güçlendirici moleküller ve etkileşime girebileceği enzimler, bulduğumuz polimorfizmler nedeniyle farklı olabilir, bu nedenle bu bakteri ile enfekte olan bitkilerde toksisite derin ve daha yoğun olabilir.

MatE proteininin 3D modelini oluşturduğumuz şablon protein, sorgu dizimizle yalnızca %21 aynı olsa da, model olarak elde ettiğimiz şekil, diğer proteinlerin tahmin edilen modellerine görsel olarak oldukça benzer. Bu nedenle, tahmin edilen modelimizin analizine güvenebiliriz.

Halkalarla birbirine bağlanan ve tüp benzeri bir biçimde düzenlenen alfa sarmalları, yapılarındaki değişikliklerin virülans faktörlerinin ve efektörlerin bitki hücresine taşınmasını etkilediğinin güçlü bir göstergesidir. Protein dizisindeki kalıntı yapısındaki farklılıklar, MatE proteininin *Pseudomonas syringae* pv. *actinidiae*'deki geni tarafından kodlandığının güçlü bir göstergesidir, bitkilerde toksisite ve enfeksiyon sürecinde önemli bir role sahiptir. Aynı zamanda bu bakteri, domates ve kivide hastalığa neden olan mekanizmada yer alan benzersiz yapısal özelliklerin nihai kanıtının yanı sıra, protein tahmini için şablonlar olarak kristal form üzerinde daha fazla araştırmaya duyulan ihtiyacı göstermektedir.

Haziran 2022, 199 sayfa.

**Anahtar kelimeler:** *Pseudomonas syringae*, indol-3-asetik asit (IAA), *matE* geni, *iaaL* geni, MATE proteini, IaaL proteini, virülans, toksisite, biyoinformatik karakterizasyon

## SUMMARY

### M.Sc. THESIS

#### IN SILICO CHARACTERIZATION OF MATE AND IAAL PROTEINS IN PSEUDOMONAS SYRINGAE AND PSEUDOMONAS SAVASTANOI

**Fatima ALIHODŽIĆ**

**Istanbul University**

**Institute of Graduate Studies in Sciences**

**Department of Molecular Biology and Genetics**

**Supervisor: Assist.Prof. Andrés Octavio ARAVENA DUARTE**

Three bacterial species from the *Pseudomonas* genus, *Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* DC3000 and *Pseudomonas savastanoi* pv. *savastanoi*, are recognized as disease-causing agents. Recent analyses show a high correlation between both virulence and disease occurrence versus the expression of *matE* and *iaaL* genes and the corresponding MatE and IaaL proteins.

These proteins are part of the Type 3 Secretion System bacterial machinery responsible for the delivery of the virulence factors of bacteria into the plant cell and are connected with the auxin hormone production, which in turn has a role in the plant disease.

As MatE protein of *Pseudomonas syringae* pv. *actinidiae* is one of the main parts of the membrane transport machinery, its characterization helps us to determine the possible role of the protein in toxicity induction.

We used different tools for the aforementioned analyzes. Using BLAST, we identified the full sequences of homologous MatE and IaaL proteins in all species containing them. Multiple sequence alignment helped us to identify the conservation level amongst proteins, polymorphisms, and similarities. The phylogenetic tree confirmed the MSA analysis results. Our protein of interest from three species has a positive phylogenetic distance from other species, and between each other.

The 3D model of the MatE proteins from different species shows a tube-like structure made of alpha-helices interconnected by loops. The main differences between the studied proteins

are the residue sequence in their loops. We found deletions and polymorphisms in the loop regions that may cause different folding patterns in the MatE proteins from the three species we analyzed, especially in the *P. syringae* pv. *actinidiae*.

We found only one ligand, the “OLC” ligand interacting with MatE of *P. syringae* pv. *actinidiae*. It binds the inner part of the tube-like structure of MatE. According to STRING analysis, MatE interacts with only two other proteins, while IaaL has a wide network of interactions. All the MatE proteins have two MatE domains of the same size.

The intergenic region of *matE* genes is highly conserved amongst the species. We found differences in the promoter regions. Polymorphisms in the promoter region might be one extra indicator of the particular characteristics of these proteins in *Pseudomonas syringae* pv. *actinidiae* and their influence on its toxicity in tomatoes and kiwi.

Polymorphisms in the loop structures of MatE proteins and promoter regions might be the reason for the protein influence on the toxicity in the tomatoes and kiwi.

June 2022, 199 pages.

**Keywords:** *Pseudomonas syringae*, indole-3-acetic acid (IAA), MATE protein, IaaL protein, virulence, toxicity, bioinformatic characterization.

## 1. INTRODUCTION

This thesis focuses on the characterization of the *matE* and *iaaL* genes of *Pseudomonas syringae* pv. *actinidiae* (Psa), *Pseudomonas syringae* pv. *tomato*, and *Pseudomonas savastanoi* pv. *savastanoi*, and the proteins they encode: MatE and IaaL. Special attention is on the MatE protein in the Psa and the analysis of its possible effect on bacterial toxicity. These genes and proteins are part of the bacterial machinery responsible for the delivery of the virulence factors of bacteria into the plant cell and are connected with the auxin hormone production, which in turn has a role in the disease process in the plant. Plants in which these genes were correlated with the disease development are tomatoes and kiwi, so this thesis is based on the protein and gene analysis and characterization of the MatE protein for a better understanding of the disease.

The beginning of civilization was related to the first crop and animal domestication. Through the agricultural means of food production, the human race made a step forward in the development and built the foundations of what we know today as modern life.

The origins of crop diseases are linked to the domestication of the plants and their development through the centuries. Some plants were domesticated centuries ago, limiting our capability to reconstruct the mechanism and cause of disease emergence. Other plants were domesticated more recently and thus serve as a valuable tool for disease studies. One such example is the kiwifruit, whose domestication began in the 1930s, with canker disease outbreaks in the 1980s caused by the bacterium *Pseudomonas syringae* pv. *actinidiae* (Psa), bringing significant loss to the world food market.

### 1.1. AUXINS

To colonize a plant successfully, pathogens must modulate basal host defenses and alter the biology of their host. Plant hormones are among the most common ways bacteria alter the host signaling mechanisms. One of such hormones is auxin, which plays a role in plant growth and development, from the cellular level, organ development, and whole plant functioning as well. Zhao (2010) showed that auxin synthesis is elevated by the tryptophan-dependent pathway and therefore characterized as a virulence factor in plant disease development. But, the exact mechanism of the auxin elevation is not well understood. It is still unknown whether

the auxin produced by the bacteria enters the plant cells, or the bacteria stimulates the auxin production in the plant (Zhao, 2010).

Auxin can be synthesized by the tryptophan-dependent and tryptophan independent pathways. Plants and bacteria, including *Pseudomonas* species, use a tryptophan-dependent pathway. An overabundance of auxins can be toxic to the plant and thus it is characterized as a virulence factor (Zhao, 2014).

## **1.2. PSEUDOMONAS GENUS**

*Pseudomonas* is a genus of Bacteria (phylum: Proteobacteria, class: Gammaproteobacteria, order: Pseudomonadales, family: *Pseudomonadaceae*). These microorganisms can impact the amount of auxin in the plant, increasing it to a toxic level, as mentioned by Weyens *et al.* (2009). Furthermore, the whole process of disease development relies on the plant's inability to defend itself from the virulence factors. These factors overtake the normal mechanism of hormone regulation, causing necrosis and death of the plant.

The *Pseudomonas* genus is one of the most complex bacterial genera. They are Gram-negative, rod-shaped, and polar-flagellated bacteria. This genus contains the largest number of species, currently including 144 species and 10 subspecies, and increasing every year. Some of the *Pseudomonas* species whose genome is determined to include *P. avellanae*, *P. savastanoi*, *P. amygdali*, *P. tomato*, and *P. meliae*.

## **1.3. PSEUDOMONAS SYRINGAE**

Out of all members of this genus, *Pseudomonas syringae* is widespread and one of the best-studied plant pathogens (Xin, Kvitko and He, 2018). There are over 50 different pathovars, of which many exert a high degree of host-plant specificity. Its pathogenicity makes it one of the most common causes of disease outbreaks and threats to global crop production. A recent example is the kiwifruit canker disease outbreak in New Zealand and Europe, caused by *Pseudomonas syringae* pv. *actinidiae*.

In the rest of this chapter, we will focus on *Pseudomonas syringae*.

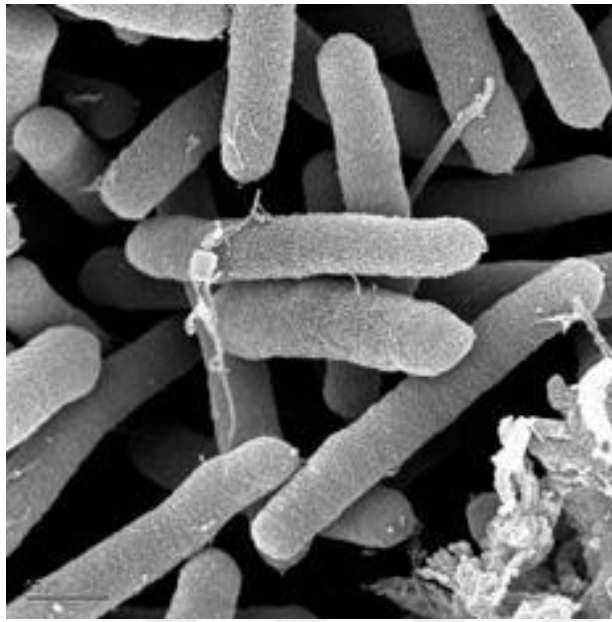


**Figure 1.1:** *Pseudomonas syringae* shown using scanning electron microscopy. It is a Gram-negative, rod-shaped, and polar-flagellated bacteria. Taken from (*Pseudomonas syringae* shown using SEM, 2021).

#### 1.4. INFECTION MECHANISM

The infection of host plants by *Pseudomonas syringae* is a complex process involving two main virulence strategies: suppression of host immunity, and creation of an aqueous apoplast to form a niche for the bacteria in the phyllosphere. After the bacteria enter the plant, it colonizes the intercellular space (apoplast) and uses the Type 3 Secretion System to deliver the effector proteins into the host cell. These effector proteins suppress the plant defense system and alter other mechanisms, triggering the release of nutrients and water from the plant cells to be used by *P. syringae*. As the bacteria multiplies and grows, it causes plant disease. During the pathogenesis of *P. syringae*, the levels of plant hormones, including the indole-3-acetic acid (IAA), increase (Djami-Tchatchou *et al.*, 2019). The disease develops during the endophytic phase, during which the bacteria enters the plant and reproduces inside the apoplast. But, the exact mechanism of the bacterial influence on the disease development is unknown (Hu *et al.*, 2010).

The apoplast is believed to be a less-than-optimal place for microbial growth, as it is a relatively nutrient- and water-poor environment (Xin *et al.*, 2016). In addition, the apoplast may contain antimicrobial compounds that are either constitutively produced or secreted into the intercellular space as part of an induced defense response. Thus, to colonize the apoplast successfully, pathogens must be able to detoxify or tolerate these antimicrobial compounds,

as well as evade or suppress further induction of host defense responses. Finally, to grow to high levels within this space, they must obtain sufficient water and nutrients from the plant cells surrounding the infection site (Kunkel and Harper, 2017).

Studies conducted by Kunkel and Harper (2017) have identified several virulence factors in *P. syringae* that are necessary for the successful infection of plant hosts after the protrusion. In general, these virulence factors include the genes *hrp* and *hrc*, effector proteins, and phytotoxins. The genes *hrp* (hypersensitive response and pathogenicity) and *hrc* (hypersensitive response and conserved) encode the structural components of the Type 3 secretion system (T3SS, TTSS) required for the translocation of bacterial effector proteins into the host cell. The effector proteins are translocated by the T3SS and presumably result in disease. The *Hrp* gene cluster also contains the genes encoding for the specific protein transport family called the MATE family, which enables the transport of virulence factors into the plant. All of them in coordination enable the transport of virulence factors into the plant and escape the plant's defense mechanisms. Furthermore, phytotoxins, such as coronatine and syringomycin, are shown to contribute to the growth of lesions and disease development (Boch *et al.*, 2002; Preiter *et al.*, 2005).

Besides the wounds, the main entry point for bacteria into the plant is the stomata. The plant has developed a defense strategy to protect itself from bacterial infections through the stomata. Upon recognition of pathogen-associated molecular patterns (PAMPs) and microbe-associated molecular patterns (MAMPs), a signaling cascade is activated in the stomatal guard cell to eventually close the stomata. This is part of the pattern-triggered immunity (PTI) in plants. Virulence factors of *P. syringae* act as a counter-defense strategy to reduce the plant's stomatal defense. One of such virulence factors is coronatine, a toxin involved in the re-opening of the stomata after its closure in response to PAMPs and salicylic acid. Coronatine is an analog of methyl jasmonic acid and can influence the endogenous antagonistic interaction between jasmonate and salicylic acid signaling pathways (Katagiri and Tsuda, 2010). Jasmonate and salicylic acid-mediated signaling pathways are related to plant resistance, prompting responses to external damage (mechanical, herbivore, or insect) and pathogen infection, thereby inducing the expression of resistance genes (Ruan *et al.*, 2019). Salicylic acid plays a central role in defense against biotrophic and hemibiotrophic pathogens, which colonize living tissue. On the other hand, jasmonate and ethylene activate defenses protecting against necrotrophs that rapidly kill plant cells to obtain nutrients. This mechanism of the hormone

defense system is known as a signaling network (Katagiri and Tsuda, 2010). This signaling network can be further modulated by several other hormones, including auxin, abscisic acid, and gibberellins. In the case of *Pseudomonas syringae*, it mostly uses auxin in the plant infection strategy (Djami-Tchatchou *et al.*, 2019).

Another plant defense strategy is the hypersensitive response, characterized by the necrosis of plant cells in the inoculated area. In this way, the bacteria are limited to the area originally inoculated and do not spread beyond its edges. Necrosis causes infected parts of the plant to die, preventing the bacterial spread to the other parts and protecting the healthy tissues. Before the plant can trigger the hypersensitive response, most often the bacteria surpasses this mechanism and overtakes the auxin hormone production (Katagiri and Tsuda, 2010).

Figure 1.2 is a graphical representation of the model of pathogen infection in resistant and susceptible plants and their corresponding responses to the infection.

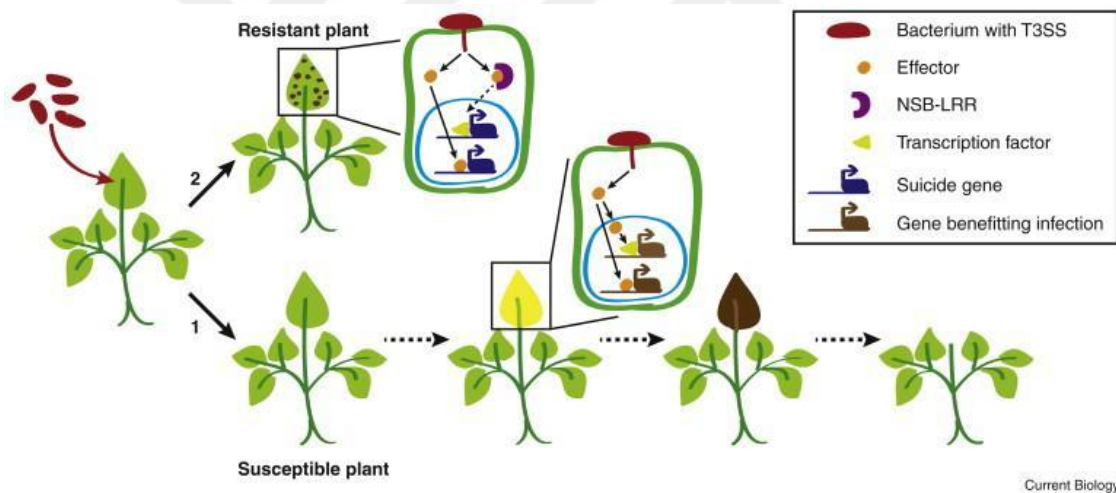


**Figure 1.2:** Immune recognition and immune response to T3SS-expressing bacteria in susceptible and resistant plants. (1) - mechanism of action of a bacteria expressing a T3SS when injecting cells of a susceptible plant. (2) - plant protection strategy with recognition of injected bacterial effectors. Photo from (Puhar and Sansonetti, 2014).

There are mainly two mechanisms of the immune response of plants to bacterial infections. The details of them are as follows.

- When a bacteria uses a T3SS to inject vulnerable plant cells, then the plant detects bacterial components and activates a protective immune response, promoting host cell survival. At the same time, effectors secreted by the bacteria can alter the target cell physiology for the benefit of bacteria, for example by acting as transcriptional activators

or by modifying host transcription factors to change gene expression (lower box in 1.2). Depending on whether plant defense or bacterial virulence mechanisms prevail, the infection by a pathogen can progress to disease.

- In contrast, detection of injected effectors in resistant plants leads to rapid target cell death. Effectors can be sequestered by promoters of suicide genes or by nucleotide-binding site leucine-rich repeat proteins, which trigger a signaling cascade. This activates the transcription of suicide genes, resulting in circumscribed necrotic lesions and prevention of bacterial spread (upper box in 1.2) (Puhar and Sansonetti, 2014).

Consistent with the effects of the Type 3 Secretion System in the suppression of stomatal closure, a recent *in vivo* imaging study conducted by Henry *et al.* (2017) showed that guard cells, which make up stomata, are targeted by *P. syringae* T3SS5 proteins (one of the components of the T3SS). Normally the plant would activate its defense mechanism, but upon the bacterial counter-mechanism, it is being silenced. Recognition of these microbe-associated molecular patterns by plants usually results in the induction of a series of basal defense responses, including a rapid oxidative burst, stomatal closure, increasing levels of defense signaling hormones, expression of defense-related genes, and production of antimicrobial compounds and lytic enzymes that act collectively to inhibit the growth of microbes in plant tissues. Nevertheless, the bacterial mechanism of infection defeats the plant's self-protection process.

Given that bacterial pathogens grow in the extracellular spaces of plant tissues, the vast majority of virulence factors are secreted, either directly into the plant cell cytosol or into the apoplast. According to Lindeberg, Cunnac and Collmer (2012), these virulence factors fall into five general classes:

- protein virulence factors, known as 'effectors', that are secreted directly into the plant cell cytosol via a specialized Type 3 Secretion System;
- cell wall-degrading enzymes;
- polysaccharides;
- low molecular weight toxins; and
- plant hormones and hormone analogs.

It is important to note that environmental factors play a significant role in *P. syringae* virulence, even though their molecular bases are still not completely understood. These factors include humidity, temperature, and coexisting microbial communities (Xin, Kvitko and He, 2018).

### 1.5. SPECIFIC PATHOVARS

We focus on two pathovars of the species *Pseudomonas syringae*, namely *Pseudomonas syringae* pv. *actinidiae* and *Pseudomonas syringae* pv. *tomato* DC3000; and on the species *Pseudomonas savastanoi* pv. *savastanoi* (*Pseudomonas syringae* pv. *Actinidiae* (bacterial canker of kiwifruit). *Invasive species compendium*, 2021).

A focus of this chapter is *Pseudomonas syringae* pv. *actinidiae* (Psa), *Pseudomonas syringae* pv. *tomato* DC3000, and *Pseudomonas savastanoi* pv. *savastanoi* species from the *Pseudomonas* genus. *Pseudomonas syringae* pv. *tomato* DC3000 and *Pseudomonas savastanoi* pv. *savastanoi* are well-studied and well-characterized species whose mechanism of action and correlation with disease occurrence is known. *Pseudomonas syringae* pv. *actinidiae* is related to those two species, but it is not well investigated. The bacteria *P. syringae* pv. *tomato* DC3000 is used for the comparison of gene similarity, interaction, and virulence effects to analyze the *matE* and *iaaL* genes and corresponding MatE and IaaL protein involved in toxicity in tomato and kiwi.

The comparison of *Pseudomonas savastanoi* (Psav) and *P. syringae* gives insight into the IAA (indole-3-acetic acid) production, which is a common mechanism in phytopathogenic bacteria and was characterized in *P. savastanoi* pv. *savastanoi*. IAA is a type of auxin which is classified as one of the virulence factors when present in high levels in the plant, disrupting plant homeostasis and enabling the bacteria to overtake the plant nutrients and water. Thus, it can be used as a reference bacterium for more detailed analysis and quality results. *P. savastanoi* was initially considered as a *Pseudomonas syringae* pathovar due to their similarity, but later was reclassified as a new species. It allows us to research the similarities and differences among two bacterial species which cause similar disease patterns in plants.

In the following sections of this thesis, we will discuss more the specific characteristics of the *Pseudomonas syringae* pv. *actinidiae* (Psa), *Pseudomonas syringae* pv. *tomato* DC3000 and *Pseudomonas savastanoi* pv. *savastanoi* (Psav).

### 1.6. *PSEUDOMONAS SYRINGAE* PV. *ACTINIDIAE*

*Pseudomonas syringae* pv. *actinidiae* is a Gram-negative, obligate aerobe, non-sporing rod bacteria. It can be found single or in pairs, or in short chains. It is motile using between one and three polar flagella.

A virulent strain *Pseudomonas syringae* pv. *actinidiae* is currently reported in Chile, China, Japan, Korea, New Zealand, Italy (2008-2009 outbreak), and other European countries including Turkey. This virulent strain is known as the pandemic strain, PsaV, Biovar 3, or Psa3. It is characterized as a highly aggressive systemic pathogen that does not produce coronatine or phaseolotoxin. PsaV is the strain responsible for the global pandemic of a kiwifruit canker disease of green-fleshed kiwifruit (*Actinidia deliciosa*) and yellow-fleshed kiwifruit (*A. chinensis*) firstly reported in Italy in 2008 (Henry *et al.*, 2017).

Colonies of PsaV on nutrient agar are translucent-white, slightly raised, glistening, and round.

### 1.7. INFECTION

This bacteria affects different parts of the plant including canes, leaves, flowers, trunks, and leaders. When a plant is infected, the extending canes show pale, translucent to dark reddish ooze from lenticels with the cracks from above olive-colored, water-soaked lesions and exude gum. These lesions elongate and the complete cane eventually becomes necrotic. On growing leaves, small, water-soaked spots are formed. Eventually, the spots become brown with chlorotic halos. Infected floral buds become brown and wither without opening. Heavy flower buds drop off. Most of the infections happen during the spring.

Symptoms of canker are first observed during the winter as droplets of ooze are being produced. During the late winter, ooze droplets increase their number and turn into a reddish-brown color. After the dormancy, natural openings are characterized by gum exudation from cracks in the bark and pruning cuts. Necrosis extends into the underlying tissue (Henry *et al.*, 2017).

In short, the symptoms include brown-black leaf spots often surrounded by a chlorotic margin, blossom necrosis, extensive twig die-back, reddening of the lenticels, extensive cankers along the main trunk and leader, and bleeding cankers on the trunk and the leader with a whitish to orange ooze (Scortichini *et al.*, 2012).

Serizawa *et al.* (1989) noted that the damage associated with *P. syringae* pv. *actinidiae* occurs

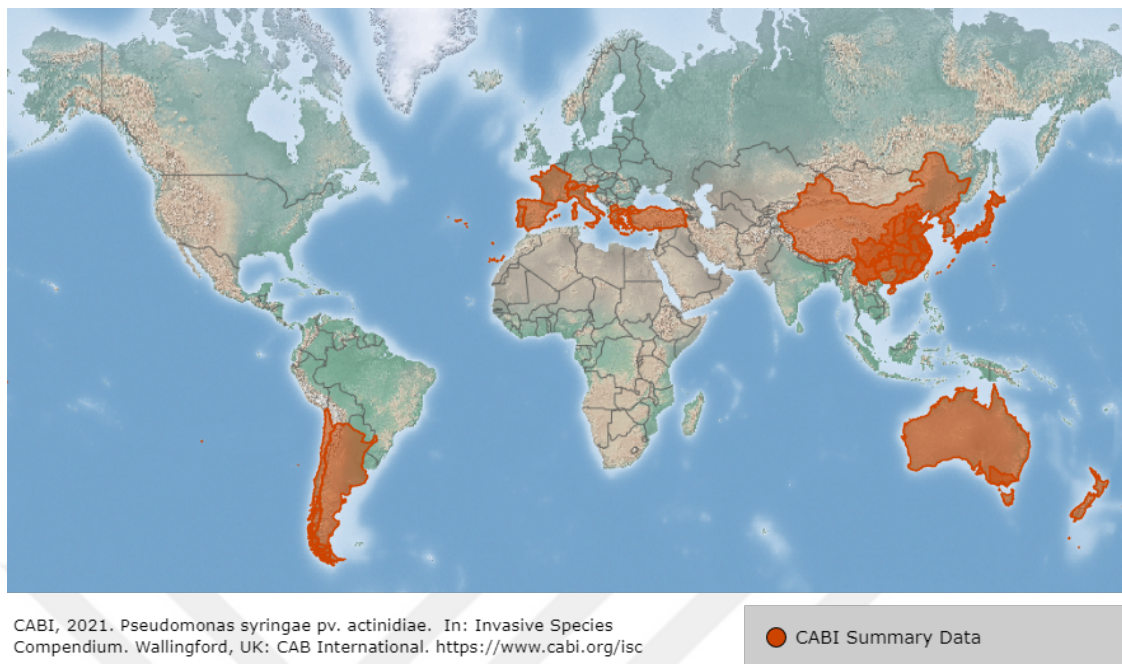


**Figure 1.3:** Distribution map of *Pseudomonas syringae* pv. *actinidiae* globally. The presence of the *P. syringae* is visible in Europe, including Turkey, Asia, Australia, and South America. Image from (*Pseudomonas syringae* pv. *Actinidiae* (bacterial canker of kiwifruit). *Invasive species compendium*, 2021).

in two phases. One phase occurs in autumn/winter and involves damage to the main vine structure and overwintering canes. The other phase occurs in spring and involves the new season's growth (leaves, flowers, and canes). The bacteria infect the host through stomata, hydathodes, lenticels, trichomes, leaf scars, or wounds, and can progress to the roots where it overwinters (Serizawa *et al.*, 1989).

As previously mentioned, environmental factors play a significant role in the virulence of *P. syringae*. In the case of *P. syringae* pv. *actinidiae*, the maximum invasiveness occurs at relatively low temperatures (10-20°C; optimum 15±3°C). The optimum temperature for the growth of *P. syringae* pv. *actinidiae* on new canes is 12-18°. Even though temperatures above 20°C are less favorable for bacterial growth, symptoms of the disease in Mediterranean countries were observed at temperatures above 25°C.

Strong winds and rainfall also favor the disease. Damage to the plants by strong winds, winter frost, or, "spring frost" enables pathogens to enter the plant through the damaged parts and cause the disease. Henry *et al.* (2017) demonstrated that populations of *Pseudomonas syringae* pv. *actinidiae* were able to survive for 21 weeks on leaves.

All of the *Pseudomonas* genus pathogens produce similar chlorotic, angular leaf lesions and infections of floral buds and flowers, but only *P. syringae* pv. *actinidiae* commonly causes the exudation of translucent gum from floral buds and the undersurfaces of leaves. Only *P. syringae* pv. *actinidiae* and *P. syringae* pv. *syringae* produce gumming, invasive, necrotic lesions which develop along canes and trunks to produce extensive cankers. The translucent gum becomes rusty brown when it is exuded from larger cankers (Henry *et al.*, 2017).

Even though the characterization of the disease is done and the bacterium is recognized as one of the most invasive pathogens in the kiwifruit production, the molecular analysis, as well as the bioinformatic characterization, is missing, especially when it comes to the protein-protein interaction in the infection process. Novel research papers indicate the extraction of the bacterial DNA sequence for further analysis, so additional data can be available for further disease characterization. It is one of the reasons why this specific bacterium is the focus of the further conducted bioinformatic analysis.

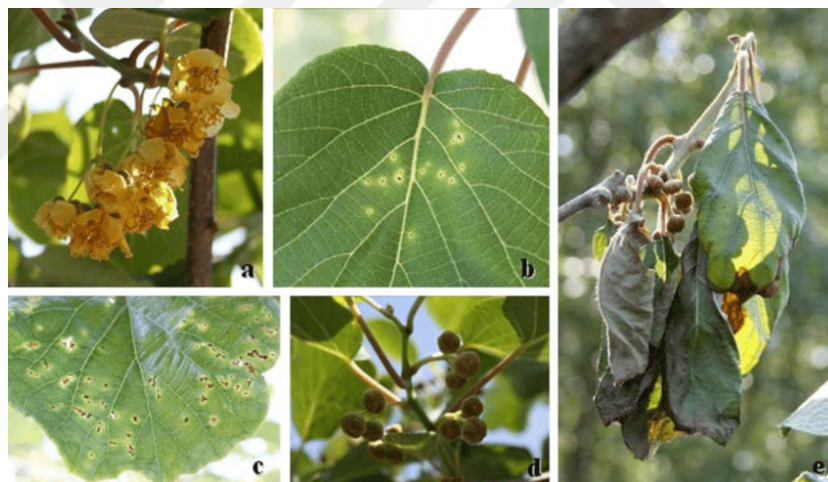


**Figure 1.4:** *Pseudomonas syringae* pv. *actinidiae* symptoms on a plant: lesions and yellow spots. The photo was taken from (*Pseudomonas syringae* pv. *Tomato DC3000* bacterial spotting on a plant leaf., 2021).

### 1.8. *PSEUDOMONAS SYRINGAE* PV. *TOMATO DC3000*

*Pseudomonas syringae* pv. *tomato* DC3000 (*PsPto*) infects tomato, *Arabidopsis thaliana*, and *Brassica* species, including turnip, mustard, collard, and cauliflower plants. Similar to *P. syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* DC3000 can grow epiphytically and endophytically on plant foliage without causing disease symptoms. In the early stages of the infective phase, *PsPto* enters the plant through wounds or natural openings and multiplies in the apoplastic space by exploiting the live host cell machinery. In this way, bacterial

survival in the apoplast is one of the key factors for the establishment of a bacterial density large enough to further infect adjacent plant tissues and cause disease (Rico and Preston, 2008).

*Pseudomonas syringae* pv. *tomato* DC3000 is found to infect almost all parts of the plant including parts of a plant in the flowering, fruiting, post-harvest, seedling, and vegetative growth stage.

Disease caused by *Pseudomonas syringae* pv. *tomato* DC3000 are various and include brown-black leaf spots which can be surrounded by the chlorotic margin and dark specks on green fruit. Dark specks on ripe fruit eventually become sunken and are surrounded by a zone of delayed ripening. Inhibition of plant growth and yield loss is present, particularly if young plants are infected (Preston, 2000). Generally, the most common symptoms include water-soaked lesions and chlorosis (see Figure 1.4, A and D)(Brooks *et al.*, 2004).

Epidemiology of the disease is seed born. *Pseudomonas syringae* pv. *tomato* DC3000 survives as a saprophyte in plant debris, soil, and leaf surfaces. Dispersion is by aerosols and rain splash. The development of disease symptoms is favored by leaf wetness and cool temperatures (13–25 °C) (Preston, 2000).

Plant apoplast is a harsh environment for bacterial survival due to the fulfillment with antimicrobial compounds, which can be performed (phytoanticipins) and inducible (phytoalexins). Antimicrobial compounds constitute chemical barriers capable of inhibiting the growth of the pathogen (Dixon, 2001).

To overcome the effect of these potentially toxic compounds and use the machinery of the plant for self-development, *Pseudomonas syringae* pv. *tomato* DC3000 developed different defense strategies, among which multidrug resistance (MDR) efflux pumps are the most widespread, together with the Type 3 Secretion System (T3SS). Multidrug resistance transporters are proteins that can recognize and pump out many different organic compounds, providing microbial resistance to antibiotics and many other antimicrobial compounds (Li and Nikaido, 2009). Multidrug resistance transporters as a defense system are found in *P. syringae* pv. *actinidiae* as well.

### 1.9. *PSEUDOMONAS SAVASTANOI* PV. *SAVASTANOI*

*Pseudomonas savastanoi* pv. *savastanoi* is a gram-negative rod-shaped bacterium with polar whips for mobility. *P. savastanoi* pv. *savastanoi* was reported to cause diseases in cultivated and wild olives and oleander.



**Figure 1.5:** *Pseudomonas syringae* pv. *tomato* DC3000 bacterial spotting on a plant leaf. Photo from (Worley *et al.*, 2013).

Olive knot disease is characterized by the formation of neoplastic tissue (knots, galls) mainly on young stems and branches. Leaves, roots, and other organs as well may occasionally be infected by the *Pseudomonas savastanoi* pv. *savastanoi*. Most infections occur through wounds on the plants, especially cracks caused by late spring frost, hail storms, wind-blown sand, agricultural practices (e.g. pruning and crop harvesting), and leaf and blossom fall. In the beginning, the knots are small (a few millimeters in diameter), pale-green wens that expand, and gradually turn greenish-brown or brown. Internally, the knots are compact, spongy tissues in which formed lysogenic cavities are filled with bacterial cells. Over time, the knots crack, partly necrotize, decay, and die within 6-8 months after their formation. In the case of heavy infections, when knots partially or completely encircle branches, the affected limbs stay dwarfed, defoliated, or get killed. On young plants in the field, the knots may cause weakness in the plant by partially encircling branches that form the structure of the trees (*Pseudomonas syringae* pv. *Actinidiae* (bacterial canker of kiwifruit). *Invasive species compendium*, 2021). On oleander, olive, and the majority of *Oleaceae*, the typical symptom is an overgrowth of the wens.

The virulence of strains of *P. savastanoi* pv. *savastanoi* and *P. savastanoi* pv. *nerii* (which is similar in the mechanism of action and structure to pv. *savastanoi*) is dependent on the production of phytohormones such as indole-3-acetic acid, and several cytokinins (Comai and Kosuge, 1982). Recent studies on Tn5-induced mutants demonstrated the presence of the *hrp* gene cluster in *P. savastanoi* pv. *savastanoi* (Sisto *et al.*, 1999). The *Hrp* gene cluster consists of the genes encoding for the specific protein transport family called the MATE family, which enables the transport of virulence factors into the plant. The presence of the gene cluster highly indicates the virulence and the strategy of the bacteria of interest. These results, together with the identification and characterization of the *hrcC* mutant of *P. savastanoi* pv. *savastanoi*, suggests that *hrp/hrc* genes play a key role in pathogenicity (*Pseudomonas syringae* pv. *Actinidiae* (bacterial canker of kiwifruit). *Invasive species compendium*, 2021). The *Hrp* gene cluster (*HrpL* operon) is present in the other species as well, such as *Pseudomonas syringae* pv. *actinidiae* and *P. syringae* pv. *tomato* DC3000, making them highly correlated and a good comparison choice for the full understanding of the virulence effects.

*P. savastanoi* pv. *savastanoi* on olive survive on the surfaces of leaves, twigs, and fruits of host plants having the highest population level in the spring and autumn. Infection occurs mainly during these seasons. Epiphytic populations of *P. savastanoi* pv. *savastanoi* together with release from active knots are the main sources of infection. Bacteria invade wounds on plants caused by leaf scars, frost injury, hail storm injury, agronomic practices, and other means. Bacteria oozing from knots may be dispersed on the plant or to other plants in the same grove by rain and aerosol formed by wind-driven rain. Another way of pathogen transmission is by propagating material, which is the main pathway by which *P. savastanoi* is distributed over long distances (*Pseudomonas syringae* pv. *Actinidiae* (bacterial canker of kiwifruit). *Invasive species compendium*, 2021).

There are many similarities between *P. savastanoi* pv. *savastanoi* which infects the olives and *P. savastanoi* pv. *nerii* which infects the oleander species. Differentiation of olive and oleander strains may be possible using molecular methods such as DNA restriction fingerprinting (Mugnai *et al.*, 1994). In this way, different protection and prevention methods could be developed.

Bacterial knot disease on olive is easily recognized when the knots are well developed. Still, in the early stages, the knots can be confused with ruptures caused by insects (by feeding or

egg-laying)(Henry *et al.*, 2017).

All three mentioned species cause similar disease outcomes and have a similar mechanism of infection. Their correlation and comparison are highly important for the understanding of the complete process and definition of the *P. syringae* pv. *actinidiae* role in disease development. All of the explanations we made are visually observable and conclusive. Our interest is to analyze the internal mechanisms, protein interactions, and gene structures and find the differences which are unique for each of the species, influencing their way of action and the spread of the disease.

### **1.10. IAA : INDOLE-3-ACETIC ACID**

One of the main common characteristics of these three different bacterial species is the ability to produce indole-3-acetic acid (IAA, a type of auxin), from the precursor tryptophan. Auxin (indole-3-acetic acid (IAA) is a type of auxin) is involved in tropisms and apical dominance (inhibition of lateral branching) in plants. Auxin also plays a role in the regulation of cell growth and elongation in stems, cell division, and differentiation, fruit development, formation of roots, and leaf fall (abscission)(Kunkel and Harper, 2017).

The main reason why we need to understand the IAA lies in the connection of the IAA with the Type 3 transport system. MatE protein, which is part of the transport machinery of the T3SS, mediates and helps in the transport of the virulence factors from the bacteria to the plant cells. More significantly, both plants and bacteria produce auxin. The exact mechanism of transport and type of the virulence factors which are transported are not well understood, but studies by Lindeberg *et al.* (2006) suggest that IAA is one of the virulence factors when produced at high levels, disrupts plant homeostasis and enables the bacteria to overtake the plant nutrients and water, elevate the amount of auxin present, and cause disease. As auxin is well known to be involved in many aspects of plant growth and development, recent findings suggest that elevated IAA levels or intensified auxin signaling promote disease development in some plant-pathogen interactions. In addition to stimulating the growth of plant cells during infection with bile-forming bacteria, auxin, and auxin signaling use the strategy to antagonize the defense reactions of plants (Kunkel and Harper, 2017).

Naturally occurring auxin indole-3-acetic acid (IAA) in plants, also known as  $\beta$ -indoleacetic acid (IAA), is formed either from the amino acid tryptophan or from the breakdown of

carbohydrates known as glycosides. IAA can be biosynthesized from tryptophan (Trp) by several pathways, but Trp-independent routes are also found and known. This hormone affects plants by its action on chemical bonds of carbohydrates comprising plant cell walls. The process permits the cells to be irreversibly altered and is accompanied by the entry of water and the synthesis of new cell wall material. As will be explained further, bacteria use the same pathway as plants to produce indole-3-acetic acid and influence the plant in a disease-causing manner.

Despite continuous research on the mechanisms of auxin production, there are still many unanswered questions regarding its biosynthesis. The molecular mechanisms regulating auxin biosynthesis at the transcriptional level and protein level are virtually unknown (Zhao, 2012). The homeostasis of auxin is mediated by a complex network of processes related to its biosynthesis, catabolism, reversible conjugation, signaling, and transport (Fu *et al.*, 2015).

Tryptophan and IAA production in bacteria is similar to the mechanism of production in plants and is commonly synthesized in plants from shikimic acid or anthranilate, as seen in Figure 1.6.

The best characterized Trp-dependent pathway in bacteria is the two-step process denoted as the indole-3-acetamide (IAM) pathway, and sequentially converts Trp to indole-3-acetamide (IAM) and then to indole-3-acetic acid (IAA), instead of standard YUCA enzyme hydrolysis as in the plant process. Enzymes leading the reaction are encoded by the *iaaM* and *iaaH* genes which are located in the same operon as the *matE* and *iaaL* genes which we will explain later in more detail (Cerboneschi *et al.*, 2016). The process is presented on 1.7.

This is compared with the Trp-independent route in which IAA may be produced from free indole by an unknown route (Wang *et al.*, 2015). Yet, the Trp-independent pathway remains unknown due to the undiscovered intermediates involved in the total pathway.

Recently, several studies have clearly shown that IAA can be a signaling molecule in microorganisms, in both IAA-producing and non-IAA-producing species. These findings raise new intriguing questions on the role of IAA in bacteria and their interaction with plants (Spaepen, Vanderleyden and Remans, 2007). A high degree of similarity between IAA biosynthesis pathways in plants and bacteria was observed. Pathway of the virulence factors transport similarity, the possible influence of the transport protein MatE on plant disease development,

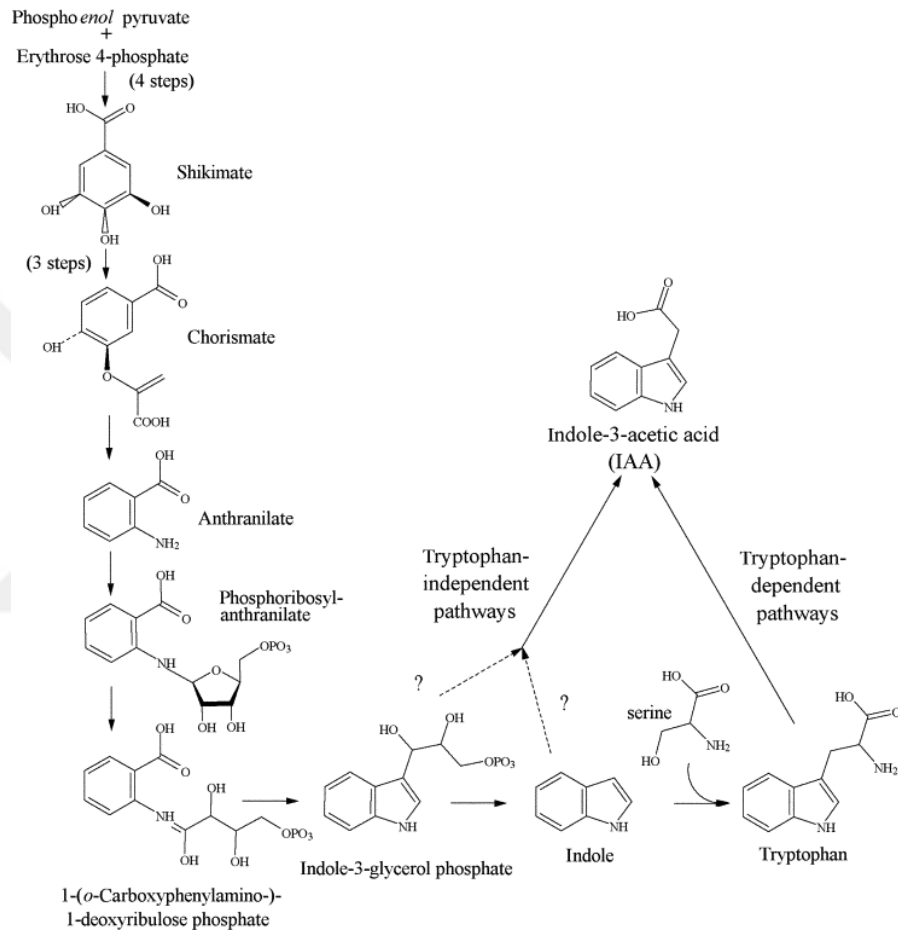


**Figure 1.6:** The Shikimic Acid Pathway in plants. Representation of a tryptophan synthesis from shikimic acid or anthranilate in plants. Different catalyzation processes by the shikimate pathway lead to tryptophan formation from indole and amino acid serine by the process of catalyzation by tryptophan synthase. Further on, by tryptophan dependent and independent pathways are formed indole-3-acetic acid (IAA). Pathway representation from (Talapatra and Talapatra, 2014).

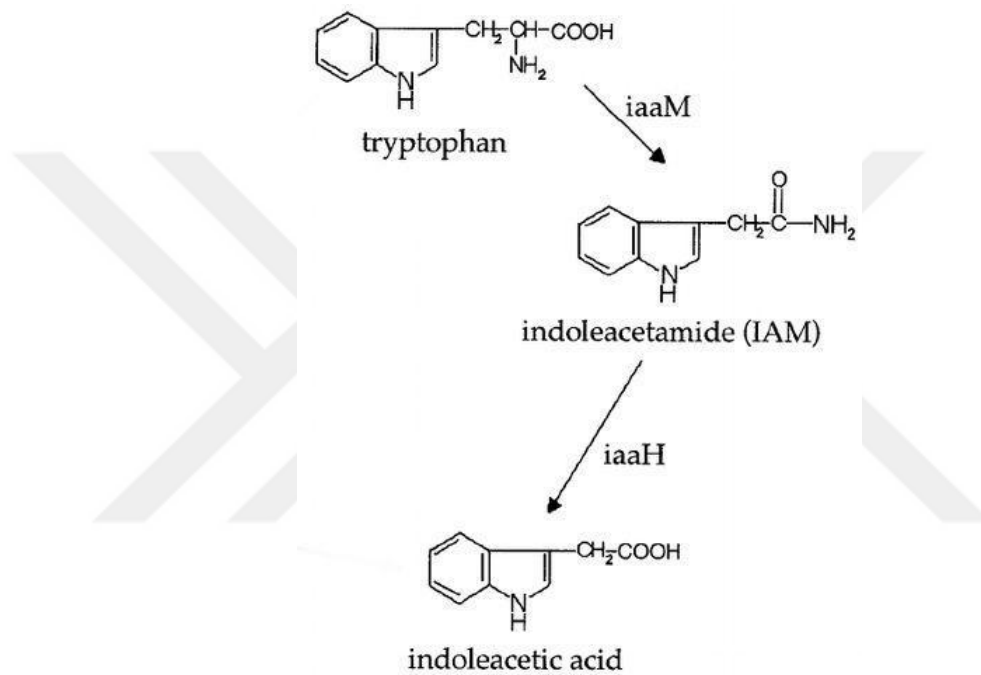


**Figure 1.7:** Indoleacetic acid (Auxin) biosynthesis: a simple two-step pathway of the tryptophan conversion into the indole-3-acetic acid in bacteria. The bacterial path of the indole-3-acetic acid production process is known as the indole-3-acetamide (IAM) pathway, where the enzymes tryptophan-2-monooxygenase (*IaaM*) and IAM hydrolase (*IaaH*) are encoded by the *iaaM* and *iaaH* genes, respectively, and sequentially transform tryptophan (Trp) to indole-3-acetamide (IAM) and then to indole-3-acetic acid (IAA), instead of standard YUCA enzyme hydrolysis as in the plant process. The end product of the pathway is indoleacetic acid. Pathway representation from (Zhao, 2012).

and bacterial genome through the protein comparison will be analyzed through different bioinformatic tools.

Another intriguing thing is the IAA production specifically by the *Pseudomonas syringae* pathovars, more precisely - *P. syringae* pv. *actinidiae* (Psa) strains. The research analyses of Chilean *P. syringae* pv. *actinidiae* isolates by Flores et al. (Flores *et al.*, 2018), McAtee et al. (McAtee *et al.*, 2018), and McCan et al. (McCann *et al.*, 2013) show that several strains have a specific way of producing indole-3-acetic acid (IAA) using the Trp-dependent pathway. A highly similar pathway was observed in plants too.

It is interesting to mention that phytopathogenic organisms tend to use the indole-3-acetamide (IAM) pathway to produce indole-3-acetic acid (IAA), whereas beneficial bacteria tend to use the indole-3-pyruvic acid (IPyA) pathway which differs in the intermediate molecules in the pathway production, but both resulting in the indole-3-acetic acid production (IAA) (Zhao, 2012).

### **1.11. INTERACTION OF PLANT AND BACTERIA**

Another important note to make is the importance of IAA (a type of auxin) for the bacteria and plants, and its influence on plant-microbe interaction. Firstly, IAA is related to bacterial toxicity, as well as drug defense. Since indole is a precursor in IAA production, it promotes bacterial excretion of toxins. It was found that indole induces the expression of multidrug exporter genes like *acrD*, *acrE*, *cusB*, *emrK*, *mdtA*, *mdtE* and *ycelL*, which protect the bacteria from extracellular toxic substances.

Secondly, indole and IAA as a final product have a role in maintaining genetic stability by contributing to the maintenance of plasmid copies. Sometimes, during the replication plasmids remain in multimers form, and this may interrupt the steady heredity of genes. To solve it, bacteria have a mechanism in which indole affects cell division until all plasmid monomeric forms are achieved.

Thirdly, indole and IAA serve as an active signal in metabolic control in bacteria. Indole can activate genes involved in amino acid metabolism, but the exact mechanism of action remains unknown (Hu *et al.*, 2010).

When it comes to the interaction between plants and bacteria, *Pseudomonas* biovars have a

beneficial effect on the plant. Bacterial IAA functions as a stimulant for root hair formation while increasing the number and length of lateral and primary roots. However, at higher concentrations, bacterial IAA can inhibit primary root growth. A well-developed root system is crucial for water and nutrient uptake, and anchoring plants in the soil. The plant reciprocates by providing the bacterial partner with exudates (containing nutrients) and a shelter so it can survive. IAA is a reciprocal signaling molecule in plant-microbe interactions. It sustains the symbiotic relationship developed between host plants and their microbial allies (Hu *et al.*, 2010).

By interacting, bacteria also have the potential to interfere with any normal auxin processes by changing the auxin amount. The impact of exogenous auxin on plant development ranges from positive to negative effects, as mentioned. The consequence for the plant is usually a function of (1) the amount of IAA produced that is available to the plant and (2) the sensitivity of the plant tissue to changes in IAA concentration (Spaepen, Vanderleyden and Remans, 2007). Thus, the whole range of different processes occurs in the plant during the infection, unavoidably changing the way the plant IAA biosynthesis mechanism happens.

Besides *Pseudomonas* species, several plant pathogens produce auxin indole-3-acetic acid (IAA) and/or virulence factors that modulate auxin host signaling. Those are: *Azospirillum*, *Bacillus*, *Bradyrhizobium*, *Enterobacter cloacae*, *Paenibacillus*, *Rhizobium*, and even *Cyanobacteria* (Spaepen and Vanderleyden, 2010).

#### **1.12. TYPE 3 SECRETION SYSTEM - T3SS**

The virulence factor of our interest is the Type 3 Secretion System (T3SS or TTSS) and its effector repertoires, which play an important role in the transmission of effector molecules from bacteria to plants in the infection process. T3SS is encoded from the HrpL operon as well as *matE* which encodes the MatE protein which is part of the T3SS machinery, and *iaaL* genes which play a role in the synthesis of IAA.

The Type 3 Secretion System is a protein secretion machinery used by many gram-negative plant bacterial pathogens to deliver effector molecules directly into the host and is at the core of the ability to cause disease (Coburn, Sekirov and Finlay, 2007).

This secretion system is a tube-like structure with ring structures embedded in the plasma membrane and a filament structure termed the hrp pilus. More specifically, the Type 3 Secre-

tion pathway is a combination of an inner membrane complex resembling the inner membrane complex of the flagellar protein secretion pathway, with an outer membrane component (Preston, 2000). The T3SS is encoded by around 20 *hrp* and *hrc* (*hrp*: hypersensitivity response and pathogenicity; *hrc*: hypersensitivity response and conserved) genes. These genes are found in several operons on the chromosome or plasmids of plant-pathogenic bacteria (Arnold, Pitman and Jackson, 2003). *Hrc* genes encoding the core components of the Type 3 Secretion pathway are highly conserved across a wide range of bacteria, opposite to the *hrp* genes which encode a mixture of structural and regulatory proteins associated with the Type 3 Secretion pathway. The Type 3 secretion genes of many bacterial pathogens are frequently clustered with other virulence and pathogenicity factors in pathogenicity islands (PAIs) or on large extrachromosomal elements (Preston, 2000). The *matE* gene in correlation with the *iaaL* gene is being found as a part of the HrpL operon of the *Pseudomonas syringae* species, a gene coding a putative multidrug and toxic compound extrusion (MATE) efflux transporters and is located upstream of the *iaaL* gene. The T3SS is encoded from the same operon as MATE efflux transporter proteins. Furthermore, *matE* gene expression was demonstrated to be T3SS-regulated, in terms of co-transcription and regulation by T3SS promoter and enhancer compounds (Spaepen, Vanderleyden and Remans, 2007).

Bacteria synthesize the T3SS effector proteins enabling the modulation of the plant cell's machinery. Effector molecules will further cause the production and increase the number of other molecules, such as auxin, which alters plants' ability to fight disease. Type 3 Secreted effector proteins are considered one of the most important classes of bacterial virulence factors. These virulence factors are characterized at a molecular level in *Pseudomonas* species, out of which the greatest role belongs to the effector proteins.

The functions of effector proteins range from interfering with the perception of microbial attack, suppressing host defense responses, activating host gene expression, modulating hormone processes (e.g. synthesis, activity, and metabolism), and signaling, to promoting tissue damage and disease development. The effector proteins have a correlated influence with the T3SS proteins in virulence induction (Xin, Kvitko and He, 2018).

Disruption of the T3SS mechanism results in loss of pathogenicity for a large number of bacterial pathogens. Most pathogenic strains that rely on the T3SS synthesize a broad spectrum (often  $\geq 30$ ) of effectors, which modulate various aspects of plant cell biochemistry,

signaling, and/or physiology (Coburn, Sekirov and Finlay, 2007).

In general, the Type 3-secreted effectors that modulate hormone homeostasis or hormone signaling are believed to promote infection and disease development by (i) manipulating regulatory crosstalk between plant defense hormone (e.g. SA, JA, and ethylene) signaling pathways to suppress host defense; (ii) stimulating physiological processes that may render host tissue more suitable for pathogen growth (e.g. release of water or nutrients into the apoplast); or (iii) promoting plant cell growth to stimulate gall, knot, canker, or pustule formation. In some interactions, manipulation of hormone biology might also result in the stimulation of plant cell death and the formation of necrotic disease lesions (Kunkel and Harper, 2017).

The mechanism of the T3SS penetration into the host plasma membrane is still unknown. Coburn et al. (Coburn, Sekirov and Finlay, 2007) suggested the most probable mechanism. Upon contact between bacteria and host cells, the syringe-like T3SS transfers proteins termed effectors from the bacterial cytosol to the cytoplasm or the plasma membrane of a single target cell. Further on, the T3SS needle/pilus may physically penetrate the membrane and/or cell wall of the eukaryotic cell, or connect to additional T3SS-associated protein complexes in the eukaryotic cell membrane and/or cell wall to provide a continuous conduit for the delivery of effector proteins into the eukaryotic cell. T3SS needle is used for the perforation of the animal cell, while pilus is used for plant cell perforation (1.8). During the process, stomatal closure is disabled by a separate class of effector proteins acting upon it. In the protrusion process of bacteria, around 25 structural and ancillary proteins are required for T3SS assembly (Coburn, Sekirov and Finlay, 2007).

### **1.13. MATE PROTEIN FAMILY, MATE AND IAAL GENES, MATE PROTEIN**

The Multidrug And Toxic Compound Extrusion (MATE) protein family encoded by the gene operon out of which of special interest for us is the *matE* gene, which plays an important role in intrinsic and acquired resistance in many bacteria. Four out of five families of MATE proteins have a function as secondary transport systems and utilize an electrochemical gradient of cations across the membrane for drug transport. One of these families, the MATE family is encoded by the *matE* gene and is implicated in the fitness and virulence of *P. syringae* pv. *tomato* (*Pst*) in tomato plants (Flores *et al.*, 2018). Thus, by analyzing the *matE* gene and encoded MatE protein, its sequence, function, and mechanism of action in the process of

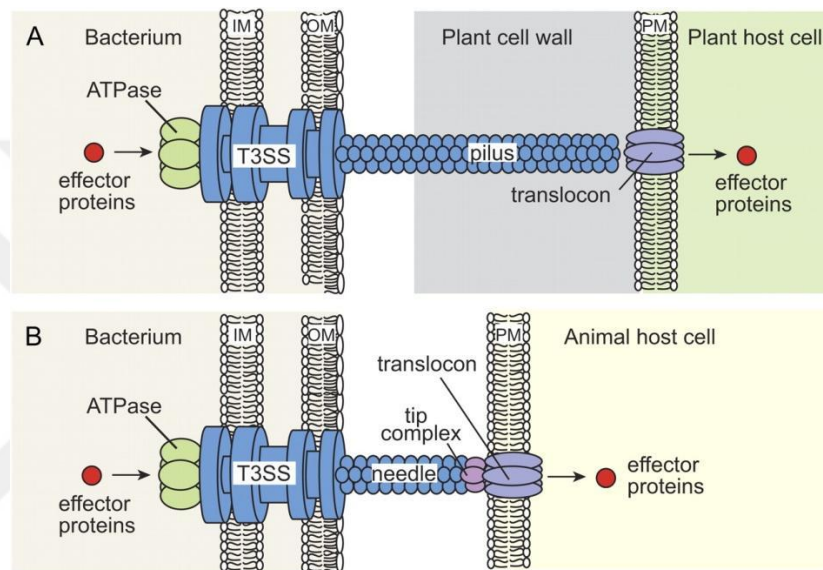


**Figure 1.8:** Schematic representation of the T3SS from the plant (A) and animal (B) pathogenic bacteria. The secretion apparatus spans both bacterial membranes and is associated with a cytoplasmic ATPase. The T3SS from plant pathogenic bacteria is connected to an extracellular pilus that presumably spans the plant cell wall. The T3SS system from animal pathogenic bacteria is associated with a short extracellular needle, which serves as a transport channel for secreted proteins. The needle is linked via the so-called tip complex to the translocon, which forms a proteinaceous channel in the host plasma membrane and allows the transport of effector proteins into the host cell cytosol. Evidence for the presence of a tip complex in plant pathogenic bacteria is still missing. IM, inner membrane; OM, outer membrane; PM, plasma membrane. Image from (Puhar and Sansonetti, 2014).

indole-3-acetic acid production in *P. syringae* pv. *tomato* and *P. savastanoi* can be made the correlation between virulence effect and general influence on plant disease development of kiwifruit and tomato species caused by *P. syringae* pv. *actinidiae*.

The crystal structures of some bacterial MATE transporters are available in protein data banks. All of the analyzed transporters are membrane proteins characterized by the presence of twelve transmembrane helices (TM 1–12), forming an internal cavity. In this pocket, several quite specific and conserved residues provide the binding sites for the substrates and for those ions ( $H^+$  or  $Na^+$ ) whose gradient across the membrane serves as an energy source (Zhang *et al.*, 2012). In a study conducted by Galán *et al.* (Galán *et al.*, 2014) on the binding free energy, calculations showed that free IAA and IAA-Lys could be realistically considered among the potential substrates transported by Psn23 MatE transport protein. Free IAA and IAA-Lys are transported across the bacterial membrane, as a counterpart of the Type 3 transport process of the MATE family in which MatE and T3SS are structural components. This mechanism depicts the importance of the genes encoding for these proteins.

The discovery of the MATE protein family was significant and revolutionary. In 2005, Otsuka *et al.* published a report which brought new light to this protein transport family. Based on sequence similarity, they found two MATE transporter genes, *hMATE1*, and *hMATE2* on human chromosome 17 (Otsuka *et al.*, 2005). The *hMATE1* gene is primarily expressed in the liver, skeletal muscle, and kidney, whereas the *hMATE2* gene is expressed in the kidney. At present, the mammalian MATE transporters are thought to mediate the final step in the excretion of toxic organic cations (Masuda *et al.*, 2006). Thus, it has become clear that the MATE transporters are present not only in bacteria but also in mammalian and plant cells, even though there is not enough detailed data on its mechanism of action in mammals. Before this, it was considered that MATE transporters are only reserved for plants, animals, and bacteria, not humans. Although the MATE transporters are widespread in many bacteria, the number of the MATE transporters in each bacterium is different.

MATE transporters are widely distributed in Gram-positive and Gram-negative bacteria, where they are usually associated with the efflux of organic cations for multidrug resistance. Conversely, MATE protein pumps found so far in plants were demonstrated to be involved in the transport of a broader range of substrates than in bacteria, and have many other roles beyond detoxification, including secondary metabolite transport, xenobiotic detoxification,

aluminum tolerance, and disease resistance. MatE proteins from the MATE protein family also regulate overall plant development by controlling phytohormone transport, terminal growth processes, and aging.

Pathogens may secrete only a few effectors proteins, as in the cases of *Pseudomonas* and *Yersinia*, or several dozen, as in the case of *Shigella*. Secretion signals are embedded within the N-termini of T3SS substrates and are not cleaved. Many, but not all T3SS effectors have chaperones that guide them to the T3SS base, where they are secreted in an ATP-dependent, unfolded state (Green and Mecsas, 2016). Some of the effector proteins bind to chaperons and deliver them to the T3SS needle (in the case of the animal cell attachment). Then, ATPase triggers the release of the effector proteins and helps them to unfold so they can pass through the narrow channel of the secretion system machinery. In the case of bacteria entering the plant, effector proteins do not bind to the chaperons but are linked to the different types of helper proteins, with the ATPase triggering the unfolding and passing through the T3SS.

Secretion in the unfolded state and total penetration into the cell is enabled by the MATE protein family as a part of the T3SS machinery. While in unfolded state, they are being transported through the Type 3 secretion channel into the cell, where they acquire folded functional state and further influence the cell and the host organism.

The structural components of T3SSs are typically encoded in a few operons, which can be found either in pathogenicity islands in the bacterial chromosome or on plasmids. In the case of the *Pseudomonas* species, T3SS is encoded in the *hrp/hrcI* operon, which encounters the *matE* gene amongst others.

The putative MATE family transporter gene is abbreviated as the *matE* gene and the protein it encodes as MatE protein. Virulence factors found in association with *P. syringae* pathovars (*P. syringae* pv. *actinidiae* and *P. syringae* pv. *tomato* DC3000), and *Pseudomonas savastanoi* pv. *savastanoi*, were also associated with the *matE* gene (Flores *et al.*, 2018). In several *P. syringae* pathovars, the *matE* gene which is correlated with the fitness and virulence of *Pseudomonas* was found in synteny with the gene *iaaL* in the process of production of indole-3-acetic acid (Flores *et al.*, 2018). The *matE* gene together with the upstream associated *iaaL* gene is part of the *hrp-hrc* T3SS gene cluster (HrpL operon) (*hrp*: hypersensitivity response and pathogenicity; *hrc*: hypersensitivity response and conserved genes - Type 3 Secretion System gene cluster) which encodes all the structural genes required for assembly of the

Type 3 Secretion System (T3SS) in which MATE proteins are part of the transport machinery. In this process, T3SS is one of the most important virulence induction machinery, without which infection would not be possible. Assembly of the T3SS enables injection of the effector proteins into the plant cell, which further promotes disease development.

The HrpL regulon to which the MATE protein family belongs is one of the most researched regions of the *P. syringae* pv. *tomato* DC3000 genome. The HrpL regulon contains genes involved in the hypersensitive response, and other genes encoding the pathogenicity (Hrp) Type 3 Secretion System (T3SS machinery), a repertoire of Type 3 effector proteins, as well as genes encoding factors unrelated to the T3SS, such as *iaaL* and the genes for coronatine synthesis.

The *matE* gene is correlated with the *iaaL* gene, belongs to a gene coding a multidrug and toxic compound extrusion (MATE) efflux transporters, and is located upstream of the *iaaL* gene in the HrpL regulon. The T3SS is encoded from the same operon as MATE efflux transporter proteins. Furthermore, *matE* gene expression was demonstrated to be T3SS-regulated, in terms of co-transcription and regulation by T3SS promoter and enhancer compounds (Spaepen, Vanderleyden and Remans, 2007).

Gene *iaaL* is widely distributed within the *P. syringae* complex, and several copies of *iaaL* are often found in close proximity to insertion sequences or encoded in plasmids, all of which provide evidence of horizontal gene transfer. In the study by Flores et al. (Flores *et al.*, 2018), it was demonstrated that the *iaaL* gene, which is highly conserved within the *P. syringae* complex, is found in synteny with the *matE* gene in the genome of *P. syringae* and *P. savastanoi* strains belonging to genomospecies 2, 3, and 4.

The existence of hrp box promoter sequences upstream of both the *iaaL* and *matE* genes was reported in *P. syringae* pv. *tomato* DC3000 (*Pto* DC3000). The divergent hrp box promoter of the *iaaL* gene binds HrpL and promotes HrpL-dependent transcription (Lam *et al.*, 2014). Data shows that both of the genes, *iaaL*, and *matE*, encode upstream hrp box-like promoters that are HrpL-dependent. More importantly, the promoter sequence of the *iaaL* gene is encoded within the 3' end of the *matE* gene in *Pto* DC3000, suggesting that it could be constrained by the genetic code (Flores *et al.*, 2018).

In terms of the promoter region of *matE*, a highly conserved hrp box motif, 5'-GGAACT-N19-

TCAG-3', was found in the strains of *Pseudomonas* genomospecies 3 and was shown that the expression from *matE* in *Pto* DC3000 may take place from two different promoters: one which is weak and constitutive and another which is stronger and HrpL-dependent (Flores *et al.*, 2018).

In *P. syringae* and *P. savastanoi* strains, the *iaaL* gene is found in synteny with an upstream *matE* gene, and the *iaaL* gene is transcribed both independently and as part of an operon with *matE* in the *P. syringae* pathovars. Deletion of either the *iaaL* or the *matE* gene results in reduced fitness and virulence of *Pseudomonas syringae* pv. *tomato* DC3000 in tomato plants (Castillo-Lizardo *et al.*, 2015). Thus, by examining the *matE* gene and protein sequence by bioinformatic analysis, comparing it with the *iaaL* sequence protein sequence from *Pseudomonas syringae* and *Pseudomonas savastanoi* can directly imply a relationship of the *matE* gene with pathogenicity and virulence effectors. Together with the *matE* gene, in the HrpL regulon are *iaaL*, *iaaM*, and *iaaH* genes (Glass and Kosuge, 1988). The location of the aforementioned genes as well as their involvement in the IAA synthesis implies their close correlation and probable pathogenicity influence correlation.

Although the *iaaL* and *matE* genes are widespread and highly conserved among *P. syringae* and *P. savastanoi* strains, a role in pathogenesis was only recently demonstrated by the *iaaL* gene by the work of two different groups, working in *Pseudomonas syringae* pv. *tomato* DC3000 and *Pseudomonas savastanoi* pv. *nerii* strain Psn23. The *matE* gene is yet to be finally characterized in plant pathogenesis (Kunkel and Harper, 2017).

Lim *et al.* (Lim, Lee and Hussein, 2011) demonstrated that as the intracellular indole-3-acetic acid (IAA) levels increase, *matE* is upregulated following transcriptional activation of the downstream gene *iaaL*, as occurs for genes that are part of the same operon. The key feature of operons in the regulation of bacterial gene networks was recently demonstrated in *E. coli*, in which the expression of a gene that is part of an operon increases with the length of the operon and according to its distance from the end of the operon during its transcription (Lim, Lee and Hussein, 2011). By increasing the so-called "transcription distance", there is more time for translation to occur during transcription, which will increase the expression of genes in the operon proportionally to their transcription distances. To confirm this scenario, in *P. syringae* species the presence of a functional Hrp promoter upstream of the *iaaL* gene involved in IAA biosynthesis further supports the role of IAA production in virulence as the *matE* and *iaaL*

genes were demonstrated to be transcribed both dependently and independently (Lim, Lee and Hussein, 2011). Hypersensitive response and pathogenicity (Hrp) promoters are known for controlling genes encoding the Hrp (Type 3 Protein Secretion) machinery and a few Type 3 effector proteins in DC3000.

The physical and functional relationship of genes, promoters, and additional factors is presented in 1.9 and 1.10 in regards to the *Pseudomonas savastanoi* pv. *nerii*, since it is the most studied group from the *Savastanoi* species and is most closely related to the *Pseudomonas savastanoi* pv. *savastanoi*, and the most accurate data were related to this bacterial strain.

In *Pseudomonas savastanoi* pv. *savastanoi* location of the *matE* gene together with the *iaaL* gene is on the plasmid pIAA1, while on *Pseudomonas syringae* pv. *tomato* DC3000 are located on chromosomes.

Figure 1.9 represents the position of *iaaL* and *matE* genes relative to the *iaaM* and *iaaH* genes of *P. savastanoi* pv. *nerii* operon, with the promoters and additional factors influencing their expression.



**Figure 1.9:** Physical relationships between IAA metabolism, T3SS, and drug efflux in *P. savastanoi* pv. *nerii* strain Psn23. Organization of the genomic region containing *iaaM/iaaH* operon and the *iaaL* and *matE* genes, including putative promoters and factors regulating their expression. The transcriptional direction of each gene is indicated in orange, yellow, and gray arrows for *iaaL*, *matE*, and the *iaaM/iaaH* operon, respectively; *hrpbox* promoters: black arrows; AuxREs: green arrows; DS element: blue arrow. Image from (Cerboneschi *et al.*, 2016).

As *matE* gene encodes for the MATE effector proteins which are part of the Type 3 transport system of bacteria and the main route of entry into the plant host and is found in the *hrp/hrc1* operon together with the genes encoding for the T3SS as well, influencing and having a role in the indole-3-acetic acid production which is the main toxic compound and disease caused in the kiwi and tomato, it is important to fully understand its sequence and protein interactions, as well as the mechanism of action in the virulence process (Lim, Lee and Hussein, 2011).

Applying different bioinformatics tools for the analysis of *matE* and *iaaL* gene sequences

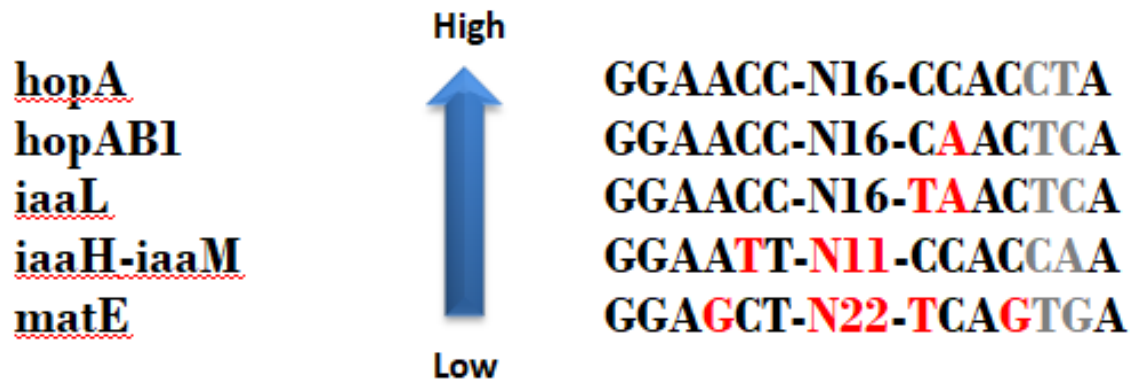


**Figure 1.10:** Physical and functional relationships between IAA metabolism, T3SS, and drug efflux in *P. savastanoi* pv. *nerii* strain Psn23. The hrp-box promoter strength was evaluated in silico for five different loci related to IAA metabolism and T3SS activation. The vertical blue arrow indicates promoter strength from the bottom (low activation) to the top (high activation). Black bold nucleotides indicate perfect identity with the hrpA promoter; red bold nucleotides underline specific mismatches. Gray bold nucleotides indicate highly conserved nucleotides.

and especially of the proteins they encode for, can be determined their effect on the toxicity in tomato and kiwi plants, and thus developed strategies as well as novel research strategies against *Pseudomonas* pathogens.

We decided to focus on the aforementioned three species of *Pseudomonas* and MatE proteins since the starting point of this research was articles mentioning the existence of these proteins correlated with the disease occurrence in tomato and kiwi species. We could not find any additional information on this correlation so we decided to analyze the mentioned proteins.

## 2. METHODS

Bioinformatics requires a specific approach to a topic, deeper and results-oriented, so every step has to be carefully done. We started this analysis by reading articles on plant bacteria which were our interests. After reading several different articles on plant bacteria, I crossed over a *Pseudomonas syringae* pv. *actinidiae* bacteria which is not investigated in detail yet. More specifically, there were mentioned *matE* and *iaaL* genes and corresponding proteins of *P. syringae* pv. *actinidiae* which are correlated, but still not fully understood in their influence on the toxicity of bacteria and disease cause. Mechanisms of action of this bacteria are still under research, so I decided to try to do some additional research on its genome and the proteins it has. Our first search was on NCBI to try to find relevant proteins and genes as previously mentioned. We extended our search to the related species from the *Pseudomonas* family for a better understanding and comparison of our species of interest. Related species were from *Pseudomonas syringae* pv. *tomato* DC3000 and *Pseudomonas savastanoi* pv. *savastanoi*. We also crossed with the mentioned related species in the articles we read as well. Finding the *matE* and *iaaL* gene sequences and their consecutive proteins led us to further research and protein analysis, as will be further mentioned.

Bioinformatics tools were used in this work to analyze *matE* and *iaaL* genes as well as the MatE and IaaL proteins and their interactions and expression influencing toxicity of *Pseudomonas syringae* and *Pseudomonas savastanoi* species; We used different tools to do the Multiple sequence alignment, Phylogenetic tree construction, 3-D structure prediction, domain identification and analysis, interactome analysis, subcellular localization of the MatE and IaaL proteins, docking site prediction, Ramachandran plot, QMEAN quality assessment, MSA and phylogenetic tree comparison between more different bacterial strains. We also did an additional analysis to check the conservation level of the intergenic sequences between *iaaL* and *matE* genes. Also, the same analysis was performed on the promoter sequences of the gene cluster to which the aforementioned genes belong.

### 2.1. RETRIEVING THE *MATE* AND *IAAL* GENE AND CORRESPONDING PROTEIN SEQUENCES

Basic Local Pairwise Sequence Alignment (BLAST) is used to identify regions of similarity that may indicate functional, structural, and/or evolutionary relationships between

two biological sequences (protein or nucleic acid). This research compared nucleic acid and protein sequences of *Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* DC3000, and *Pseudomonas savastanoi* pv. *savastanoi*. By using BLAST we could find regions of similarity between our protein and species of interest with other biological sequences. Also, we were comparing sequences of interest to sequence databases and checked for the statistical significance of the results we got. Having an insight into the sequence of similarity of *Pseudomonas* strains as well as their cotranscription and location in the same operon could help us to better understand the virulence effects and mechanism of action the bacteria are using in the infection process. It would be an additional step in solving the “puzzle” of the disease occurrence in the kiwifruit and tomato. The first step to do was to search our genes and proteins of interest from three organisms (*Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* str. DC3000, and *Pseudomonas savastanoi* pv. *savastanoi*) in the NCBI and got the initial protein sequences belonging to these species. The sequences we got are the query sequences from the 3.1, and after running the BLAST analysis we could see that our query sequences were not complete. Rather, we got the full-length sequences using the tblastn tool. We inserted MatE and IaaL protein sequences into the tool (WP\_003380685.1 sequence of the AMP binding protein, identified as an indoleacetic acid lysine ligase - IaaL, and WP\_003380687.1 for the multidrug transporter MatE protein). We chose the reference sequence genome to be full and limited to *Pseudomonas syringae*, excluded environmental samples and took the max target to be 100. The results we obtained gave us the complete sequences of the MatE and IaaL proteins in *Pseudomonas* species (3.3 and 3.8).

To get the full sequences of the proteins, we did pairwise sequence alignment. The aim was also to check if our sequences of interest were present only in the respective organisms and if the sequences we have are complete. We run BLAST again, taking the parameters to be non-redundant proteins, working with the PSI-BLAST, max target sequences to be 10, and using BLOSUM45 as a matrix parameter.

The results we got are summarized and presented in the 3.3 and 3.8 as well as in the Appendix 11.

Our next step was multiple sequence alignment of the sequences we got.

## 2.2. MULTIPLE SEQUENCE ALIGNMENT (MSA)

MSA is used to align and compare three or more biological sequences of similar length. From the output of MSA applications, homology can be inferred and the evolutionary relationship between the sequences studied. Some of the well-known and available MSA applications are ClustalW, Kalign, MAFFT, MUSCLE, T-Coffee, and COBALT from NCBI (Hall, 2013). For our research, we chose ClustalW, T-Coffee, and COBALT analysis. By this step, we aim to align different bacterial sequences of *Pseudomonas* species, especially sequences related to the HrpL operon which contains *matE* and *iaaL* genes, and see how similar they are.

ClustalW is the first server that was used in this analysis, and during the procedure, we used MatE and IaaL protein sequences from three species of interest as a focus group (*P. syringae* pv. *actinidiae*, *P. syringae* pv. *tomato* DC3000, and *P. savastanoi* pv. *savastanoi*). Also, for this analysis additional sequences of the same/highly similar proteins were found in other organisms from the same species, since the same proteins are found in other organisms as well. During the NCBI taxonomy search, we have found out that sequences such as those of our proteins of interest are found under the different names in different organisms. We decided to use them for comparison and conservation analysis. The T-coffee and COBALT servers used the same protein sequences as for the ClustalW. Analyses were done twice in some cases, adding the outgroup organisms and running the program, and without the addition of the outgroup organisms. Parameters used in the ClustalW, T-coffee, and COBALT analysis were standard. The protein sequences used are those that have a 100% match with our firstly found partial protein sequences and can be found in 3.5 for IaaL proteins, 3.9 for MatE proteins, and Appendix 11. The procedure was simple, entering the sequences of proteins of interest or accession numbers of the proteins into the online tools of ClustalW, T-coffee, and COBALT with the standard parameters and running the program. The results we got are automatically sent to us and are used in further analysis. Full multiple sequence alignment by ClustalW for IaaL proteins is available in the Appendix 1, for the MatE proteins Appendix 2. For the T-coffee MSA analysis, IaaL results are available in the Appendix 3, and for the MatE in the Appendix 4. For the COBALT MSA analysis, full alignment for the IaaL proteins is available in the Appendix 5, and for the MatE proteins in the Appendix 6.

After making an MSA and finding a difference in the sequence structure of our proteins of interest, we needed a phylogenetic tree to investigate the distance and correlation between them.

### 2.3. PHYLOGENETIC TREE CONSTRUCTION

By creating a phylogenetic tree (phylogeny), would be obtained diagram that depicts the lines of evolutionary descent of different species, organisms, or genes from a common ancestor. Thus, we were able to see the evolutionary correlation and distance between the *Pseudomonas* strains of interest. Building a phylogenetic tree requires four distinct steps: (Step 1) identify and acquire a set of homologous DNA or protein sequences, (Step 2) align those sequences, (Step 3) estimate a tree from the aligned sequences, and (Step 4) present that tree in such a way as to convey the relevant information to others (Hall, 2013). The phylogenetic tree we made depicts the relation between different *Pseudomonas* strains, and specifically to see how and where is our sequence of interest – *Pseudomonas syringae* pv. *actinidiae*. There would also be another phylogenetic tree depicting the relatedness of *Pseudomonas* strains with other bacterial strains for a more broad and general comparison.

We made different phylogenetic trees for the IaaL and MatE proteins from *Pseudomonas* species with outgroup organisms that have the same proteins. The aim was to check their general phylogenetic relatedness and evolutionary distance. While searching the taxonomically related proteins, we wanted to find related and unrelated species containing the IaaL and MatE protein sequences. Firstly, we wanted to see how phylogenetically related IaaL and MatE proteins, separately, are inside the same genus and species, and with the outgroup that has a similar similarity percentage. We got the proteins and species for the broad phylogenetic trees with outgroup organisms for IaaL and MatE proteins by BLAST tool excluding genus by genus and species by species, to find as many as possible different species which have our proteins of interest. The first BLAST analysis was for the *Pseudomonas* group genomospecies 3 without excluding anything. The reason why we chose genomospecies 3 is that our proteins of interest belong to these genomospecies. The next step was to exclude the *Pseudomonas* group genomospecies 3. Best matching proteins were taken for the phylogenetic tree sequence analysis. Exclusion for each next step was as follows: exclusion of *Pseudomonas* group, Proteobacteria, Cyanobacteria, Bacteria, Archaea, and Procaryotes. For each time we excluded a certain group, we chose the proteins of the highest percentage similarity to be included in the phylogenetic tree. Names and descriptions of the proteins are given in the Appendix 12.

While searching for the closest outgroup organisms for the MatE proteins and excluding the *Pseudomonas* species, we got only *Glycine max* to be an outgroup organism having a MatE protein. Excluding only *Pseudomonas syringae* pv. *actinidiae* taxa 1096, we got results to be

the proteins of *P. syringae* pv. *actinidiae* from other taxa and with low percentage identity, so that specific sequence was used in the phylogenetic tree as well. When excluding bacteria, we got again the *Pseudomonas* strains. Accession numbers and descriptions of the proteins used for the phylogenetic tree are available in the Appendix 13 and 14.

When including Eucharina (taxid: 289280) for the MatE protein, we got no results.

With the phylogenetic tree construction, we could validate that our protein of interest, MatE from *Pseudomonas syringae* pv. *actinidiae* has amino acid residue polymorphisms from the other proteins, and the phylogenetic tree showed it as a distinct protein from others. Thus, our next step was a 3D protein prediction.

#### **2.4. 3D STRUCTURE PROTEIN PREDICTION AND VALIDATION**

Tertiary protein structure is the three-dimensional shape of the protein of interest and the most important level of protein folding. 3D folding of the protein dictates its function and its interaction with other proteins. The three-dimensional structure is formed by many different bonds (interactions) between R groups of amino acids which make the upper side chains of the protein. Sidechains make the strand of the molecule bend and loop around in a more complicated three-dimensional form. Interactions between R groups include hydrogen bonding, ionic bonding, dipole-dipole interactions, and London dispersion forces (Zedalis and Eggebrecht, 2018).

To find the native structure of proteins, the physicochemical principles as well as identifying the lowest free-energy states are considered as the starting point. For the prediction of target proteins with unknown structures, the bioinformatics-based methods have earned considerable success (Farhadi, 2018). Protein structure prediction methods are mainly classified into three types: ab Initio folding, comparative (homology) modeling, and threading. Each mentioned method may be applied for a protein structure, depending on the existence of related experimental structures that are deposited in the Protein Data Bank (PDB) or NCBI. Since the sequences of interest for this research are available in the PDB and NCBI, as well as in the UniProt data bank, the necessary methods could be applied. The methods mentioned require the usage of the different tools/software for the protein structure prediction.

Each process of the method application is divided into several steps, starting from the nucleotide sequence which is retrieved from the data bank. The next step is the insertion into

one of the software and following the further proposed procedure of sequence comparison, homology search, creation of the protein model, and final possible prediction of the tertiary protein structure.

There are several available different tools for tertiary protein prediction out of which many use homology analysis. Some of them are as follows.

PHYRE2 – which uses the alignment of hidden Markov models via Hhsearch to significantly improve the accuracy of alignment and detection rate. It also incorporates a new ab initio folding simulation called “Poing” to model regions of proteins with no detectable homology (Kelley *et al.*, 2015).

Swiss-model - An automated comparative protein modeling server. Requires additional viewer tools such as DeepView – Swiss-PdbViewer.

There are few other software for 3D protein prediction, and each of them has its way of processing the data, so the time it takes to finish the processing differs. Some of the methods take up to a few minutes while others take more than a few hours. It all depends on the sequence length, availability of the homolog proteins, and some other software specifications. In our research, we used PHYRE2 and the Swiss model. Using both software was simple. For the Swiss-model analysis, the MatE protein sequence (WP\_003380687.1) was inserted into the query tab, and search results were further analyzed. When uploading the protein sequence into the server, we had two options offered. One was to find the best template for our model and the second was to build a model by an automatically assigned template. When firstly searching for a template that would be the best available match to our sequence, the best result was 6idp.1.A template from *Vibrio cholerae*, the first one in a row of 50 offered templates. When we choose a second option, to automatically build a model, the server chooses the best available template and builds according to it. Even then, the template model taken by the server was the same as the template we chose. For the PHYRE2 analysis, the procedure was similar. MatE protein sequence is entered into the query field, and the “search” option is chosen. There was no need for pre-template identification, but the template was automatically assigned. The template model according to which the MatE protein model is built is c4z3pA, with a 100% confidence level, and with no data on which organism it is found. The template used for the Swiss-model server also had a 100% confidence level. When we enter the details about the model built by PHYRE2, there is an available list of other templates which can be

used for the MatE protein prediction. The sequence identity of the template with our MatE protein is 20%. It was possible to download results directly from the software. Together with the 3D model available, we got the QMEAN value of the protein, additional info on the sequence identity, coverage, and Ramachandran plot analysis automatically done by the Swiss model.

## **2.5. IDENTIFICATION OF PROTEIN DOMAINS**

Protein domains often correspond to structural domains which are self-stabilizing and fold independently of the rest of the protein chain. They may occur independently or as part of complex multidomain protein architectures which evolve by domain accretion, domain loss, or domain recombination (*Domain*, 2018). The majority of domains annotated in UniProtKB/Swiss-Prot are defined by the InterPro resources PROSITE, Pfam, and SMART. Additionally, the InterProScan server has been used to analyze the MatE and IaaL protein so the double check on domains can be obtained. By knowing protein domains and their 3D structure, many different antibacterial means can be developed for the protection of the kiwifruit and tomato species. During the process, a standard procedure was used. The sequence of the MatE protein (WP\_003380687.1) has been inserted into the servers and the results are summarized and presented. Once the sequence is inserted into the server, the procedure is automatic and gives the best scoring results with the additional info on the domain found.

## **2.6. PREDICTION OF THE INTERACTOME**

The theoretical analysis of interactions depends on heterogeneous sources of biological information, such as sequence and structural databases, literature, and experimental data. The main databases containing experimental information about protein-protein interactions are the Database of Interacting Proteins (DIP), STRING, and InterProScan. The literature data on selected protein sequences are available from the iHOP and STRING databases. In the Protein Data Bank (PDB) database, one can find the three-dimensional structures of protein complexes, while in SCOP, one can find protein domains, whereas protein families in PFAM. It is possible to find the homologous proteins to each of a given protein's interacting partners using tools like PSI-BLAST (Plewczyński and Ginalski, 2009). For our purposes, we used the PSI-BLAST server.

The sequence of the MatE protein (WP\_003380687.1) and IaaL protein (WP\_003380685.1) of the *Pseudomonas syringae* pv. *actinidiae* is used for the interactome prediction by inserting them into the aforementioned databases starting from the STRING and comparing the results with other databases. The procedure was automatic, the server found the interacting proteins and additional info on the overall procedure. The results we got were an indicator of the MatE and IaaL protein interactions. Knowing the MatE and IaaL protein interactions enabled us to work on the ideas of prevention of those interactions and so the disease development. We choose the confidence level for the analysis to be 0.900.

## 2.7. SUBCELLULAR LOCALIZATION OF MATE AND IAAL PROTEINS

In recent years, many technical advances have been made in the prediction of subcellular localization of proteins, and there are three groups of methods for predicting the subcellular localization of multilabel proteins from various organisms. The first group of the methods is part of the „pLoc-mGneg“ web-server for the subcellular localization of proteins of gram-negative bacteria, and the second group of methods is the “LocTree3“ server (Shen, Tang and Guo, 2019). We used both of them to determine the MatE and IaaL proteins' cellular location.

The sequence of MatE protein (WP\_003380687.1) and IaaL protein (WP\_003380685.1) is run through the mentioned web servers. The procedure was automatic, we inserted the MatE protein and IaaL protein sequences separately into the server and used standard procedure and parameters. Results gave us insight into the position of the MatE and IaaL protein in the transmembrane area since the *matE* gene encodes for the Type 3 secretion system proteins which are part of the transmembrane protein family and is in a cluster with the *iaaL* gene. The results we got from two servers were analyzed and compared. IaaL protein is not found in the *Glycine max* species.

## 2.8. PREDICTION OF DOCKING SITES

Docking is a molecular modeling technique that is used to predict how a protein (an enzyme) interacts with small molecules (ligands). The ability of a protein (enzyme) and nucleic acid to interact with small molecules to form a supramolecular complex plays a major role in the dynamics of the protein, which may enhance or inhibit its biological function (Roy, Kucukural and Zhang, 2010). In our research, docking can be performed by ClusPro and RCSB PDB software by placing the rigid molecules or fragments into the protein's active site

using different approaches like clique-searching, geometric hashing, or pose clustering. The first step of docking is the generation of the composition of all possible conformations and orientations of the protein paired with the ligand. The second step is that the scoring function takes input and returns a number indicating favorable interaction (Shen, Tang and Guo, 2019).

ClusPro represents the first fully automated, web-based program for the computational docking of protein structures. Users may upload the coordinate files of two protein structures through ClusPro's web interface, or enter the PDB codes of the respective structures, which ClusPro will then download from the PDB server (Comeau *et al.*, 2004).

RCSB PDB is another database that we used for the ligand-binding activities and analysis of the proteins of interest since it has its ligand data bank and can show the interaction of the ligands with the proteins of the same or highly similar structure.

Firstly the RCSB PDB was used to find all of the possible and suitable ligands for the proteins of interest, and then analysis was performed using ClusPro software. Other software could be used in a case where ClusPro or RCSB PDB do not provide desired results. Standard procedure was used. The sequence of interest of the MatE protein (WP\_003380687.1) was inserted into the server. Standard automatically assigned parameters are used. The analysis took a few hours to be completed. After the results showed the best possible result, the ligand was marked and protein interaction analyzed. Also, some additional information on the ligands is available once the server finishes the procedure.

## **2.9. RAMACHANDRAN PLOT AND QMEAN QUALITY ASSESSMENT**

A Ramachandran plot is a way to visualize energetically favored regions for backbone dihedral angles against amino acid residues in protein structure. Histograms are used to count  $\Phi$  (Phi; C-N-CA-C) /  $\Psi$  (Psi; N-CA-C-N) occurrences for all displayed categories. The number of observed  $\Phi$  /  $\Psi$  pairs determines the contour lines (*Structure assesment*, no date).

QMEAN is a composite estimator based on different geometrical properties and provides both global (i.e. for the entire structure) and local (i.e. per residue) absolute quality estimates based on one single model.

The Ramachandran plot analysis was joined with the 3D protein prediction analysis by the Swiss-model server. The Swiss model is an accurate tool and gives automatic Ramachandran

analysis with the protein model predicted. Thus, there was no need to do a separate analysis for it. Standard procedures and automatic parameters are used in the process of 3D protein prediction. The sequence has been inserted into the server and the results analyzed.

## **2.10. MSA AND PHYLOGENETIC TREE COMPARISON BETWEEN MORE DIFFERENT BACTERIAL STRAINS**

The MatE protein sequences of the bacterial strains of interest such as *P. syringae* pv. *actinidiae* and *P. syringae* pv. *tomato* DC3000, as well as *P. savastanoi* pv. *savastanoi* are inserted into the multiple sequence alignment applications such as ClustalW and T-Coffee for the comparison together with many more similar sequences, so we can see the broader picture of sequence relatedness. In many cases, the input set of query sequences is assumed to have an evolutionary relationship by which they share a linkage and are descended from a common ancestor. From the resulting MSA, sequence homology can be concluded and phylogenetic analysis can be conducted to describe the sequences' shared evolutionary origins (Chang *et al.*, 2008). The most commonly used tool for phylogenetic tree construction and MSA is ClustalW. MSA between more different strains is performed following the standard procedure. Sequences are taken from the NCBI, tax id number 103796 for *Pseudomonas syringae* pv. *actinidiae*. Strains with different matching values with our sequence of interest are taken for the MSA analysis. Furthermore, one by one strain and species are canceled from the search, so we got more distant and different strains that contain the same *matE* gene. They are also taken to construct MSA and a phylogenetic tree.

In 2.1 are descriptions and accession numbers of the focus protein sequences from *Pseudomonas* species used in this research and bioinformatic analysis.

**Table 2.1:** MatE and IaaL protein accession numbers from three *Pseudomonas* species of interest. The focus is on the MatE and IaaL proteins. The Focus species are *Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* DC 3000, and *Pseudomonas savastanoi* pv. *savastanoi*. These are the full protein sequences which are used throughout the main analyzes after the BLAST of homologous sequences was performed. The first step was with the partial sequences, so after we found the full sequences, the analysis continued. Partial sequences are in the brackets.

| Species                                                  | IaaL                           | MatE                           |
|----------------------------------------------------------|--------------------------------|--------------------------------|
| <i>Pseudomonas syringae</i> pv. <i>actinidiae</i>        | WP_003380685.1<br>(AXK59710.1) | WP_003380687.1<br>(AXK59714.1) |
| <i>Pseudomonas syringae</i> pv. <i>tomato</i><br>DC 3000 | WP_007247550.1<br>(AAO53915.1) | WP_005782319.1<br>(AAO53914.1) |
| <i>Pseudomonas savastanoi</i> pv.<br><i>savastanoi</i>   | WP_057456329.1<br>(RMS74473.1) | WP_019742037.1<br>(RMS74471.1) |

## 2.11. CONSERVATION ANALYSIS OF THE *IAAL* AND *MATE* INTERGENIC AND PROMOTER SEQUENCES

Analysis of the conservation of intergenic and promoter sequences of *iaaL* and *matE* genes was performed following the BLAST analysis. Firstly, we got the whole genome sequences of different organisms which contain *matE* and *iaaL* genes. These sequences are from the initial step of retrieving the MatE and IaaL proteins and gene sequences from NCBI. The next step was to extract the intergenic parts of the sequences between our two genes of interest so that further analysis could be done. We used the BLAST database to extract only the intergenic and promoter sequences. To extract them, we used `tblastn` and inserted *matE* and *iaaL* gene sequences. From other parameters in this step, we choose the option **ref.seq.genome(full)** for the analysis, limited our search only to the *Pseudomonas syringae* species, and excluded environmental samples. Our maximum target was to be 100. By this step, `tblastn` translated the sequence of the genome into a protein and compared our amino acid sequences to other sequences. Comparing them, the server found the location of the genes we were interested in. We listed the results we got in the table as input data into the excel sheet which is used and available in the appendix. By sorting the results we got into the sheet, we had an insight into where one gene finishes and another gene starts, so the length of the intergenic sequence could be determined. We got it to be 118 nucleotides. After extraction of those specific parts

of the sequences by the program, we were able to perform MSA in the ClustalW program to check for the conservation level of our sequences. For the ClustalW analysis, we took 118 nucleotides. We took 118 nucleotides length of the sequence because we showed that between the end of the *matE* gene and the start of the *iaaL* gene is 118 nucleotides long intergenic sequence. We took exactly that part for our analysis. The same procedure of extraction is done for the promoter region of our genes. Since they are both encoded from the same promoter, we needed an upstream part of the sequence, since promoters are generally located upstream of genes. The length of the sequence we used for the promoter conservation analysis was 100 nucleotides. We performed ClustalW analysis to check the conservation level of the promoter sequence. The full promoter sequence for the MatE and IaaL proteins is available in the Appendix 9, and for the intergenic sequence in the Appendix 8.

### 3. RESULTS

The results we got from different analyzes are shown in the following parts. Each separate analysis is explained, results are presented in the tables and figures, and connected with the explanations and descriptions in the discussion part. The analyzes are ordered according to the complexity of the procedures and following step by step the clues about the thesis aim.

The starting point is three organisms (*Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* str. DC3000, and *Pseudomonas savastanoi* pv. *savastanoi*). We found them in NCBI by description from the articles we read. We referred to the (Flores *et al.*, 2018) and results from the NCBI led us to the proteins AXK59710.1, AAO 53915.1, and RMS 74473.1, all of the partial sequences of indoleacetic acid lysine ligase protein, and proteins AXK59714.1, AAO53914.1 and RMS74471.1 corresponding to multidrug transporter MatE protein, also described as a membrane protein. A detailed explanation of this protocol can be found in the Materials and methods part. Sequences we first identified in articles we read and started our analyzes with are PCR products, and this is the reason why they are incomplete. They were a good starting point, so we could see what our next step would be.

#### 3.1. COMPARISON BETWEEN PROTEINS

Table 3.1 and 3.2 show the BLAST results of protein alignment of *Pseudomonas syringae* pv. *actinidiae* with two other species of interest: *Pseudomonas syringae* pv. *tomato* DC3000 3.1, and *Pseudomonas savastanoi* pv. *savastanoi* for the *iaaL* and *matE* genes and corresponding proteins 3.2. The goal of the BLAST analysis is to check how similar are the sequences of the given proteins in terms of the amino-acid sequence comparison, so their structural, function and evolutionary relationship can be determined.

**Table 3.1:** Results of comparing protein sequences of IaaL as found in *Pseudomonas syringae* pv. *actinidiae* (AXK59710.1) versus *Pseudomonas syringae* pv. *tomato* DC3000 (AAO53915.1) and *Pseudomonas savastanoi* pv. *savastanoi* (ACC69167.1)

|                                                                                                                                              | Sequence<br>length | Total<br>score | Query<br>cover | Percentage Accessed<br>identity | length |
|----------------------------------------------------------------------------------------------------------------------------------------------|--------------------|----------------|----------------|---------------------------------|--------|
| <i>Pseudomonas syringae</i> pv.<br><i>actinidiae</i> (AXK59710.1) vs<br><i>Pseudomonas syringae</i> pv.<br><i>tomato</i> DC3000 (AAO53915.1) | 353<br>aa/395 aa   | 818            | 100%           | 100%                            | 395    |
| <i>Pseudomonas syringae</i> pv.<br><i>actinidiae</i> (AXK59710.1) vs<br><i>Pseudomonas savastanoi</i> pv.<br><i>savastanoi</i> (RMS74473.1)  | 353 aa/<br>146 aa  | 283            | 39.0%          | 98.6%                           | 146    |

**Table 3.2:** Results of comparing protein sequences of MatE as found in *Pseudomonas syringae* pv. *actinidiae* (AXK59713.1) versus *Pseudomonas syringae* pv. *tomato* DC3000 (RMU94288.1) and *Pseudomonas savastanoi* pv. *savastanoi* (RBSW01000328.1).

|                                                                                                                                              | Sequence<br>length | Total<br>score | Query<br>cover | Percentage Accessed<br>identity | length |
|----------------------------------------------------------------------------------------------------------------------------------------------|--------------------|----------------|----------------|---------------------------------|--------|
| <i>Pseudomonas syringae</i> pv.<br><i>actinidiae</i> (AXK59714.1) vs<br><i>Pseudomonas syringae</i> pv.<br><i>tomato</i> DC3000 (AAO53914.1) | 424<br>aa/479 aa   | 741            | 99.0%          | 91.0%                           | 479    |
| <i>Pseudomonas syringae</i> pv.<br><i>actinidiae</i> (AXK59714.1) vs<br><i>Pseudomonas savastanoi</i> pv.<br><i>savastanoi</i> (RMS74471.1)  | 424 aa/<br>467 aa  | 729            | 100%           | 87.7%                           | 467    |

We have achieved high query coverage of 100% of the IaaL and 99% of the MatE protein BLAST analysis for the *Pseudomonas syringae* pv. *actinidiae* versus *Pseudomonas syringae* pv. *tomato* DC3000. Even though the query cover is high, the percentage identity of the sequences is lower than query coverage and is 91% for the MatE protein. Accessed length

is the longer sequence length, even though the protein sequences used in this analysis are incomplete. In the case of *Pseudomonas syringae* pv. *actinidiae* versus *Pseudomonas savastanoi* pv. *savastanoi* the query cover for the IaaL protein is low, only 39%, while for the MatE protein it is 100%. Percentage identity is again less than 100%, indicating the difference between the proteins compared.

We conclude that these sequences seem to be incomplete and thus we need to find the corresponding full sequences.

For our further analysis, we needed to use complete protein sequences, so we performed a homology search for our proteins of interest and found the complete matching proteins.

## **3.2. IAAL PROTEINS**

### **3.2.1. Full Sequences of the Proteins**

Homology search gave us full sequences of our proteins of interest, so we were able to continue a more detailed analysis. Once we got the full sequences of MatE and IaaL proteins, we validated them by pairwise sequence alignment. We checked if the sequences were only in the respective organisms, and if the sequences were complete. BLAST homology search results are available in the table 3.3.

**Table 3.3:** Protein homology of partial and full sequences of IaaL proteins. First three results of the search for homologs of each partial protein AXK59710.1, AAO53915.1, and RMS74473.1 (indoleacetic acid lysine ligase - IaaL) from *Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* str. DC3000 and *Pseudomonas savastanoi* pv. *savastanoi*. We confirm that all of them are highly similar to AMP-binding proteins in other organisms in the same genus. There are no gaps in any of these alignments. All the queries were strictly contained in the full-length proteins

| Query      | Subject        | Taxonomy                                         | Pct<br>iden-<br>tity | Mis<br>match | Query<br>cover-<br>age |
|------------|----------------|--------------------------------------------------|----------------------|--------------|------------------------|
| AXK59710.1 | WP_087674970.1 | <i>Pseudomonas syringae</i>                      | 100                  | 0            | 100                    |
| AXK59710.1 | WP_003380685.1 | <i>Pseudomonas syringae</i>                      | 100                  | 0            | 100                    |
| AXK59710.1 | WP_005737399.1 | <i>Pseudomonas syringae</i><br>group             | 99.15                | 3            | 99.72                  |
| AAO53915.1 | WP_007247550.1 | <i>Pseudomonas syringae</i><br>group             | 100                  | 0            | 100                    |
| AAO53915.1 | WP_122285283.1 | <i>Pseudomonas syringae</i><br>group genomosp. 3 | 99.747               | 1            | 100                    |
| AAO53915.1 | WP_024667964.1 | <i>Pseudomonas syringae</i><br>group             | 99.747               | 1            | 100                    |
| RMS74473.1 | WP_057433268.1 | <i>Pseudomonas syringae</i><br>group             | 100                  | 0            | 100                    |
| RMS74473.1 | WP_057456329.1 | <i>Pseudomonas savastanoi</i>                    | 99.747               | 1            | 99.75                  |
| RMS74473.1 | WP_044322280.1 | <i>Pseudomonas syringae</i><br>group             | 99.494               | 2            | 99.75                  |

BLAST compares two sequences to determine their homology, similarity, and evolutionary relatedness. It scans for local segments of a sequence attempting to find similar regions and align them, so the comparison and further analysis can be done. In our case, the analysis of gene and protein sequences of *iaaL* and *matE* genes has been done from three sequences (separate analysis of one query sequence with two existing sequences). The analysis is done on the gene and protein sequences of *iaaL* and *matE* genes of *Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* DC3000, and *Pseudomonas savastanoi* pv. *savastanoi*, as well as other organisms from *Pseudomonas* species, have proteins matching

our proteins of interest to a specific level. The main results are the new sequences that extend the original ones.

From the results of the homology analysis, we got a good starting point for further analysis and investigation. After getting the sequences we could analyze in detail, the next step was multiple sequence alignment, which will help us to see more amino acid structures of the proteins and possible polymorphisms between them. Polymorphic changes in the structure of the proteins would help us differentiate between the sequences and have an influence on the toxicity of the bacteria we are focusing on. This is the reason why each performed procedure is significant for us.

### **3.2.2. Multiple Sequence Alignment**

Sequences used in the following analyses were retrieved from NCBI (GenBank) and UniProt and then aligned using the ClustalW and T-coffee programs.

#### **3.2.2.1. Why multiple alignments?**

The homology and evolutionary relationship of IaaL and MatE proteins can be determined by a multiple sequence alignment.

After aligning sequences containing our proteins of interest and determining their homology and identity, we could see that there is a high similarity between them. We used sequences that are found in different organisms but have our proteins of interest. The organisms we used are the top-scoring BLAST results we got when searching for the homologous proteins in outgroup organisms. Detailed steps are explained in the Materials and methods part.

For our research, it is important to see the homology between the sequences according to which we analyze our sequence of interest. It enables us to conclude the possible role and influence of the specific *actinidiae* strain on toxicity.

By using multiple sequence alignment, it is possible to see the conserved sequences, analyze their similarity and check for the parts which are different.

Also, according to the multiple sequence alignment, it is possible to draw further analysis and conclusions on function similarities and differences, as well as on their influence on virulence and disease development. It is due to the various protein functions in the organism, and variants or small changes which are significant for the function of the proteins and complete cells.

All of the aforementioned possibilities by multiple sequence alignment helped us to further analyze our proteins of interest and draw conclusions on their influence on the bacterial toxicity and disease occurrence in tomato and kiwi.

#### **3.2.2.2. *Why these sequences in particular?***

The sequences used correspond to the *Pseudomonas* species and are listed in 3.3.

We used the protein sequences resulting from the BLAST analysis shown in table 3.3, and similar proteins found in distant bacteria as “outgroup” organisms in the analysis. The outgroup was chosen among the top-scoring BLAST results excluding proteins from organisms in the *Pseudomonas* genus. The BLAST table is in Appendix 11.

#### **3.2.3. MSA Results of IaaL Proteins**

MSA show all the sequences, one in each line, and the gaps necessary to align them. When the sequences are too long to fit in the text, they continue in the next block, in the same order.

In the last line an asterisk (\*) indicates that the residue is the same across all query proteins; a colon (:) indicates that the residues are different but highly similar across query proteins, and a dot (.) indicates that residues are different but similar to some degree, across query proteins (Sievers *et al.*, 2011).

The first analysis was done using ClustalW on the IaaL proteins from various species with and without outgroup organisms. The outgroup organisms are the top-scoring organisms from the BLAST search for the homologous organisms containing our proteins of interest. Accession numbers of all sequences used in this analysis are available in Appendix 1.

First, we did ClustalW MSA of multiple organisms from different species containing the IaaL protein, to see the sequence match and conservation level. We got a result showing many different residues across the species, and with the conservation level to be low due to the polymorphisms present. This might be due to the difference between the species used, so we did a ClustalW with only *Pseudomonas* species, to check their sequence identity and conservation level.

In 3.2 is shown ClustalW multiple sequence alignment of the IaaL proteins without the outgroup organisms. In this way, it is possible to see better alignment and fewer gaps between close organisms from *Pseudomonas* species.

|                   |                                                               |     |
|-------------------|---------------------------------------------------------------|-----|
| Salmonella        | -----MTQRDP                                                   | 6   |
| Escherichia       | -----MQPRITAITGCQAFDYTAFRDSVRISSHVYSTTGDSGSLTQRDP             | 44  |
| Arsenophonus      | MLSIYKEEMVWNTVKDKIATLFLQPAISQWYQQFSQSIPVMEKK-----GYLAKNDQ     | 53  |
| Erwinia           | -MSAYDVEKEWLRVAGIAAETHQDADLRGFTYQDFRKWVPVMDKD-----GFTAQTER    | 52  |
| P. savastanoi     | -MEDSHDCLYRSRISSTAANKHQNNDFEGFTYQDFRTHVPIMDKD-----GFAAQTER    | 52  |
| P. syringae_group | -MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRTHVPIMDK-----GFAAQTER      | 52  |
| P. syringae_NR    | -MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRKHVPIMDK-----GFAAQTER      | 52  |
| P. syringae       | -MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRKHVPIMDK-----GFAAQTER      | 52  |
| P. avellanae      | -MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRKHVPIMDK-----GFAAQTER      | 52  |
| P. savastanoi_NR  | -----THQNDFEGFTYQDFRTHVPIMDK-----GFTAQTER                     | 33  |
| P. amygdali       | -MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRKHVPIMDK-----GFTAQTER      | 52  |
| P. savastanoi_NR1 | -MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRKHVPIMDK-----GFTAQTER      | 52  |
|                   | :: :                                                          |     |
| Salmonella        | A----AEQIPMLLGATSGTTGAMKMISVRLKSSKDAVSASEHNLIHNLRSRQVFSAGDVV  | 62  |
| Escherichia       | A----AEQIPMLLGATSGTTGAMKMISVRLKSSKDAVSASEHNLIHNLRSRQVFSAGDVV  | 100 |
| Arsenophonus      | SLHLLNSKNSYLIGFTSGTSGTLKRCLYKMSGDVYVVKG-----CEEYNIFCSTDTC     | 105 |
| Erwinia           | CVA--ENERNCLIGFTSGTSGNLKRCYYYDCEVDEDSSR----SNVFRSNGFIQPGDRC   | 106 |
| P. savastanoi     | CLE--RNERNCLIGFTSGTSGNLKRCYYYDCEVDEDSSL----SNVFRSNGFIPGDRC    | 106 |
| P. syringae_group | CLE--RNERNCLIGFTSGTSGNLKRCYYYDCEVDEDSSR----SNVFRSNGFIQPGDRC   | 106 |
| P. syringae_NR    | CLE--RNERNCLIGFTSGTSGNLKRCYYYDCEVDEDSSR----SNVFRSNGFIQPGDRC   | 106 |
| P. syringae       | CLE--RNERNCLIGFTSGTSGNLKRCYYYDCEVDEDSSR----SNVFRSNGFIQPGDRC   | 106 |
| P. avellanae      | CLE--RNERNCLIGFTSGTSGNLKRCYYYDCEVDEDSSR----SSVFRSNGFIQPGDRC   | 106 |
| P. savastanoi_NR  | CLE--RNERNCLIGFTSGTSGNLKRCYYYDCEVDEDSSR----SNVFRSNGFIQPGDRC   | 87  |
| P. amygdali       | CLE--RNERNCLIGFTSGTSGNLKRCYYYDCEVDEDSSR----SNVFRSNGFIQPGDRC   | 106 |
| P. savastanoi_NR1 | CLE--RNERNCLIGFTSGTSGNLKRCYYYDCEVDEDSSR----SNVFRSNGFIQPGDRC   | 106 |
|                   | . .: *: * ***: * : * . . . . . . . . : *                      |     |
| Salmonella        | GNLFTINLFSSLHYSACEILKHCRAHVVPVGDIALLRKSHFDLADIQLSTLFGVPATLV   | 122 |
| Escherichia       | GNLFTINLFSSLHYSACEILKHCRAHVVPVGDIALLRKSHFDLADIQLSTLFGVPATLV   | 160 |
| Arsenophonus      | ANLFTINLFSALHCLFSNVASACGANIISIGDFNVLAASEHFDALIDVNTNVLIGIPCEII | 165 |
| Erwinia           | ANLFTINLFSALNNITMMAGNCGAHVVSVDITLLTRSHFEALKEIKLNVLLGVPSTIL    | 166 |
| P. savastanoi     | ANLFTINLFSALNNTITMMAGNCGAHVVSVDITLTKSHFEALNSIKLNVLLGVPSTIL    | 166 |
| P. syringae_group | ANLFTINLFSALNNTITMMAGNCGAHVVSVDITLTKSHFEALNSIKLNVLLGVPSTIL    | 166 |
| P. syringae_NR    | ANLFTINLFSALNNTITMMAGNCGAHVVSVDITLTKSHFEALNSIKLNVLLGVPSTIL    | 166 |
| P. syringae       | ANLFTINLFSALNNTITMMAGNCGAHVVSVDITLTKSHFEALNSIKLNVLLGVPSTIL    | 166 |
| P. avellanae      | ANLFTINLFSALNNTITMMAGNCGAHVVSVDITLTKSHFEALNSIKLNVLLGVPSTIL    | 166 |
| P. savastanoi_NR  | ANLFTINLFSALNNTITMIAGNCGAHVVSVDITLTKSHFEALNSIKLNVLLGVPSTIL    | 147 |
| P. amygdali       | ANLFTINLFSALNNTITMMAGNCGAHVVSVDITLTKSHFEALNSIKLNVLLGVPSTIL    | 166 |
| P. savastanoi_NR1 | ANLFTINLFSALNNTITMMAGNCGAHVVSVDITLTKSHFEALNSIKLNVLLGVPSTIL    | 166 |
|                   | .*****: : : : * *::: :*: : : .*: * : : .*:*: . :              |     |

**Figure 3.1:** ClustalW multiple sequence alignment of IaaL proteins from *P. syringae* pv. *actinidiae*, *P. syringae* pv. *tomato* DC3000, *P. savastanoi* pv. *savastanoi*, and additional sequences of interest. Sequence Alignment of three IaaL proteins for indoleacetic acid lysine ligase protein from *Pseudomonas syringae* pv. *actinidiae*, indoleacetate-lysine ligase protein from *Pseudomonas syringae* pv. *tomato* DC3000, and IaaL protein from *Pseudomonas savastanoi* pv. *savastanoi* which is the focus of this research was created by the ClustalW program. This figure also includes the sequences of *Salmonella enterica*, *Escherichia coli*, *Arsenophonus nasoniae*, and *Erwinia mallotivora* as an outgroup organism for better insight into the differences between the total length sequences. The start part of the protein has a less similar sequence and conserved amino acids, even though it might contain the same main functional domain of the protein. Comparison with the outgroup organisms resulted in many gaps.

|                                |                                                             |     |
|--------------------------------|-------------------------------------------------------------|-----|
| P. syringae_groupNR3           | MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQNFVRKHPIMDKEGFTAQTERCLERNERN | 60  |
| P. syringae_groupNR4           | MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRTHVPIMDKEGFTAQTERCLERNERN | 60  |
| P. savastanoi1                 | MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRTHVPIMDKEGFTAQTERCLERNERN | 60  |
| P. savastanoi                  | MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRTHVPIMDKEGFTAQTERCLERNERN | 60  |
| P. syringae_genomosp3          | MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRTHVPIMDKEGFAAQTERCLERNERN | 60  |
| P. syringae_groupNR            | MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRTHVPIMDKEGFAAQTERCLERNERN | 60  |
| P. syringae_tomatoDC3000       | MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRTHVPIMDKEGFAAQTERCLERNERN | 60  |
| P. syringae_tomatoT1           | MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRTHVPIMDKEGFAAQTERCLERNERN | 60  |
| P. syringae_groupNR2           | MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRTHVPIMDKEGFAAQTERCLERNERN | 60  |
| P. syringae_group              | MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRTHVPIMDKEGFAAQTERCLERNERN | 60  |
| P. syringae_NR                 | MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRTHVPIMDKEGFAAQTERCLERNERN | 60  |
| P. syringae                    | MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRTHVPIMDKEGFAAQTERCLERNERN | 60  |
| *****; **, *****; *****; ***** |                                                             |     |
| P. syringae_groupNR3           | CLIGFTSGTSGNLKRCYYYYDCEVDESSRSNVFRSNGFIQPGDRCANLFTINLFSALNN | 120 |
| P. syringae_groupNR4           | CLIGFTSGTSGNLKRCYYYYDCEVDESSRSNVFRSNGFIQPGDRCANLFTINLFSALNN | 120 |
| P. savastanoi1                 | CLIGFTSGTSGNLKRCYYYYDCEVDESSRSNVFRSNGFIQPGDRCANLFTINLFSALNN | 120 |
| P. savastanoi                  | CLIGFTSGTSGNLKRCYYYYDCEVDESSRSNVFRSNGFIQPGDRCANLFTINLFSALNN | 120 |
| P. syringae_genomosp3          | CLIGFTSGTSGNLKRCYYYYDCEVDESSRSNVFRSNGFIQPGDRCANLFTINLFSALNN | 120 |
| P. syringae_groupNR            | CLIGFTSGTSGNLKRCYYYYDCEVDESSRSNVFRSNGFIQPGDRCANLFTINLFSALNN | 120 |
| P. syringae_tomatoDC3000       | CLIGFTSGTSGNLKRCYYYYDCEVDESSRSNVFRSNGFIQPGDRCANLFTINLFSALNN | 120 |
| P. syringae_tomatoT1           | CLIGFTSGTSGNLKRCYYYYDCEVDESSRSNVFRSNGFIQPGDRCANLFTINLFSALNN | 120 |
| P. syringae_groupNR2           | CLIGFTSGTSGNLKRCYYYYDCEVDESSRSNVFRSNGFIQPGDRCANLFTINLFSALNN | 120 |
| P. syringae_group              | CLIGFTSGTSGNLKRCYYYYDCEVDESSRSNVFRSNGFIQPGDRCANLFTINLFSALNN | 120 |
| P. syringae_NR                 | CLIGFTSGTSGNLKRCYYYYDCEVDESSRSNVFRSNGFIQPGDRCANLFTINLFSALNN | 120 |
| P. syringae                    | CLIGFTSGTSGNLKRCYYYYDCEVDESSRSNVFRSNGFIQPGDRCANLFTINLFSALNN | 120 |
| *****; *****; *****; *****     |                                                             |     |
| P. syringae_groupNR3           | ITTMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQHGHHIE   | 180 |
| P. syringae_groupNR4           | ITTMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQHGHHIE   | 180 |
| P. savastanoi1                 | ITTMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQHGHHIE   | 180 |
| P. savastanoi                  | ITTMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQHGHHIE   | 180 |
| P. syringae_genomosp3          | ITTMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQHGHHID   | 180 |
| P. syringae_groupNR            | ITTMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQHGHHID   | 180 |
| P. syringae_tomatoDC3000       | ITTMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQHGHHID   | 180 |
| P. syringae_tomatoT1           | ITTMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQHGHHID   | 180 |
| P. syringae_groupNR2           | ITTMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQHGHHID   | 180 |
| P. syringae_group              | ITTMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQHGHHID   | 180 |
| P. syringae_NR                 | ITTMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQHGHHID   | 180 |
| P. syringae                    | ITTMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQHGHHID   | 180 |
| *****; *****; *****; *****     |                                                             |     |

**Figure 3.2:** ClustalW analysis of the IaaL proteins from *Pseudomonas* species without out-group organisms. When only *Pseudomonas* species are compared, there is a much higher conservation level between the organisms and sequence identity. No gaps present and single mutation events.

For the additional check for protein interaction, the same analysis with an outgroup organism is done using the COBALT program from NCBI, so the data we got are presented in 3.3.

|                |     |                                                                                  |     |
|----------------|-----|----------------------------------------------------------------------------------|-----|
| WP_195918284.1 | 1   | -----MTQRDPAAE-----QIPMLLGATSGTTGAMK                                             | 26  |
| WP_157940369.1 | 1   | -----MQPRITAITGCAFDYTAFRDSVRIS[7]SGSLTQRDPAAE-----QIPMLLGATSGTTGAMK              | 64  |
| WP_051296966.1 | 1   | mLSIYKEEMVWNTVKDKIATLFLPAISQWYQQFSQIPVMEK                                        | 77  |
| WP_034939628.1 | 1   | -MSAYDVEKEWLRVAGIAAETHQDADLRGFTYQDFRKHVPVMDK                                     | 74  |
| WP_082301237.1 | 1   | -MEDSHDCLRYSRISSTAAKIHQNNDFEGFTYQDFRTHVPIMDK                                     | 74  |
| WP_007247550.1 | 1   | -MTAYDVEKEWSRISNTAAKTHQNNDFEGFTYQDFRTHVPIMDK                                     | 74  |
| WP_087674970.1 | 1   | -MTAYDVEKEWSRISNTAAKTHQNNDFEGFTYQDFRKHVPIMDK                                     | 74  |
| WP_019331605.1 | 1   | -MTAYDVEKEWSRISNTAAKTHQNNDFEGFTYQDFRKHVPIMDK                                     | 74  |
| WP_005613862.1 | 1   | -MTAYDVEKEWSRISNTAAKTHQNNDFEGFTYQDFRKHVPIMDK                                     | 74  |
| WP_004662044.1 | 1   | -----THQNNDFEGFTYQDFRTHVPIMDK                                                    | 55  |
| WP_005735201.1 | 1   | -MTAYDVEKEWSRISNTAAKTHQNNDFEGFTYQDFRKHVPIMDK                                     | 74  |
| WP_057442603.1 | 1   | -MTAYDVEKEWSRISNTAAKTHQNNDFEGFTYQDFRKHVPIMDK                                     | 74  |
| WP_195918284.1 | 27  | MISVRLKSSK-DaVSASEHNLIhnIRSRQVFSAGDVVGNLFTINLFSSLHYSACEILKHCRAHVVPVGDIALLRKSHFDF | 105 |
| WP_157940369.1 | 65  | MISVRLKSSK-DaVSASEHNLIhnIRSRQVFSAGDVVGNLFTINLFSSLHYSACEILKHCRAHVVPVGDIALLRKSHFDF | 143 |
| WP_051296966.1 | 78  | RCLYKMSGDVyNvKGCEEYNIFc-----STDTCANLFTINLLFAHCLFSNVASACGANIISIGDFNVLASEHFDA      | 148 |
| WP_034939628.1 | 75  | RCYYYYDCEV-D-EDSSRSNVF---RSNGFIQPGDRCANLFTINLFSALNNTITMMAGNCGAHVVSVDITLLTRSHFEA  | 149 |
| WP_082301237.1 | 75  | RCYYYYDCEV-D-EDSSRSNVF---RSNGFIQPGDRCANLFTINLFSALNNTITMMAGNCGAHVVSVDITLLTRSHFEA  | 149 |
| WP_007247550.1 | 75  | RCYYYYDCEV-D-EDSSRSNVF---RSNGFIQPGDRCANLFTINLFSALNNTITMMAGNCGAHVVSVDITLLTRSHFEA  | 149 |
| WP_087674970.1 | 75  | RCYYYYDCEV-D-EDSSRSNVF---RSNGFIQPGDRCANLFTINLFSALNNTITMMAGNCGAHVVSVDITLLTRSHFEA  | 149 |
| WP_019331605.1 | 75  | RCYYYYDCEV-D-EDSSRSNVF---RSNGFIQPGDRCANLFTINLFSALNNTITMMAGNCGAHVVSVDITLLTRSHFEA  | 149 |
| WP_005613862.1 | 75  | RCYYYYDCEV-D-EDSSRSNVF---RSNGFIQPGDRCANLFTINLFSALNNTITMMAGNCGAHVVSVDITLLTRSHFEA  | 149 |
| WP_004662044.1 | 56  | RCYYYYDCEV-D-EDSSRSNVF---RSNGFIQPGDRCANLFTINLFSALNNTITMMAGNCGAHVVSVDITLLTRSHFEA  | 130 |
| WP_005735201.1 | 75  | RCYYYYDCEV-D-EDSSRSNVF---RSNGFIQPGDRCANLFTINLFSALNNTITMMAGNCGAHVVSVDITLLTRSHFEA  | 149 |
| WP_057442603.1 | 75  | RCYYYYDCEV-D-EDSSRSNVF---RSNGFIQPGDRCANLFTINLFSALNNTITMMAGNCGAHVVSVDITLLTRSHFEA  | 149 |
| WP_195918284.1 | 266 | QKLPEAI-VQIVLSLSDERQEIMDIKICGLSHAVDLSELEDEICHLDPDIYEAYMKRGRINLTVVEYNALLFSPRGKCL  | 344 |
| WP_157940369.1 | 304 | QKLPEAI-VQIVLSLSAERQEIMDIKICGLSHAVDLSELEDEICHLDPDIYEAYMKRGRINLTVVEYNALLFSPRGKCL  | 382 |
| WP_051296966.1 | 308 | KFHPADISIQIHLVAKDTRVIVVWPTGIFSPQIKLLEKELTEIPDIYESQKLNQGYLVVEVDPIQYIRSVRGKIL      | 387 |
| WP_034939628.1 | 309 | QTLDRSLEIQVHLSVTDESKELVTVFVQASEVSEDERQRIEAAIAAIPDINEAYQKQGSVAVVAKDARDYAVSERGKML  | 388 |
| WP_082301237.1 | 309 | QTLGRSLEIQVHLSVTEERKELVTVFVQASEVSEDERVRIETAIADIPDIKEAYQKQGSVSVLRKRDARDYAVSERGKML | 388 |
| WP_007247550.1 | 309 | QTLGRSLEIQVHLSVTDARKELVTVFVQASEVSEDERARIETAIADIPDINEAYQKQGSVSVLRKRDARDYAVSERGKML | 388 |
| WP_087674970.1 | 309 | QTLGRSLEIQVHLSVTDERKELVTVFVQASEVSEDERTRIETAIADIPDINEAYQKQGSVSVLRKRDARDYAVSERGKML | 388 |
| WP_019331605.1 | 309 | QTLGRSLEIQVHLSVTDARKELVTVFVQASEVSEDERTRIETAIADIPDINEAYQKQGSVSVLRKRDARDYAVSERGKML | 388 |
| WP_005613862.1 | 309 | QTLGRSLEIQVHLSVTDARKELVTVFVQASEVSEDERTRIETAIADIPDINEAYQKQGSVSVLRKRDARDYAVSERGKML | 388 |
| WP_004662044.1 | 290 | QTLGRSLEMQVHLSVTEARKELVTVFVQASEVSEDERARIETAIADIPDINEAYQKQGSVSVLRKRDARDYAVSERGKML | 369 |
| WP_005735201.1 | 309 | QTLGRPLEIQVHLSVTEARKELVTVFVQASEVSEDERARIETAIADIPDINEAYQKQGSVSVLRKRDARDYAVSERGKML | 388 |
| WP_057442603.1 | 309 | QTLGRPLEIQVHLSVTEARKELVTVFVQASEVSEDERARIETAIADIPDINEAYQKQGSVSVLRKRDARDYAVSERGKML | 388 |

**Figure 3.3:** COBALT analysis of multiple sequence alignment of IaaL proteins of interest from *P. syringae* pv. *actinidiae*, *P. syringae* pv. *tomato* DC3000, *P. savastanoi* pv. *savastanoi*, with additional proteins from other species. Sequences used are in the same order as sequences of proteins from ClustalW alignment for IaaL proteins with outgroup organisms. A list of the sequence identifiers is available in the appendix. Proteins analyzed are indoleacetic acid lysine ligase protein from *Pseudomonas syringae* pv. *actinidiae*, indoleacetate-lysine ligase protein from *Pseudomonas syringae* pv. *tomato* DC3000, IaaL protein from *Pseudomonas savastanoi* pv. *savastanoi*, and IaaL proteins from *Pseudomonas* species, *Arsenophonus*, *Erwinia*, and *Escherichia*. The amino acids marked in red show the same parts of the sequences which have been aligned, meaning they are homologous and conserved among the aligned species. Dark (gray) parts of the sequence are indicators of the existing gaps in the sequences aligned. Sequences colored blue indicate polymorphism locations. The complete alignment is available in the Appendix 5.

COBALT analysis showed similar results when performed by aligning various species. Gaps and polymorphisms are present, and no conservation level is available by this analysis indicated by an asterisk (\*) and dots (: and .). What is interesting to mention is that COBALT analysis showed the existence of the 7 and 11 additional residues in the alignment of sequences that are not visible and available in the sequence format in rows 64 and 77. This might indicate

possible different protein conformation, but it will be tested further on. The red color of the alignment indicates homologous sequences, so the level of conservation is the same as in the previous analysis.

For additional analysis and validation of the alignment of different species, and to see if we can check for the aforementioned 7 and 11 residues of sequences that might influence the protein, we used the T-Coffee tool for multiple sequence alignment. The results are presented in 3.4.

T-coffee analysis did not reveal the additional residues we wanted to check, but the conservation level is matching with the previous analyzes, and our results were confirmed by different tools. Since not all tools enable naming the sequences by specific identifiers but only by original sequence accession number, we made a table with the list of all protein sequences used in the analysis.

|                |                                                                    |
|----------------|--------------------------------------------------------------------|
| WP_195918284.1 | MT-----RITAITGCAFDYTAFRDSVRISSHVYSTTGDSGSLTQRDPAA--QRDPAA-         |
| WP_157940369.1 | MQ-P-----RITAITGCAFDYTAFRDSVRISSHVYSTTGDSGSLTQRDPAA--Y-LAKNDQSLHLL |
| WP_051296966.1 | MSIYKEEMVWNTVKDKIATLFLPAISQWYQQFSQSIQVMEKKG----F-TAQTERCVAE-       |
| WP_034939628.1 | MS-AYDVEKEWLRVAGIAAETHQDADLRGFTYQDFRKWVPVMDKDG----F-TAQTERCVAE-    |
| WP_057442603.1 | MT-AYDVEKEWSRISNTAAKTHQNNDFEGFTYQNFVKHVPIMDKG----F-TAQTERCLER-     |
| WP_057442603.1 | MT-AYDVEKEWSRISNTAAKTHQNNDFEGFTYQNFVKHVPIMDKG----F-TAQTERCLER-     |
| WP_005735201.1 | MT-AYDVEKEWSRISNTAAKTHQNNDFEGFTYQNFVKHVPIMDKG----F-TAQTERCLER-     |
| WP_004662044.1 | -----THQNNDFEGFTYQDFRTHVPIMDKG----F-TAQTERCLER-                    |
| WP_005613862.1 | MT-AYDVEKEWSRISNTAAKTHQNNDFEGFTYQDFRKHVPIMDKG----F-AAQTERCLER-     |
| WP_019331605.1 | MT-AYDVEKEWSRISNTAAKTHQNNDFEGFTYQDFRKHVPIMDKG----F-AAQTERCLER-     |
| WP_087674970.1 | MT-AYDVEKEWSRISNTAAKTHQNNDFEGFTYQDFRKHVPIMDKG----F-AAQTERCLER-     |
| WP_007247550.1 | MT-AYDVEKEWSRISNTAAKTHQNNDFEGFTYQDFRTHVPIMDKG----F-AAQTERCLER-     |
| WP_082301237.1 | ME-DSHCLRYSRISSTAANKIHQNNDFEGFTYQDFRTHVPIMDKG----F-AAQTERCLER-     |
| cons           | ::                                                                 |
| WP_195918284.1 | EQ-IPMLLGATSGTTGAMKMSVRLKSSKDA--VSASEHNLIHNLRSRQVFSAGDVVGNLFTI     |
| WP_157940369.1 | EQ-IPMLLGATSGTTGAMKMSVRLKSSKDA--VSASEHNLIHNLRSRQVFSAGDVVGNLFTI     |
| WP_051296966.1 | NSKNSYLIGFTSGTGLKRCLYKMSGDVYVNVKGCE-EY----NI-----FCSTDTCANLFTI     |
| WP_034939628.1 | NE-RNCLIGFTSGTSGNLKRCYYYDDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI    |
| WP_057442603.1 | NE-RNCLIGFTSGTSGNLKRCYYYDDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI    |
| WP_057442603.1 | NE-RNCLIGFTSGTSGNLKRCYYYDDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI    |
| WP_005735201.1 | NE-RNCLIGFTSGTSGNLKRCYYYDDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI    |
| WP_004662044.1 | NE-RNCLIGFTSGTSGNLKRCYYYDDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI    |
| WP_005613862.1 | NE-RNCLIGFTSGTSGNLKRCYYYDDCEVDE--DSS-RS---SVFRSNGFIQPGDRCANLFTI    |
| WP_019331605.1 | NE-RNCLIGFTSGTSGNLKRCYYYDDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI    |
| WP_087674970.1 | NE-RNCLIGFTSGTSGNLKRCYYYDDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI    |
| WP_007247550.1 | NE-RNCLIGFTSGTSGNLKRCYYYDDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI    |
| WP_082301237.1 | NE-RNCLIGFTSGTSGNLKRCYYYDDCEVDE--DSS-LS---NVFRSNGFILPGDRCANLFTI    |
| cons           | ::** *****: . . . : . *                                            |
| WP_195918284.1 | EKIARMINQKLPEAI-VQIVLSLSDERQEIMDIKICGLSHAVDLSELEDEICHLPDIYEAYM     |
| WP_157940369.1 | EKIARMINQKLPEAI-VQIVLSLSAERQEIMDIKICGLSHAVDLSELEDEICHLPDIYEAYM     |
| WP_051296966.1 | ETIVAVVHKFHPADISIQIHLVAKDTREVITVWVPTGIFSPQIKLLEKELTEIPDIYESKQ      |
| WP_034939628.1 | SIIQQAIIKQTLDRSLEIQVHLSVTDESKELVTVFVPATEVSEDERARIEAIAIPDINEAYQ     |
| WP_057442603.1 | GIIQQTIIKQTLGRPLEIQVHLSVTEARKELVTVFVQASEVDEEERARIETAIADIPDINEAYQ   |
| WP_057442603.1 | GIIQQTIIKQTLGRPLEIQVHLSVTEARKELVTVFVQASEVDEEERARIETAIADIPDINEAYQ   |
| WP_005735201.1 | GIIQQAIIKQTLGRPLEIQVHLSVTEARKELVTVFVQASEVDEDERARIETAIADIPDINEAYQ   |
| WP_004662044.1 | GIIQQAIIKQTLGRSLEMQVHLSVTEARKELVTVFVQASEVDEDERARIETAIADIPDINEAYQ   |
| WP_005613862.1 | GIIQQAIIKQTLGRSLEIQVHLSVTDARKELVTVFVQASEVDEDERTRIETAIADIPDINEAYQ   |
| WP_019331605.1 | GIIQQAIIKQTLGRSLEIQVHLSVTDARKELVTVFVQASEVDEDERTRIETAIADIPDINEAYQ   |
| WP_087674970.1 | GIIQQAIIKQTLGRSLEIQVHLSVTDARKELVTVFVQASEVDEDERTRIETAIADIPDINEAYQ   |
| WP_007247550.1 | GIIQQAIIKQTLGRSLEIQVHLSVTDARKELVTVFVQASEVDEDERARIETAIADIPDINEAYQ   |
| WP_082301237.1 | GIIQQTIIKQTLGRSLEIQVHLSVTEERKELVTVFVQASEVDEDERVRIETAIADIPDIKEAYQ   |
| cons           | * :: : *: **: : : : : : * : :*** *                                 |

**Figure 3.4:** MSA by T-Coffee of *iaaL* gene-encoded IaaL proteins from *Pseudomonas*, *Erwinia*, *Arsenophonus*, and *Escherichia* species. Sequences used are in the same order as sequences of proteins from ClustalW alignment for IaaL proteins with outgroup organisms. A list of the sequence identifiers is available in the appendix. Proteins analyzed are indoleacetic acid lysine ligase protein from *Pseudomonas syringae* pv. *actinidiae*, indoleacetate-lysine ligase protein from *Pseudomonas syringae* pv. *tomato* DC3000, IaaL protein from *Pseudomonas savastanoi* pv. *savastanoi*, and IaaL proteins from *Pseudomonas* species, *Erwinia*, *Arsenophonus*, and *Escherichia*. Under the sequences are symbols indicating the conservation level of the sequences. Asterisk (\*) indicates that the residue is the same across all query proteins; a colon (:) indicates that the residues are different but highly similar across query proteins, and a dot (.) indicates that residues are different but similar to some degree, across query proteins.

**Table 3.4:** Accessions for the MSA and T-coffee for IaaL and MatE proteins.

| Accession      | Label in phylogenetic tree |
|----------------|----------------------------|
| WP_003380685.1 | P.syringae                 |
| WP_005737399.1 | P.syringae_group           |
| WP_007247550.1 | P.syringae_groupNR         |
| WP_024667964.1 | P.syringae_groupNR2        |
| WP_044322280.1 | P.syringae_groupNR3        |
| WP_057433268.1 | P.syringae_groupNR4        |
| WP_057456329.1 | P.savastanoi               |
| WP_087674970.1 | P.syringae_NR              |
| WP_122285283.1 | P.syringae_genomosp3       |
| AXK59710.1     | P.syringae_actinidae       |
| AAO53915.1     | P.syringae_tomatoDC3000    |
| EEB56700.1     | P.syringae_tomatoT1        |
| RMS74473.1     | P.savastanoi1              |
| WP_195918284.1 | Salmonella                 |
| WP_157940369.1 | Escherichia                |
| WP_051296966.1 | Arsenophonus               |
| WP_034939628.1 | Erwinia                    |
| WP_082301237.1 | P.savastanoi               |
| WP_019331605.1 | P.syringae                 |
| WP_005613862.1 | P.avellanae                |
| WP_004662044.1 | P.savastanoi_NR            |
| WP_005735201.1 | P.amygdali                 |
| WP_057442603.1 | P.savastanoi_NR1           |

Together with the sequence alignment, we got the percent identity matrix of the protein sequences we analyzed. Besides checking for the visual comparison of sequences, the percent identity matrix helps us to see the percentage similarities between aligned sequences, thus providing an additional way of sequence analysis and homology. Percentage identity gives us reliable information on the matching level of sequences used, so we can determine which sequences to continue working on, and which are the closest to our proteins of interest.

### 3.2.4. Identity Matrix

Percent identity matrix for *iaaL* gene-encoded proteins is listed in 3.5.

Percent identity matrix shows the similarity between the proteins of interest from the *iaaL* genes, indoleacetic acid lysine ligase protein from *Pseudomonas syringae* pv. *actinidiae*, indoleacetate-lysine ligase protein from *Pseudomonas syringae* pv. *tomato* DC3000, and IaaL protein from *Pseudomonas savastanoi* pv. *savastanoi*, as well as from the additionally used protein sequences from *Pseudomonas* species for the better comparison.

**Table 3.5:** Percent Identity Matrix table for proteins encoded by *iaaL* genes by ClustalW. Proteins compared and listed are *Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* DC3000, and *Pseudomonas savastanoi* pv. *savastanoi* for *iaaL* gene.

|                                                                                               | Indoleacetic acid<br>lysine ligase ( <i>P. s.</i> pv.<br><i>actinidiae</i><br>WP_003380685.1) | Indoleacetate-lysine<br>ligase ( <i>P. s.</i> pv. <i>tomato</i><br>DC3000;<br>WP_007247550.1) | IaaL ( <i>P. s.</i> pv.<br><i>savastanoi</i><br>WP_057456329.1) |
|-----------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|-----------------------------------------------------------------|
| Indoleacetic acid<br>lysine ligase ( <i>P. s.</i> pv.<br><i>actinidiae</i><br>WP_003380685.1) | 100.00                                                                                        | 97.45                                                                                         | 97.20                                                           |
| Indoleacetate-lysine<br>ligase ( <i>P. s.</i> pv. <i>tomato</i><br>DC3000;<br>WP_007247550.1) | 97.45                                                                                         | 100.00                                                                                        | 97.20                                                           |
| IaaL ( <i>P. s.</i> pv.<br><i>savastanoi</i><br>WP_057456329.1)                               | 97.20                                                                                         | 97.20                                                                                         | 100.00                                                          |

**Table 3.6:** Extended Identity Matrix table for IaaL proteins by ClustalW. The outgroup organisms are included. Accession numbers are available in 3.4. The percentage identity we got by the matrix shows differences between the proteins analyzed. Outgroup organisms have less identity than the organisms from the *Pseudomonas* species.

|              |       |       |       |       |       |       |       |       |       |       |       |       |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Salmonella   | 100   | 99.43 | 32.46 | 34.39 | 36.42 | 36.71 | 36.71 | 36.42 | 36.13 | 36.71 | 36.71 | 36.42 |
| Escherichia  | 99.43 | 100   | 31.1  | 32.89 | 35.54 | 35.28 | 35.28 | 35.01 | 34.75 | 35.92 | 35.54 | 35.28 |
| Arsenophonus | 32.46 | 31.1  | 100   | 39.39 | 38.36 | 39.64 | 39.64 | 39.9  | 39.64 | 40.05 | 39.39 | 39.39 |
| Erwinia      | 34.39 | 32.89 | 39.39 | 100   | 81.01 | 86.33 | 86.33 | 86.58 | 86.58 | 86.7  | 85.57 | 85.32 |
| P.savastanoi | 36.42 | 35.54 | 38.36 | 81.01 | 100   | 90.89 | 90.38 | 90.13 | 90.13 | 93.62 | 90.89 | 91.14 |
| P.syringae   | 36.71 | 35.28 | 39.64 | 86.33 | 90.89 | 100   | 96.96 | 98.23 | 97.97 | 97.34 | 96.96 | 96.71 |
| _group       |       |       |       |       |       |       |       |       |       |       |       |       |
| P.syringae   | 36.71 | 35.28 | 39.64 | 86.33 | 90.38 | 96.96 | 100   | 98.48 | 97.97 | 96.01 | 95.95 | 95.7  |
| _NR          |       |       |       |       |       |       |       |       |       |       |       |       |
| P.syringae   | 36.42 | 35.01 | 39.9  | 86.58 | 90.13 | 98.23 | 98.48 | 100   | 98.73 | 96.28 | 96.46 | 96.2  |
| P.avellanae  | 36.13 | 34.75 | 39.64 | 86.58 | 90.13 | 97.97 | 97.97 | 98.73 | 100   | 96.28 | 96.2  | 95.95 |
| P.savastanoi | 36.71 | 35.92 | 40.05 | 86.7  | 93.62 | 97.34 | 96.01 | 96.28 | 96.28 | 100   | 97.61 | 97.34 |
| _NR          |       |       |       |       |       |       |       |       |       |       |       |       |
| P.amygdali   | 36.71 | 35.54 | 39.39 | 85.57 | 90.89 | 96.96 | 95.95 | 96.46 | 96.2  | 97.61 | 100   | 99.24 |
| P.savastanoi | 36.42 | 35.28 | 39.39 | 85.32 | 91.14 | 96.71 | 95.7  | 96.2  | 95.95 | 97.34 | 99.24 | 100   |
| _NR1         |       |       |       |       |       |       |       |       |       |       |       |       |

Percent identity matrix shows a different level of the identity of our sequences, indicating the possible protein three-dimensional difference which might influence the complete process of toxicity and disease development. Following the results we got, our next step was a phylogenetic tree which shows the evolutionary distance between the organisms so we can see which proteins are more related to our proteins of interest.

### 3.2.5. Phylogenetic Tree

Phylogenetic trees give us insight into the lines of evolutionary descents of the proteins from species of interest that have a common ancestor. The patterns of branching in the tree show us the highest similarity among the targeted species from common ancestors.

The first phylogenetic tree is for the IaaL proteins of *Pseudomonas* family with the outgroup organisms, so the genomic distance can be seen. All of the aligned organisms contain the

IaaL protein.

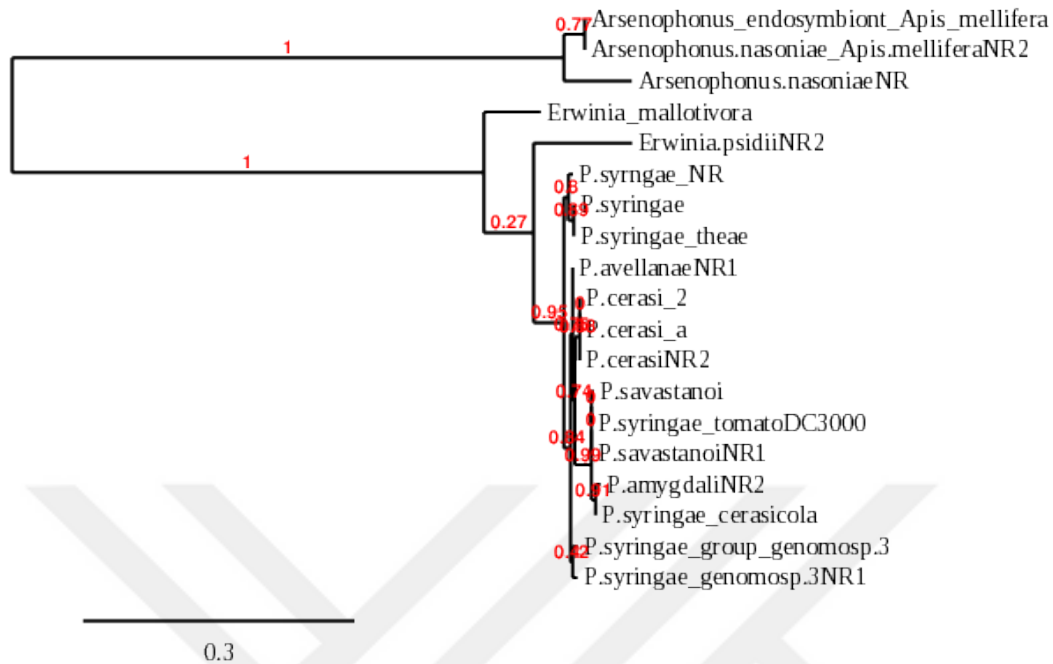


**Figure 3.5:** Phylogenetic tree for the IaaL proteins of *Pseudomonas* family including outgroups of *Arsenophonus* and *Erwinia* genus. IaaL proteins are present in all the species used for the phylogenetic tree construction, but the genetic and evolutionary distance between the organisms is high. *Erwinia psidii* is more closely related to the *Pseudomonas* family but still belongs to the outgroup organisms. The branch length is proportional to the number of substitutions per site. The numbers in red are confidence levels and represent the statistical significance for the node positioning as a distance from the root of the tree between the organisms compared. It also represents the number of differences between sequences.

We did a phylogenetic tree with the outgroup organisms which have the same IaaL protein to see the distance between them. It was interesting to see, because IaaL protein is present in the organisms outside the *Pseudomonas* species, and might have an influence on the toxicity as well in their host organisms. To have a better insight into *Pseudomonas* species, we did another phylogenetic tree only for *Pseudomonas* and *Erwinia* as the closest outgroup organism that has an IaaL protein.

As we can see in the 3.6, there is a differentiation inside the species, and branching distance between more similar and distant organisms. This was an additional reason to take IaaL proteins from these species as well for the analysis, to search for the sequence difference from our three species of interest with them, and thus analyze further until we answer the question of influence on toxicity. List of all the species from *Pseudomonas* we used in the analysis and their accession numbers are listed in the 3.7 and in the Appendix 12.

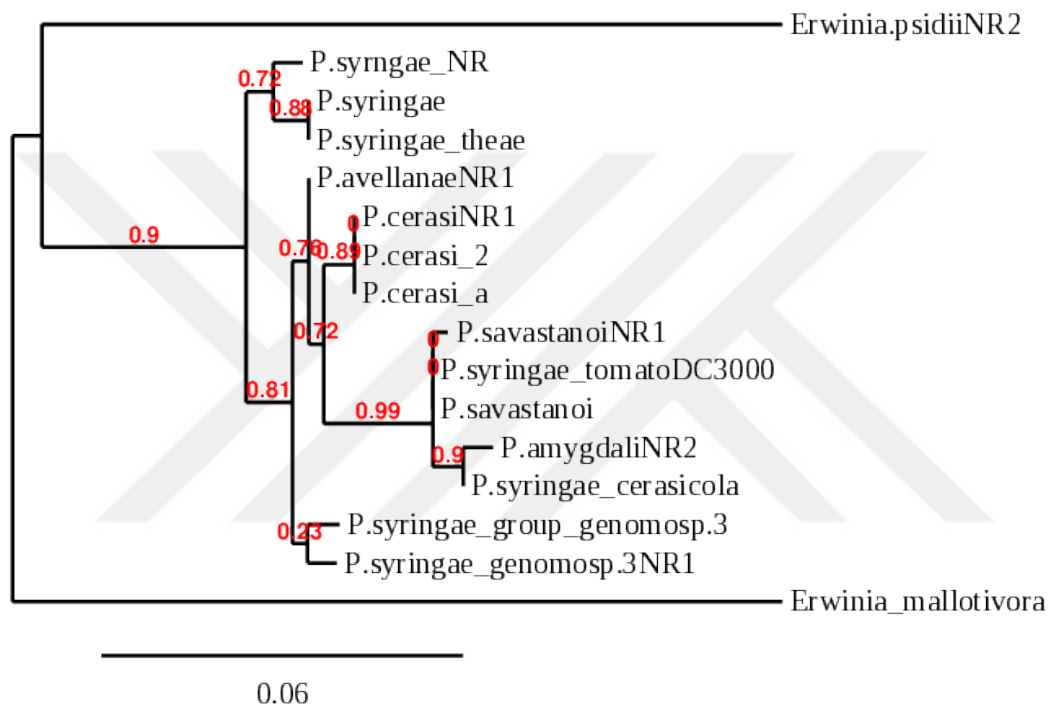


**Figure 3.6:** Phylogenetic tree of IaaL proteins from *Pseudomonas* family, including *Erwinia mallotivora* as only outgroup. Distances between the *Pseudomonas* members according to the IaaL proteins are small, indicating their common or close origin, as they belong to the same species. The branch length is proportional to the number of substitutions per site. The numbers in red are confidence intervals and represent the statistical significance for the node positioning as a distance from the root of the tree between the organisms compared. It also represents the number of differences between sequences.

**Table 3.7:** Code names of the IaaL proteins for the T-coffee and ClustalW analysis.

| Accession             | Label in tree                  | Scientific name                                             |
|-----------------------|--------------------------------|-------------------------------------------------------------|
| <b>WP_003380685.1</b> | <b>P.syringae</b>              | <b><i>Pseudomonas syringae</i> pv. <i>actinidiae</i></b>    |
| WP_005737399.1        | P.syringae_group               | <i>Pseudomonas syringae</i> group                           |
| <b>WP_007247550.1</b> | <b>P.syringae_tomatoDC3000</b> | <b><i>Pseudomonas syringae</i> pv. <i>tomato</i> DC3000</b> |
| WP_024667964.1        | P.syringae_groupNR2            | <i>Pseudomonas syringae</i> group                           |
| WP_044322280.1        | P.syringae_groupNR3            | <i>Pseudomonas syringae</i> group                           |
| WP_057433268.1        | P.syringae_groupNR4            | <i>Pseudomonas syringae</i> group                           |
| <b>WP_057456329.1</b> | <b>P.savastanoi</b>            | <b><i>Pseudomonas savastanoi</i></b>                        |
| WP_087674970.1        | P.syringae_NR                  | <i>Pseudomonas syringae</i>                                 |
| WP_122285283.1        | P.syringae_genomosp3           | <i>Pseudomonas savastanoi</i>                               |
| AXK59710.1            | P.syringae_actinidae           | <i>Pseudomonas syringae</i> pv. <i>actinidiae</i>           |
| AAO53915.1            | P.syringae_tomatoDC3000        | <i>Pseudomonas syringae</i> pv. <i>tomato</i> str. DC3000   |
| EEB56700.1            | P.syringae_tomatoT1            | <i>Pseudomonas syringae</i> pv. <i>tomato</i> T1            |
| RMS74473.1            | P.savastanoi1                  | <i>Pseudomonas savastanoi</i>                               |
| WP_195918284.1        | Salmonella                     | <i>Salmonella enterica</i>                                  |
| WP_157940369.1        | Escherichia                    | <i>Escherichia coli</i>                                     |
| WP_051296966.1        | Arsenophonus                   | <i>Arsenophonus nasoniae</i>                                |
| WP_034939628.1        | Erwinia                        | <i>Erwinia mallotivora</i>                                  |

Phylogenetic trees show us differentiation between the organisms, so our next step was to check the three-dimensional structure of the proteins, to have a better insight into the protein structure and its functionality.

### 3.2.6. 3D Structure Prediction

Three-dimensional protein structure prediction requires understanding the interaction between amino acid bonds, protein folding techniques, and different aspects of the functional properties

of the proteins. The function of the protein is directly dependent on its three-dimensional form since it represents its structure formed by many different bonds (interactions) between R groups of amino acids which make the upper side chains of the protein. Sidechains can interact with other molecules, forming different compounds with various functions. Prediction of the 3D form of the protein represents an important step in the understanding of novel proteins and species, their regulations, and possible therapeutic settings when it is to the disease-causing bacteria.

For the sake of research and additional results, the same analysis by the Swiss model is done for the Indoleacetic acid lysine ligase protein of *Pseudomonas syringae* pv. *actinidiae* as for the MatE protein from different organisms. By having different models it is possible to see similarities, and differences and draw possible function similarities of the proteins. As aforementioned, the *iaaL* gene and consecutive protein are starting points of the detailed analysis of the *matE* gene and gene-encoded protein of *Pseudomonas syringae* pv. *actinidiae*. Thus, by understanding the *iaaL* gene and protein and its structure, interactions and promoter region which is common for both genes, we can better understand and analyze the *matE* gene and protein and fulfill the aim of this thesis.

It is important to mention that we analyzed IaaL and MatE protein secondary structure sequences and found a difference between them. Thus, our assumption was that we will have a different 3D structure as well, in terms of the number of helices, loops, and beta strands. In fact, we got visually similar proteins consisting of the many alpha helices interconnected by loops, with the beta strand which are not present in the MatE proteins. The protein we create by 3D modeling is compared with other sequences and the best available template is automatically found by the server. This analysis is important for the visual comparison of sequences, even though these two proteins have different functions. All of the MATE efflux proteins have an alpha-helical structure organized in a tube-like formation and are interconnected by loops. This cross-species analysis shows that our protein of interest conserves its structural formation across the species, indicating the same function and influence on the toxicity in the species we analyzed. It is important to have in mind that the 3D protein model has been built according to the best possible template found, so it is the best prediction we could get, not the 100% accurate model of the protein.

We followed the logic of the research and after making a 3D model of the IaaL protein and its

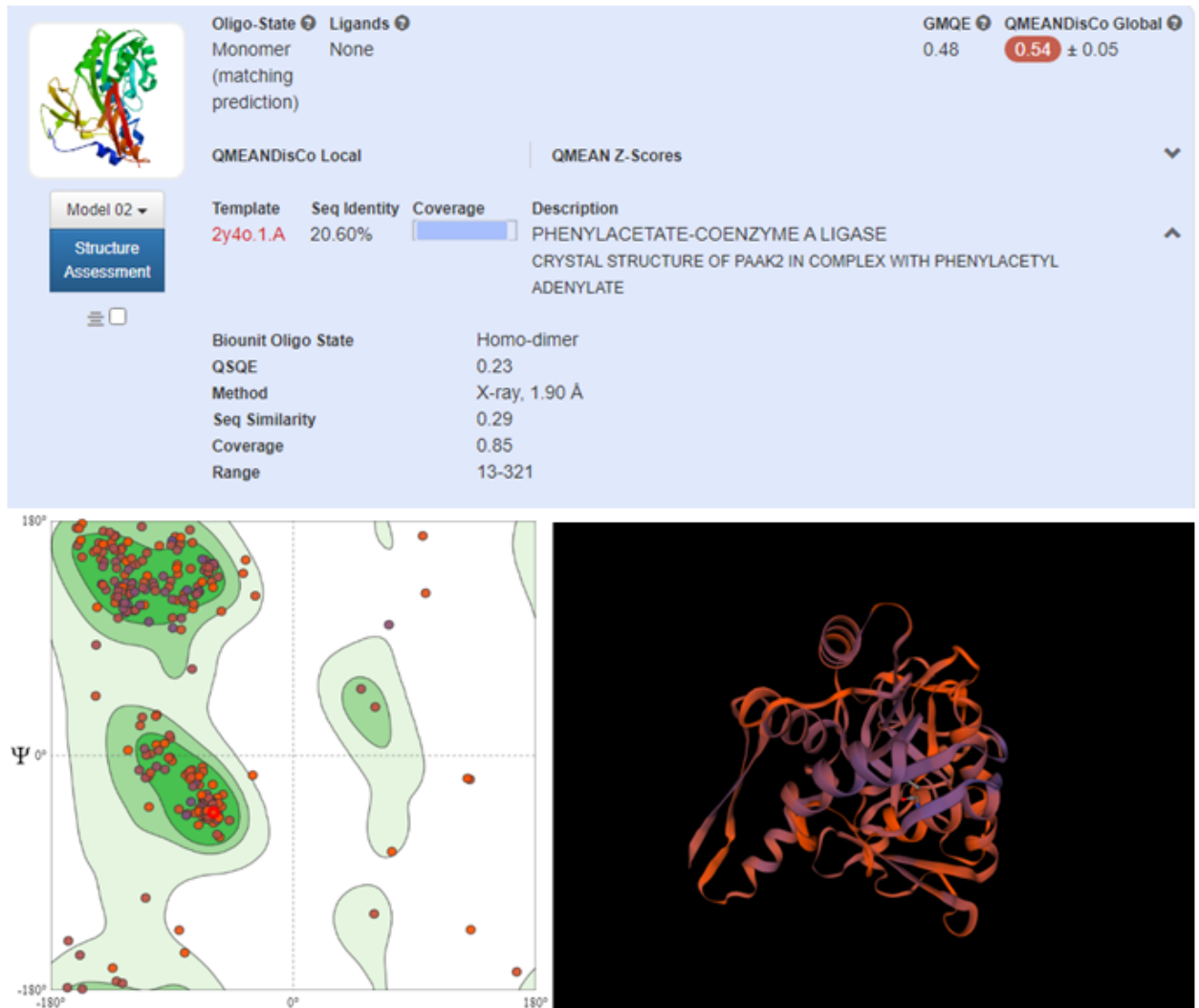


**Figure 3.7:** Swiss-model and Ramachandran plot of the Indoleacetic acid lysine ligase 3D protein from *Pseudomonas syringae* pv. *actinidiae* (WP\_003380685.1). The 3D model shows alpha-helices connected with the beta-sheets, monomers with two types of structural units. The global model quality estimate value (GMQA) is low due to the template's low match value. Amino acid residues are shown as red and purple dots, corresponding to the 3D protein image on the right side. Residues that are more distantly located from other red/purple dots are assigned residues at the end of the alpha helices and usually have the angle values nontypical for the alpha helix. Red dots which are more concentrated in the green area are residues having angle values typical for the alpha helix.

analysis, the next step was the analysis of MatE protein in our species of interest, as well as in the other organisms containing the mentioned protein.

### **3.3. MATE**

#### **3.3.1. Full Sequences of the Proteins**

Once we got the sequences of MatE and IaaL proteins, we validated them by doing a pairwise sequence alignment. We checked if the sequences of MatE protein were only in the respective organisms, and if the sequences were complete.



**Table 3.8:** Protein homology of partial and full sequences of MatE proteins. The first three results of the search for homologs of each partial protein AXK59714.1, AAO53914.1, and RMS74471.1 (multidrug transporter MatE protein, also described as membrane protein) from *Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* str. DC3000 and *Pseudomonas savastanoi* pv. *savastanoi*. There are no gaps in any of these alignments.

| Query      | Subject        | Taxonomy                                                | Pct identity | Mis match | Query coverage |
|------------|----------------|---------------------------------------------------------|--------------|-----------|----------------|
| AXK59714.1 | WP_003380687.1 | <i>Pseudomonas syringae</i>                             | 100          | 0         | 100            |
| AXK59714.1 | WP_017700567.1 | <i>Pseudomonas syringae</i>                             | 99.292       | 3         | 99.76          |
| AXK59714.1 | WP_017710541.1 | <i>Pseudomonas syringae</i>                             | 99.528       | 2         | 99.76          |
| AAO53914.1 | WP_007247551.1 | <i>Pseudomonas syringae</i> group                       | 100          | 0         | 100            |
| AAO53914.1 | WP_230692474.1 | <i>Pseudomonas syringae</i> group<br><i>genomosp. 3</i> | 99.782       | 1         | 99.78          |
| AAO53914.1 | WP_005763530.1 | <i>Pseudomonas amygdali</i>                             | 99.782       | 1         | 99.78          |
| RMS74471.1 | WP_005782319.1 | <i>Pseudomonas syringae</i> group                       | 100          | 0         | 100            |
| RMS74471.1 | WP_054995848.1 | <i>Pseudomonas syringae</i> group                       | 99.782       | 1         | 100            |
| RMS74471.1 | WP_019742037.1 | <i>Pseudomonas savastanoi</i>                           | 99.782       | 1         | 99.78          |

BLAST compares two sequences to determine their homology, similarity, and evolutionary relatedness. We did the same analysis for the IaaL proteins to find the complete protein sequences since those we firstly found are PCR products and are not reliable for our analysis. We used BLAST to find homologous sequences, and the full sequences we got are presented in the table {#tbl:R1.1}. Once we got the full sequences, we were able to continue the analysis. We can see that sequences are present in different organisms, but all of them belong

to the *Pseudomonas* species. Even so, our focus remains on the *Pseudomonas syringae* pv. *actinidiae*, and *Pseudomonas syringae* pv. *tomato* DC3000, and *Pseudomonas savastanoi* pv. *savastanoi*.

The main result of the homology search of the MatE proteins is the new sequences that extend the original ones.

Due to the complete protein sequences, we had to do a multiple sequence alignment to check for their identity and conservation level. We did three MSA analyzes, ClustalW, COBALT, and T-coffee, as in the case of IaaL proteins.

### 3.3.2. MSA Results for matE

The following ClustalW analysis is for the MatE proteins encoded by various species. The focus is on the MatE efflux protein from *Pseudomonas syringae* pv. *actinidiae*, its homologs in *Pseudomonas syringae* pv. *tomato* DC3000 and *Pseudomonas savastanoi* pv. *savastanoi*, and the matE protein detoxification from *Glycine max* (Soybean) as an outgroup organism containing the MatE protein. We decided to focus on these three species and MatE proteins since the starting point of this research was articles mentioning the existence of these proteins correlated with the disease occurrence in tomato and kiwi species. We could not find any additional information on this correlation so we decided to analyze the mentioned proteins. The results of ClustalW MSA analysis are presented in 3.8.

Including the *Glycine max* as an outgroup organism resulted in the gaps and low conservation level as the final outcome of the analysis. Our first assumption was that *Glycine max* does not contain the MatE protein homolog to the *Pseudomonas* at all, but further analysis proved us wrong, even though there are many differences in the sequence structure.

Our next analysis was performed by a COBALT tool, excluding the *Glycine max* outgroup organism, so we can see the conservation level and homology of only *Pseudomonas* species. The results we got are presented in the following 3.9.

MSA COBALT analysis of *Pseudomonas* species shows us high conservation levels of the protein sequences, with the difference in the start region with several polymorphisms. Overall protein structure is conserved, but we wanted to check also for the polymorphisms visible in the 3D structure of the sequence between different organisms. The first step further was a T-coffee MSA analysis of the MatE protein sequences including the *Glycine max* species.

**Figure 3.8:** ClustalW multiple protein sequence alignments for the MatE protein from *P. syringae* pv. *actinidiae*, *P. syringae* pv. *tomato* DC3000, *P. savastanoi* pv. *savastanoi* and *Glycine max* as a focus with additional *Pseudomonas* species. Alignment has been done for the MatE efflux family protein from *Pseudomonas syringae* pv. *actinidiae*, MATE efflux protein from *Pseudomonas syringae* pv. *tomato* DC3000, MATE efflux protein from *Pseudomonas savastanoi* pv. *savastanoi*, and matE protein detoxification from *Glycine max* (Soybean). Sequences compared are listed by order of SOYBN for Soybean matE protein sequence, mate for *P. s.* pv. *actinidiae*, AXK59714.1 for *P. s.* pv. *tomato* DC3000, and MATE for *P. s.* pv. *savastanoi*. Many gaps are present at the beginning of the sequence with a very low conservation level. *Glycine max* is an outgroup organism and only contains a conserved domain part of the protein, the same domain as the *Pseudomonas* species. The complete sequence is available in the 2.

|                                |     |                         |                             |                                      |     |
|--------------------------------|-----|-------------------------|-----------------------------|--------------------------------------|-----|
| <a href="#">WP_003380687.1</a> | 1   | MAVTKAVFTEGSVLRHVVFMTGT | SATSLLSIFLIEVLTVVYIAMLRDDSL | LAFAVAKTLMFFLISMILGIAVAVST           | 77  |
| <a href="#">WP_017700567.1</a> | 1   | MAVTKAVFIEGSVLRHVVFMTGT | SATSLLSIFLIEVLTVVYIAMLRDDSL | LAFAVAKTLMFFLISMILGIAVAVST           | 77  |
| <a href="#">WP_017710541.1</a> | 1   | MAVTKAVFTEGSVLRHVVFMTGT | SATSLLSIFLIEVLTVVYIAMLRDDSL | LAFAVAKTLMFFLISMILGIAVAVST           | 77  |
| <a href="#">WP_007247551.1</a> | 1   | MAVTKAVFTEGSVLRHVVFMTST | SATSLLSIFLIEVLTVVYIAMLRDDSL | LAFAVAKTLMFFIISMILGIAVAVSS           | 77  |
| <a href="#">WP_230692474.1</a> | 1   | MAVTKAVFTEGSVLRHVVFMTST | SATSLLSIFLIEVLTVVYIAMLRDDSL | LAFAVAKTLMFFIISMILGIAVAVSS           | 77  |
| <a href="#">WP_005763530.1</a> | 1   | MAVTKAVFTEGSVLRHVVFMTST | SATSLLSIFLIEVLTVVYIAMLRDDSL | LAFAVAKTLMFFIISMILGIAVAVSS           | 77  |
| <a href="#">WP_005782319.1</a> | 1   | MAVTKAVFTKGSMLRHVVFMST  | SATSLLSIFLIEVLTVVYIAMLRDDSL | LAFAVAKTLMFFLISMILGIAVAVST           | 77  |
| <a href="#">WP_054995848.1</a> | 1   | MAVTKAVFTKGSMLRHVVFMST  | SATSLLSIFLIEVLTVVYIAMLRDDSL | LAFAVAKTLMFFLISMILGIAVAVST           | 77  |
| <a href="#">WP_019742037.1</a> | 1   | MAVTKAVFTKGSMLRHVVFMST  | SATSLLSIFLIEVLTVVYIAMLRDDSL | LAFAVAKTLMFFLISMILGIAVAVST           | 77  |
| <a href="#">WP_003380687.1</a> | 78  | AVSRSLGSAAPDQARRLASSG   | LLLGIFATVCGAALELVLDPAIRML   | GGEGNSFTAAREYLWLTLPASVLVAVELLCVQI    | 157 |
| <a href="#">WP_017700567.1</a> | 78  | AVSRSLGSAAPDQARRLASSG   | LLLGILATVCGAALELVLDPAIRML   | GGEGNSFTAAREYLWLTLPASVLVAVELLCVQI    | 157 |
| <a href="#">WP_017710541.1</a> | 78  | AVSRSLGSAAPDQARRLASSG   | LLLGIFATVCGAALELVLDPAIRML   | GGEGNSFTAAREYLWLTLPASVLVAVELLCVQI    | 157 |
| <a href="#">WP_007247551.1</a> | 78  | TVSRALGSGVPDQARRLASSG   | LLLGIIATVCCAVLELIFLDSAIQML  | GGEGSSFTAAREYLWLTLPASVLVAVEQLCVQI    | 157 |
| <a href="#">WP_230692474.1</a> | 78  | TVSRALGSGVPDQARRLASSG   | LLLGIIATVCCAVLELIFLDSAIQML  | GGEGSSFTAAREYLWLTLPASVLVAVEQLCVQI    | 157 |
| <a href="#">WP_005763530.1</a> | 78  | TVSRALGSGVPDQARRLPSSG   | LLLGIIATVCCAVLELIFLDSAIQML  | GGEGSSFTAAREYLWLTLPASVLVAVEQLCVQI    | 157 |
| <a href="#">WP_005782319.1</a> | 78  | TVSRALGSGVPDQARRLASSG   | VLLGSMATVCGAALELVLDQAIRML   | GAGNSFTAAREYLWLTLPASVLVAVEQLCVQI     | 157 |
| <a href="#">WP_054995848.1</a> | 78  | TVSRALGSGVPDQARRLASSG   | VLLGSMATVCGAALELVLDQAIRML   | GAGNSFTAAREYLWLTLPASVLVAVEQLCVQI     | 157 |
| <a href="#">WP_019742037.1</a> | 78  | TVSRALGSGVPDQARRLASSG   | VLLGSMATVCGAALELVLDQAIRML   | GAGNSFTAAREYLWLTLPASVLVAVEQLCVQI     | 157 |
| <a href="#">WP_003380687.1</a> | 158 | LRSAGFGKQAMWILLTSTA     | AVAIADPLLIIVFDLGLVGAGIAYLI  | SSLVSACLGFYYVHRVARLTSGVSLTHLSGDISII  | 237 |
| <a href="#">WP_017700567.1</a> | 158 | LRSAGFGKQAMWILLTSTA     | AVAIADPLLIIVFDLGLVGAGIAYLI  | SSLVSACLGFYYVHRVARLTSGVSLTHLSGDISII  | 237 |
| <a href="#">WP_017710541.1</a> | 158 | LRSAGFGKQAMWILLTSTA     | AVAIADPLLIIVFDLGLVGAGIAYLI  | SSLVSACLGFYYVHRVARLTSGVSLTHLSGDISII  | 237 |
| <a href="#">WP_007247551.1</a> | 158 | LRSAGFGKQAMWTLTATA      | AVAIADPLLIIVFDLGLAGAGVAYLI  | SSLVSACLGFYYVHRVARLTCTRVSLTHLSGDISII | 237 |
| <a href="#">WP_230692474.1</a> | 158 | LRSAGFGKQAMWTLTATA      | AVAIADPLLIIVFDLGLAGAGVAYLI  | SSLVSACLGFYYVHRVARLTCTRVSLTHLSGDISII | 237 |
| <a href="#">WP_005763530.1</a> | 158 | LRSAGFGKQAMWTLTATA      | AVAIADPLLIIVFDLGLAGAGVAYLI  | SSLVSACLGFYYVHRVARLTCTRVSLTHLSGDISII | 237 |
| <a href="#">WP_005782319.1</a> | 158 | LRSAGFGKQAMWTLTATA      | AVAIADPLLIIVFDLGLVGAGIAYLI  | SSLVSACLGFYYVHRVAHLTCTRVSLKNLSGDIRNI | 237 |
| <a href="#">WP_054995848.1</a> | 158 | LRSAGFGKQAMWTLTATA      | AVAIADPLLIIVFDLGLVGAGIAYLI  | SSLVSACLGFYYVHRVAHLTCTRVSLKNLSGDIRNI | 237 |
| <a href="#">WP_019742037.1</a> | 158 | LRSAGFGKQAMWTLTATA      | AVAIADPLLIIVFDLGLVGAGIAYLI  | SSLVSACLGFYYVHRVAHLTCTRVSLKNLSGDIRNI | 237 |

**Figure 3.9:** COBALT analysis for multiple sequence alignment of the proteins of interest for MatE proteins from *P. syringae* pv. *actinidiae*, *P. syringae* pv. *tomato* DC3000, *P. savastanoi* pv. *savastanoi* as a focus protein, and additional proteins from the *Pseudomonas* species: MatE efflux family protein from *Pseudomonas syringae* pv. *actinidiae*, MATE efflux protein from *Pseudomonas syringae* pv. *tomato* DC3000, and MATE efflux protein from *Pseudomonas savastanoi* pv. *savastanoi*. The amino acids marked in red show the same parts of the sequences which have been aligned, meaning they are homologous and conserved among the aligned species. Dark (gray) parts are indicators of the existing gaps among the aligned sequences. Sequences listed are by order as aforementioned. The complete alignment is available in the Appendix 6.



**Figure 3.10:** MSA by T-Coffee of *matE* gene-encoded proteins from four species of interest, *P. syringae* pv. *actinidiae*, *P. syringae* pv. *tomato* DC3000, *P. savastanoi* pv. *savastanoi*, and additional proteins from the *Pseudomonas* species. The fourth sequence is due to the additional comparison and presence of the *matE* gene in the sequence. The name of the proteins on the left side of 3.10 is by order as previously mentioned. A pink color indicates good alignments, yellow average, green, and purple bad alignments. Full sequences of T-coffee MSA analysis for the *matE* gene encoded proteins are available in the Appendix 4.

MSA results we got on the T-coffee server are consistent with the results we got by ClustalW. We have a high level of the average homology value and conservation of the sequence, thus giving us a reason to continue our research. Average sequence homology indicates the difference in the sequence structure, so the protein's function might be different as well. Protein's folding might be influenced, so several polymorphisms might play a role in a protein and bacterial toxicity influence on the plant, which was our aim to investigate.

With the sequence homology and conservation, we also have the percent identity matrix to show us the reliable numeral data of the identity of sequences we analyze, so we can build our next steps in the analysis.

### **3.3.3. Identity Matrix**

Percent identity matrix for *matE* gene-encoded proteins MatE are listed in 3.9.

The percent identity matrix shows the similarity between the proteins of interest from the MatE proteins from *Pseudomonas syringae* pv. *actinidiae* as a focus of our research and additional protein sequences we analyzed.

**Table 3.9:** Identity Matrix table for proteins encoded by *matE* genes by ClustalW. Proteins compared and listed are by order of *Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* DC3000, and *Pseudomonas savastanoi* pv. *savastanoi* for *iaaL* gene. For *matE* gene: *Glycine max* (Soybean), *Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* DC3000, and *Pseudomonas savastanoi* pv. *savastanoi*.

|                                                                                                  | matE protein                                                                    |                                                                                                  |                                                                                |                                                                                    |
|--------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
|                                                                                                  | detoxification<br>from <i>Glycine</i><br><i>max</i> (Soybean)<br>NP_001238722.1 | MatE efflux<br>family protein ( <i>P.</i><br><i>s.</i> pv. <i>actinidiae</i> )<br>WP_003380687.1 | MATE efflux<br>protein ( <i>P. s.</i><br>pv. <i>tomato</i> )<br>WP_005782319.1 | MATE efflux<br>protein ( <i>P. s.</i><br>pv. <i>savastanoi</i> )<br>WP_019742037.1 |
| matE protein<br>detoxification<br>from <i>Glycine</i><br><i>max</i> (Soybean)<br>NP_001238722.1  | 100.00                                                                          | 23.61                                                                                            | 22.83                                                                          | 22.88                                                                              |
| MatE efflux<br>family protein ( <i>P.</i><br><i>s.</i> pv. <i>actinidiae</i> )<br>WP_003380687.1 | 23.61                                                                           | 100.00                                                                                           | 88.68                                                                          | 87.37                                                                              |
| MATE efflux<br>protein ( <i>P. s.</i> pv.<br><i>tomato</i> )<br>WP_005782319.1                   | 22.83                                                                           | 88.68                                                                                            | 100.00                                                                         | 90.33                                                                              |
| MATE efflux<br>protein ( <i>P. s.</i> pv.<br><i>savastanoi</i> )<br>WP_019742037.1               | 22.88                                                                           | 87.37                                                                                            | 90.33                                                                          | 100.00                                                                             |

**Table 3.10:** Extended Identity Matrix table for MatE proteins by ClustalW. An outgroup species of *Glycine max* is included in the matrix. The accession numbers of the organisms used are available in 3.11. The results show different percentage identities between the organisms, with only three proteins having 100% identity amongst them. Differences in the identity matrix also suggest the mutations in the sequence, which might influence the 3D protein folding pattern. With the results we got, we did a phylogeny analysis to see the evolutionary distance between our proteins.

|                                 |       |       |       |       |       |       |       |       |       |       |       |       |       |
|---------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Glycine_<br>max                 | 100   | 24.93 | 24.93 | 24.93 | 24.87 | 24.66 | 24.39 | 24.66 | 24.66 | 24.66 | 24.39 | 24.39 | 24.39 |
| P.savastanoi<br>_MP3            | 24.93 | 100   | 99.56 | 99.78 | 99.78 | 87.55 | 87.55 | 87.77 | 87.77 | 87.77 | 88.59 | 88.37 | 88.14 |
| P.syringae<br>_group_MT4        | 24.93 | 99.56 | 100   | 99.78 | 99.78 | 87.99 | 87.99 | 88.21 | 88.21 | 88.21 | 88.81 | 88.59 | 88.37 |
| P.syringae<br>_group_MT2        | 24.93 | 99.78 | 99.78 | 100   | 100   | 87.77 | 87.77 | 87.99 | 87.99 | 87.99 | 88.59 | 88.37 | 88.14 |
| P.savastanoi<br>_MatEefflux     | 24.87 | 99.78 | 99.78 | 100   | 100   | 87.77 | 87.77 | 87.99 | 87.99 | 87.99 | 88.59 | 88.37 | 88.14 |
| P.syringae_gr<br>_genomosp3.MT5 | 24.66 | 87.55 | 87.99 | 87.77 | 87.77 | 100   | 99.56 | 99.78 | 99.78 | 99.78 | 90.38 | 90.16 | 90.38 |
| P.amygdali<br>_MT               | 24.39 | 87.55 | 87.99 | 87.77 | 87.77 | 99.56 | 100   | 99.78 | 99.78 | 99.78 | 90.38 | 90.16 | 90.38 |
| P.syringae<br>_group_MT3        | 24.66 | 87.77 | 88.21 | 87.99 | 87.99 | 99.78 | 99.78 | 100   | 100   | 100   | 90.6  | 90.38 | 90.6  |
| P.tomato<br>DC3000_MatEefflux   | 24.66 | 87.77 | 88.21 | 87.99 | 87.99 | 99.78 | 99.78 | 100   | 100   | 100   | 90.6  | 90.38 | 90.6  |
| P.tomatoT1<br>_MatEefflux       | 24.66 | 87.77 | 88.21 | 87.99 | 87.99 | 99.78 | 99.78 | 100   | 100   | 100   | 90.6  | 90.38 | 90.6  |
| P.syringae<br>_MT1              | 24.39 | 88.59 | 88.81 | 88.59 | 88.59 | 90.38 | 90.38 | 90.6  | 90.6  | 90.6  | 100   | 99.55 | 99.11 |
| P.syringae<br>_MP2              | 24.39 | 88.37 | 88.59 | 88.37 | 88.37 | 90.16 | 90.16 | 90.38 | 90.38 | 90.38 | 99.55 | 100   | 99.11 |
| P.syringae<br>_MP1              | 24.39 | 88.14 | 88.37 | 88.14 | 88.14 | 90.38 | 90.38 | 90.6  | 90.6  | 90.6  | 99.11 | 99.11 | 100   |

From the percent identity matrix we can see that our proteins of interest have different levels of identity. Thus, looking for indications of their possible differences in structure and folding, we made a phylogenetic tree to check how distant are the protein sequences we analyzed, in terms of mutations, insertions, and deletions.

### 3.3.4. Phylogenetic Tree

Phylogenetic trees give us insight into the lines of evolutionary descents of proteins from the species of interest from a common ancestor. By aligning sequences of different species containing our proteins of interest in the *Pseudomonas* genus, we got the results presented in the 3.11.

By aligning MatE and protein sequences from the *Pseudomonas* genus, plus the (*Glycine max*) Soybean species, we got the phylogenetic tree with the details of branching and distances.

Soybean is included due to the presence of the *matE* gene and MatE protein, as well as an outgroup control since much data over the function of the MATE proteins, were assigned.

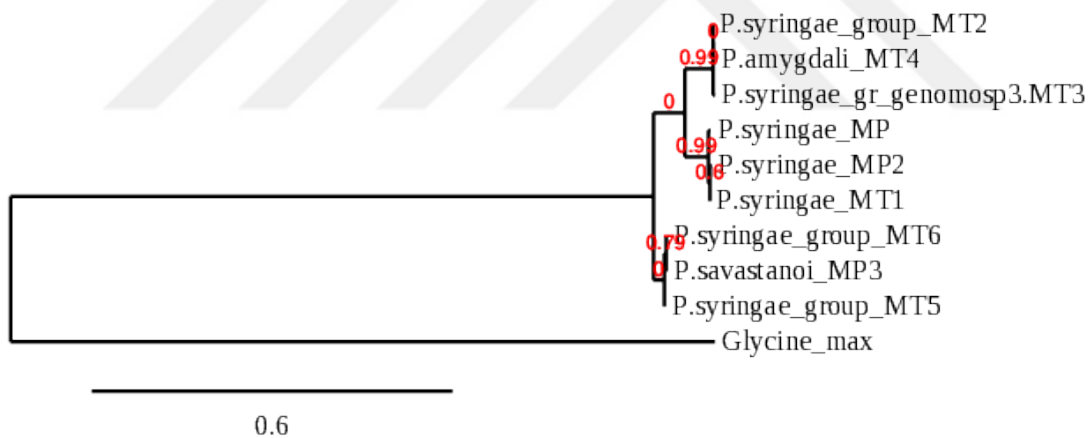


**Figure 3.11:** Phylogenetic tree of MatE proteins from *Pseudomonas* family including *Glycine max* as an outgroup organism. *Glycine max* encodes MatE protein, so we use it as an outgroup organism. The distance between matE proteins encoded by *Glycine max* and *Pseudomonas* family is significant. We chose *Glycine max* as an outgroup organism after we searched the AlphaFold database, and found that *Glycine max* has the highest number of proteins associated with the MatE family protein (9% of 531 143 proteins). Accession numbers and species names are available in the Appendix 13.

The phylogenetic tree of the MatE proteins from different species shows the big distance between the *Glycine max* and *Pseudomonas*. Distances between the organisms from *Pseudomonas* species are more related, so we decided to do the tree only for them. The results are presented in the 3.12.

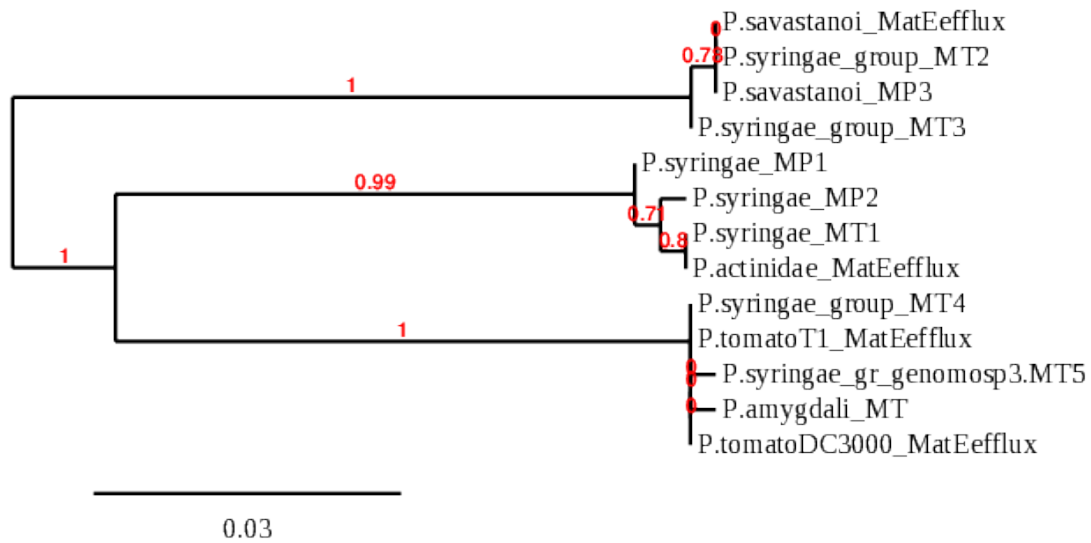


**Figure 3.12:** Phylogenetic tree of MatE proteins in *Pseudomonas* family. The aim of this tree is to see the evolutionary distance between the proteins found in all the related *Pseudomonas* species. Most of the proteins are relatively close, which was expected knowing that they are encoded by organisms from the same genus. Accession numbers for each entry in the phylogenetic tree are listed in the Appendix 14.

The phylogenetic tree for the MatE proteins from *Pseudomonas* species shows us the “grouping” of the more closely related organisms. We can see three main groups of the organisms containing the MatE protein, all of them being closely related but still with significant differences in the evolutionary distance analysis. The focus of the next analysis is to make a 3D model of one protein from each of the three groups, so we can have a more detailed insight into their sequence structure. In the table 3.11 are all of the accession numbers and name tags of the proteins used in the analysis.

**Table 3.11:** Code names of the MatE proteins for the T-coffee and ClustalW analysis.

| Accession      | Label in tree               | Scientific name                                              |
|----------------|-----------------------------|--------------------------------------------------------------|
| WP_003380687.1 | P.syringae_MT1              | <i>Pseudomonas syringae</i> pv.<br><i>actinidiae</i>         |
| WP_005763530.1 | P.amygdali_MT               | <i>Pseudomonas amygdali</i>                                  |
| WP_005782319.1 | P.syringae_group_MT2        | <i>Pseudomonas savastanoi</i> pv.<br><i>savastanoi</i>       |
| WP_007247551.1 | P.syringae_group:MT3        | <i>Pseudomonas syringae</i> group                            |
| WP_017700567.1 | P.syringae_MP1              | <i>Pseudomonas syringae</i>                                  |
| WP_017710541.1 | P.syringae_MP2              | <i>Pseudomonas syringae</i> pv.<br><i>actinidiae</i>         |
| WP_019742037.1 | P.savastanoi_MP3            | <i>Pseudomonas savastanoi</i>                                |
| WP_054995848.1 | P.syringae_group:MT4        | <i>Pseudomonas syringae</i> group                            |
| WP_230692474.1 | P.syringae_gr:genomosp3.MT5 | <i>Pseudomonas syringae</i> group<br>genomosp.3              |
| AXK59714.1     | P.actinidae:MatEefflux      | <i>Pseudomonas syringae</i> pv.<br><i>actinidiae</i>         |
| AAO53914.1     | P.tomatoDC3000:MatEefflux   | <i>Pseudomonas syringae</i> pv. <i>tomato</i><br>str. DC3000 |
| EEB56701.1     | P.tomatoT1:MatEefflux       | <i>Pseudomonas syringae</i> pv. <i>tomato</i><br>T1          |
| RMS74471.1     | P.savastanoi:MatEefflux     | <i>Pseudomonas savastanoi</i>                                |
| NP_001238722.1 | Glycine_max                 | <i>Glycine max</i>                                           |

### 3.3.5. 3D Structure Protein Prediction and Validation, SWISS-MODEL and Phyre2

Three-dimensional protein structure prediction require understanding the interaction between amino acid bonds, protein folding techniques, and different aspects of the functional properties of the proteins. The function of the protein is directly dependent on its three-dimensional form since the 3D shape represents protein structure formed by many different bonds (interactions) between R groups of amino acids which make the upper side chains of the protein. Sidechains can interact with other molecules, forming different compounds with various functions. Also, even small polymorphisms in the sequence structure can influence the way of protein folding,

altering its function. For us, it was important to analyze these changes, since they might help us understand the way MatE protein influences the toxicity in *actinidia*.

The first MatE model predicted was made by the Swiss-model program, and the results obtained are in 3.13. Models are made with the sequences of the MatE proteins of *Pseudomonas* species. The template model according to which the analysis is performed is *Vibrio cholerae*, assigned automatically by the server as the best available model. Unfortunately, template sequence identity and our protein sequence of interest are only 21% identical, but it was the template with the highest matching score available in the database. Proteins used for making the 3D models are: WP\_005782319.1, WP\_017710541.1, and WP\_003380687.1. The first model is made by the SWISS-MODEL prediction tool, using the WP\_003380687.1 MatE protein.

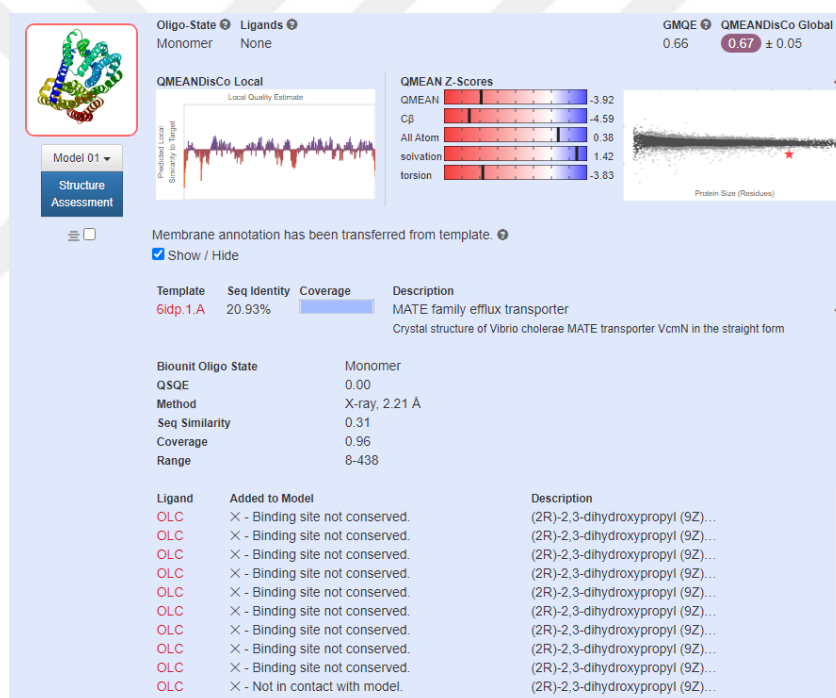


**Figure 3.13:** SWISS-MODEL results for the MatE protein (WP\_003380687.1) of *Pseudomonas syringae* pv. *actinidia* including query sequence identity with the template sequence and coverage of the alignment. The template used for the analysis is from the *Vibrio cholerae*, and was the best available matching template the server could find. The coverage and the method by which the analysis is performed are good, as well as the QMEAN value. The model which we got is a monomer with no ligands presented.

The Swiss model of MatE protein gave us a 3D approximate structure, but the template used had only 21% of sequence identity with our protein of interest. The tube-like structure of the protein is the first insight into the protein's overall shape. According to the model we built,



**Figure 3.14:** SWISS-MODEL of the MatE protein model-template model of *Pseudomonas syringae* (WP\_003380687.1)\*. Model-template alignment represents the query and template alignment with the assigned encircled parts 3.14 which depict alpha-helices with the amino acids of which they consist. There is a noticeable low level of sequence identity between the query sequence and a template.

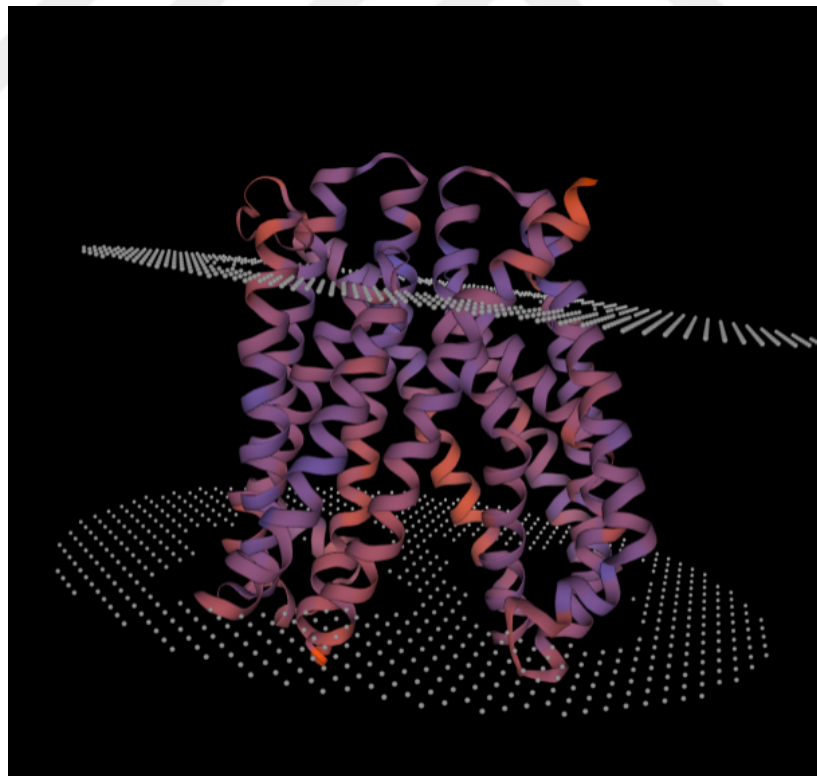


**Figure 3.15:** MatE (WP\_003380687.1) 3D protein shape and its alpha-helical structure of *Pseudomonas syringae* pv. *actinidiae*. The model consists of 10 helices orbiting and interconnected by loops. Helices are organized in a tube-like formation and are of closely uniform length. The model is made by the Swiss-model program.

we have a basis for the further analyzes we can perform.

The second program used for the 3D protein prediction was Phyre2. The reason why two programs were used is that they offer different additional tools and analyzes, as well as for the comparison of the programs and results. Also, we used them as control tools for accuracy and different server abilities to perform. Phyre2 models of the proteins are presented in 3.16.

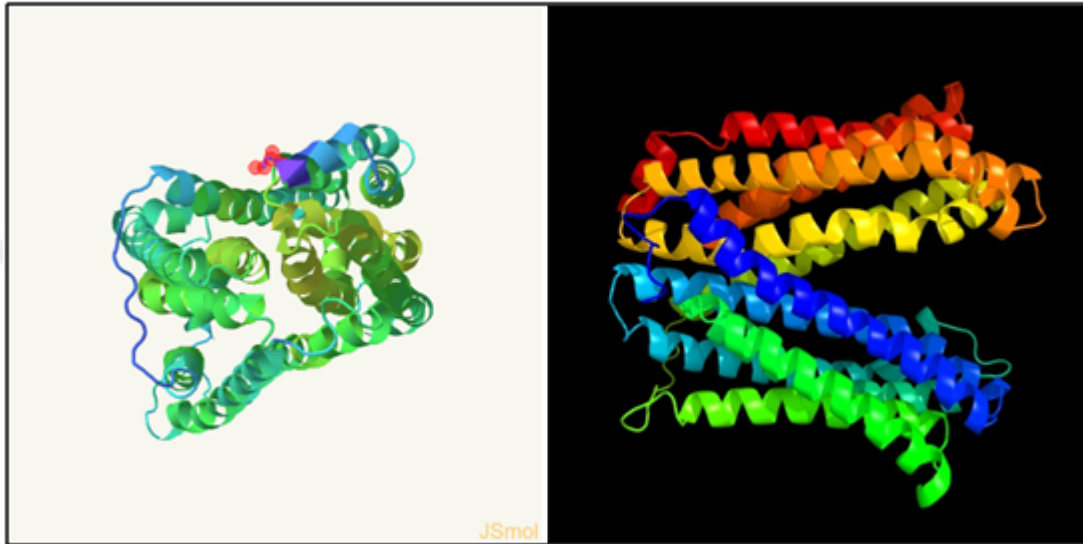


**Figure 3.16:** PHYRE2 3D model of the MatE protein sequence WP\_003380687.1 of *Pseudomonas syringae* pv. *actinidiae*. The identity of the query with the template is 21%. It is low due to the lack of a good template. The number of residues aligned is 426, which shows that the complete sequence has been aligned. The red color in the first 3D model marks the active site of the protein.

After we found out the residue differences between our MatE protein (WP\_003380687.1) of *Pseudomonas syringae* pv. *actinidiae* and other proteins, we wanted to check its 3D structure, so we can visualize the protein.

We took three proteins from the *Pseudomonas* group to see their 3D structure and analyze the structural differences and possible folding changes. Differences in the folding pattern might influence the protein function, which makes our analysis significant for the aim of this research.

Due to the range of this research and variations of MatE proteins in structure, the influence on the toxicity of *Pseudomonas* species can be significant. To compare it, we did the SWISS-MODEL analysis for the MatE protein of three proteins of *Pseudomonas* species including our MatE protein of interest from *Pseudomonas syringae* pv. *actinidiae*. The results we got are presented with the 3D models for the comparison in 3.23, 3.24 and 3.25.



**Figure 3.17:** Aligned secondary structure of MatE protein sequence WP\_003380687.1 of *Pseudomonas syringae* pv. *actinidiae*, consisting of the alpha helices, and the primary protein structure. We got the results from the Phyre2 server. The five lines in each block represent the predicted secondary structure, the query sequence, the template sequence, the template known secondary structure, and the template predicted secondary structure. Grey area marks conserved matching residues. Gaps and mismatch polymorphisms are present, due to the low sequence identity of the protein and template. The predicted structure of the protein consists of the helices, even though further analyzes are needed for an accurate protein model. For the analysis is used the Phyre2 server.

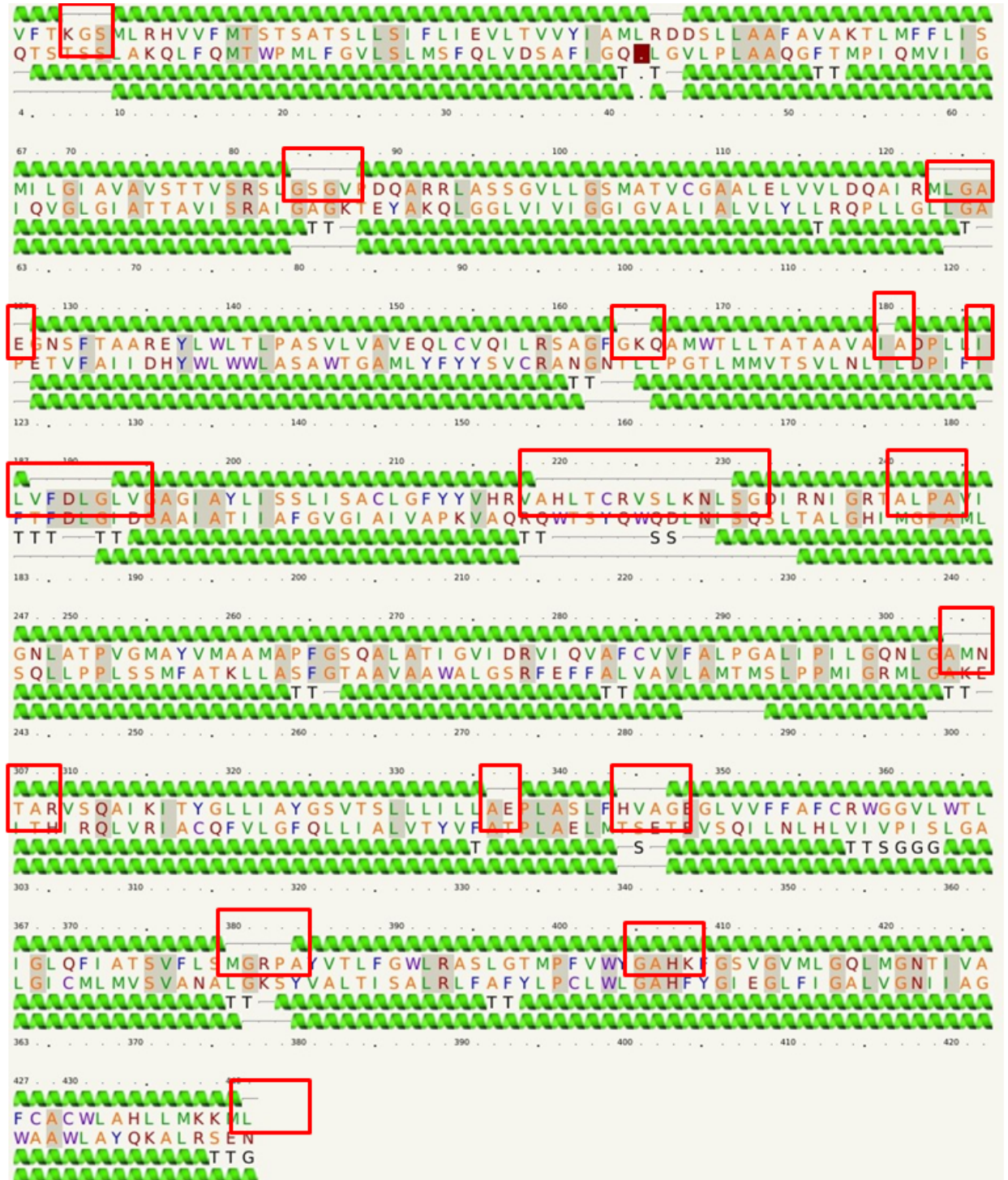


**Figure 3.18:** Aligned secondary structure of the MatE protein WP\_005782319.1 of *Pseudomonas savastanoi* pv. *savastanoi* with red encircled residues different from two other proteins aligned from the same species. Encircled parts are compared with two other proteins, and all of the differences are related to the loops. The position in the loop of some residue mutations or polymorphisms in one protein is different from the loop residue position in the second and third protein. Since residue differences are in the loop, they might influence the changes in 3D protein folding and function.

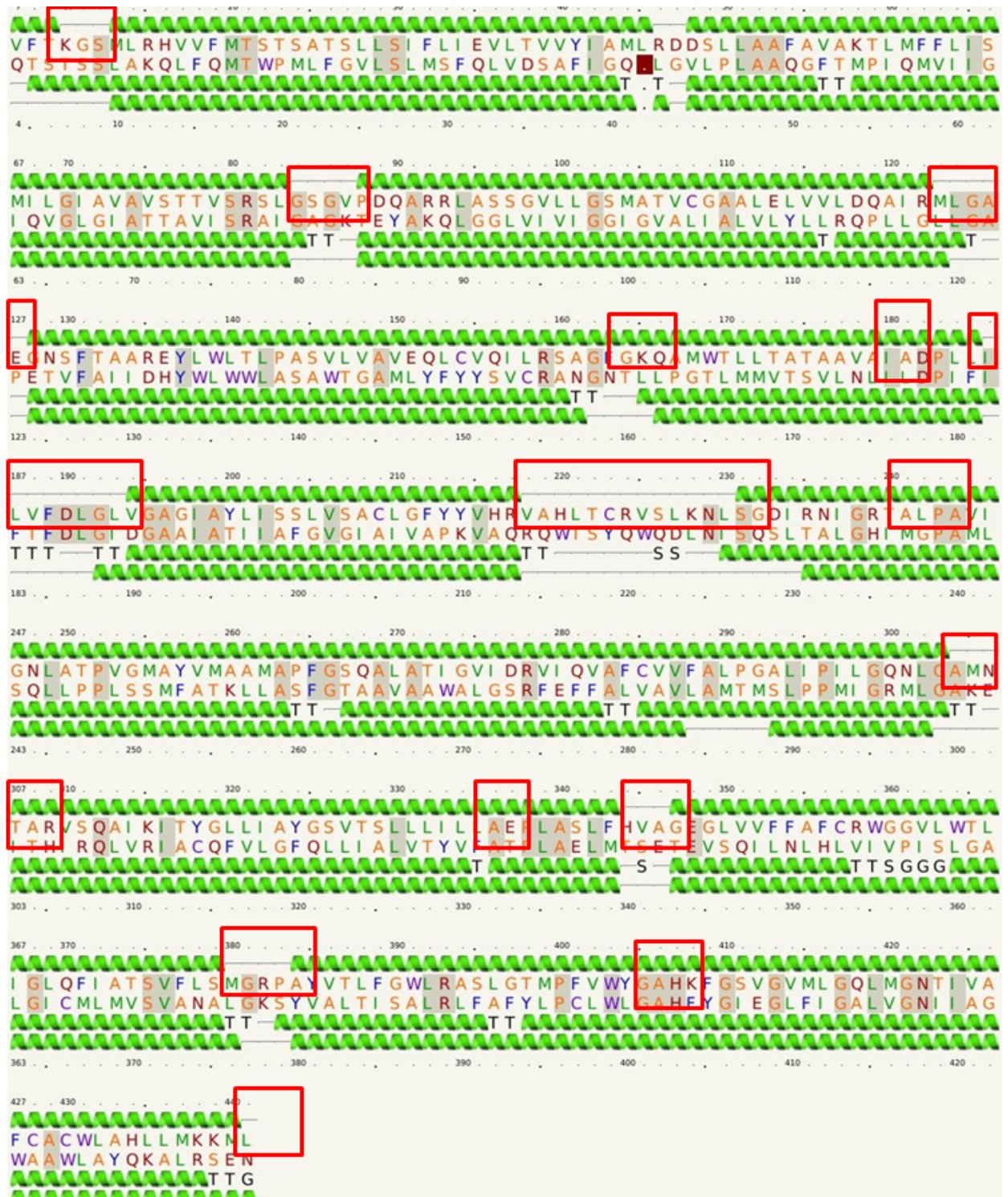


**Figure 3.19:** Aligned secondary structure of MatE protein WP\_054995848.1 of *Pseudomonas syringae* pv. *tomato* DC3000 with red encircled residues different from two other proteins aligned from the same species. Some differences in residues of WP\_054995848.1 and WP\_005782319.1 protein are common against the WP\_003380687.1 protein. Despite common differences in residues, they have differences between them.

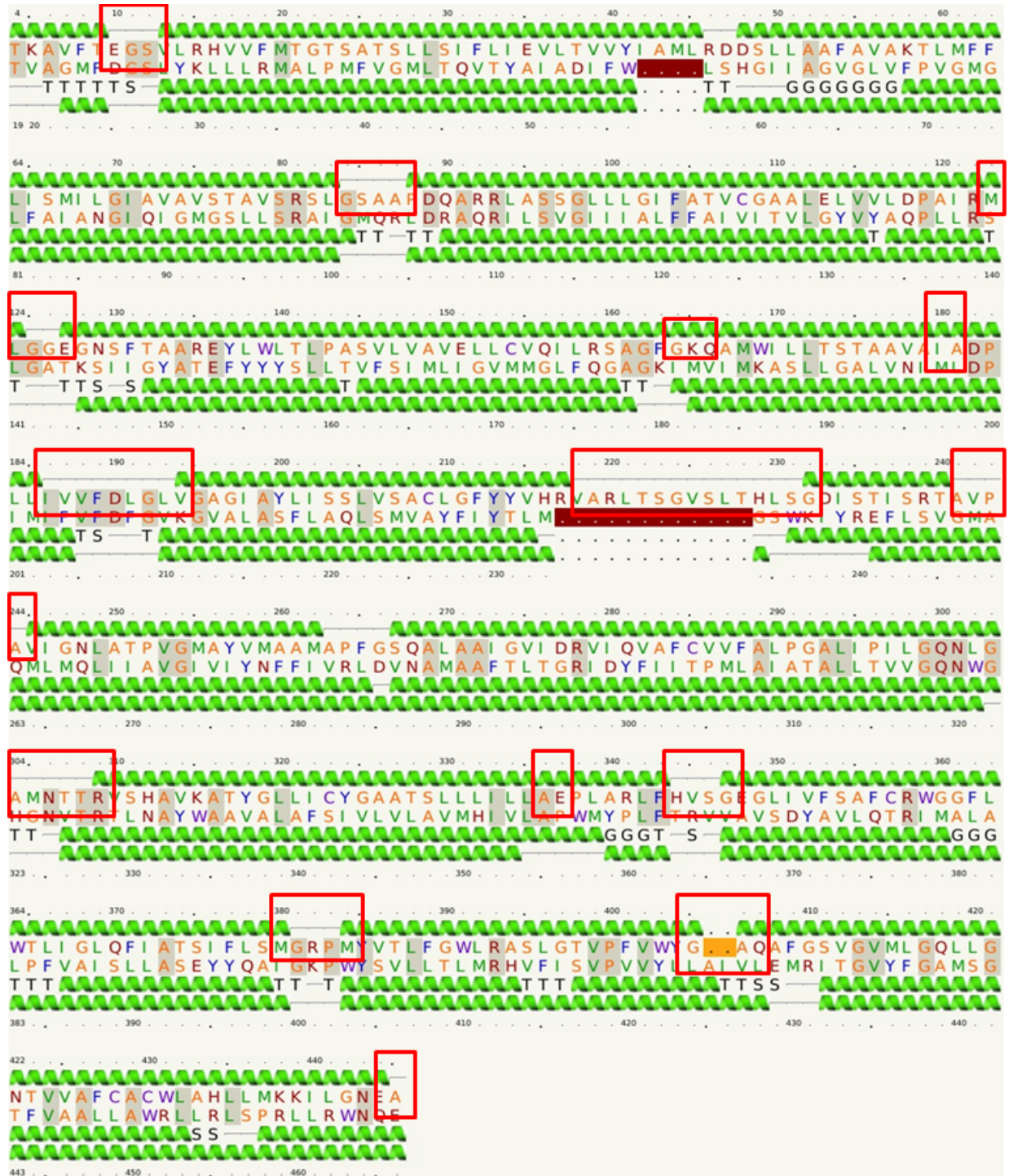


**Figure 3.20:** Aligned secondary structure of MatE protein WP\_003380687.1 of *Pseudomonas syringae* pv. *actinidiae* with red encircled residues different from two other proteins aligned from the same species. Different residues in the structure of this MatE protein might be the cause of specific changes in protein 3D folding formation, and bacterial toxicity on the plant after the infection occurs.

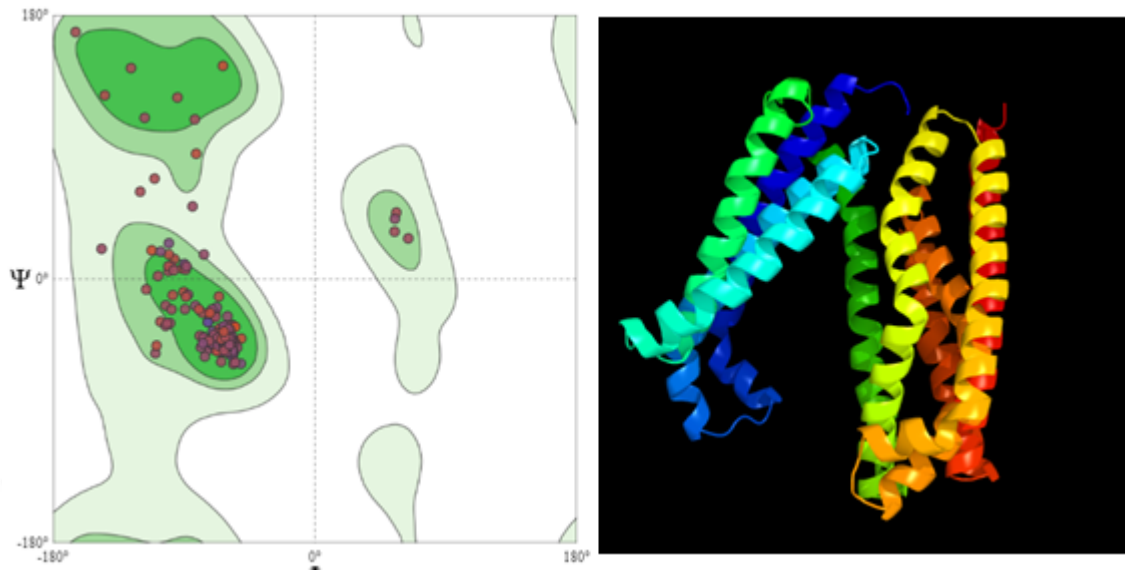


**Figure 3.21:** PHYRE2 3D model and Ramachandran plot of the MatE protein (WP\_003380687.1) of *Pseudomonas syringae* pv. *actinidiae* based on the template 6idp.1.A. 362 residues (99% of the sequence) are modeled with 100.0% confidence using the highest scoring template, the crystal structure of *Vibrio cholerae* mate transporter vcmn in the bent2 form, which has 20.93% sequence identity respect to MatE in *Pseudomonas syringae* pv. *actinidiae*. The predicted structure shows a tube-like structure of the protein consisting of alpha helices connected by loops.

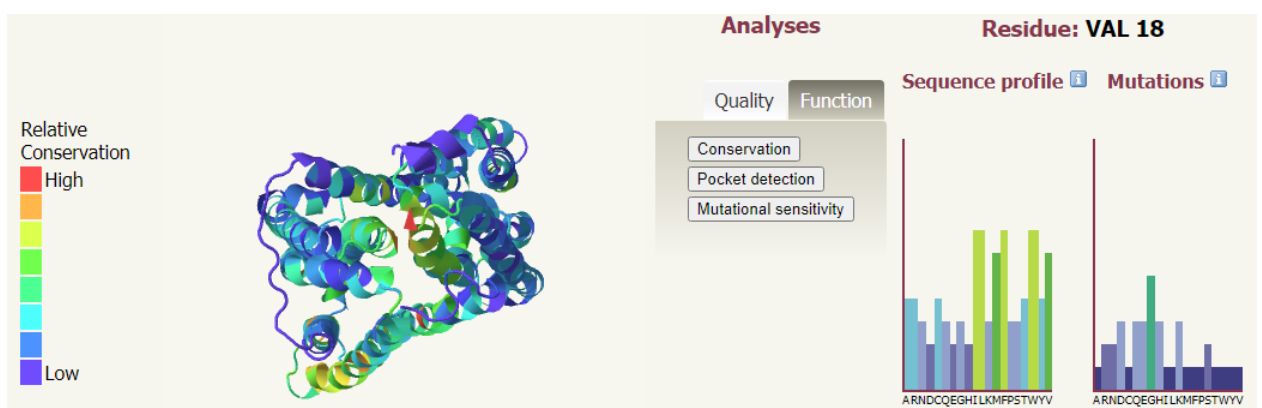


**Figure 3.22:** PHYRE2 3D model of the MatE protein of *Pseudomonas syringae* pv. *actinidiae* (WP\_003380687.1) with emphasis on the conservation level of the sequence. The purple/blue color indicates low conservation. This result is coherent with what we found using SWISS-MODEL protein prediction analysis in terms of sequence identity and 3D model prediction, as we can see on the 3.15 .

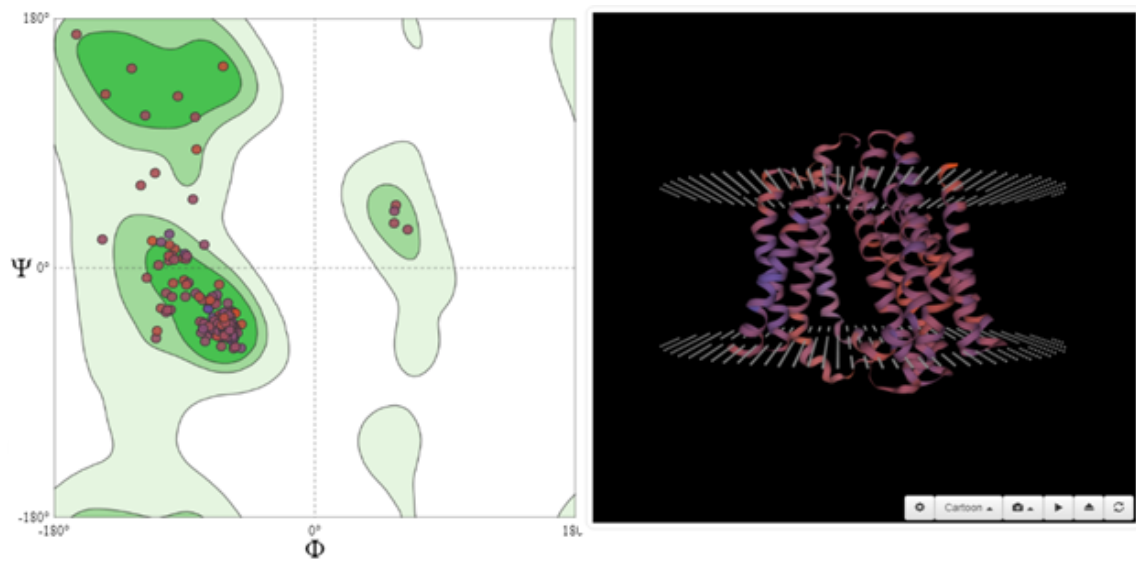


**Figure 3.23:** Ramachandran plot and 3D protein predictions of MatE protein of *Pseudomonas syringae* pv. *actinidiae* (WP\_003380687.1).

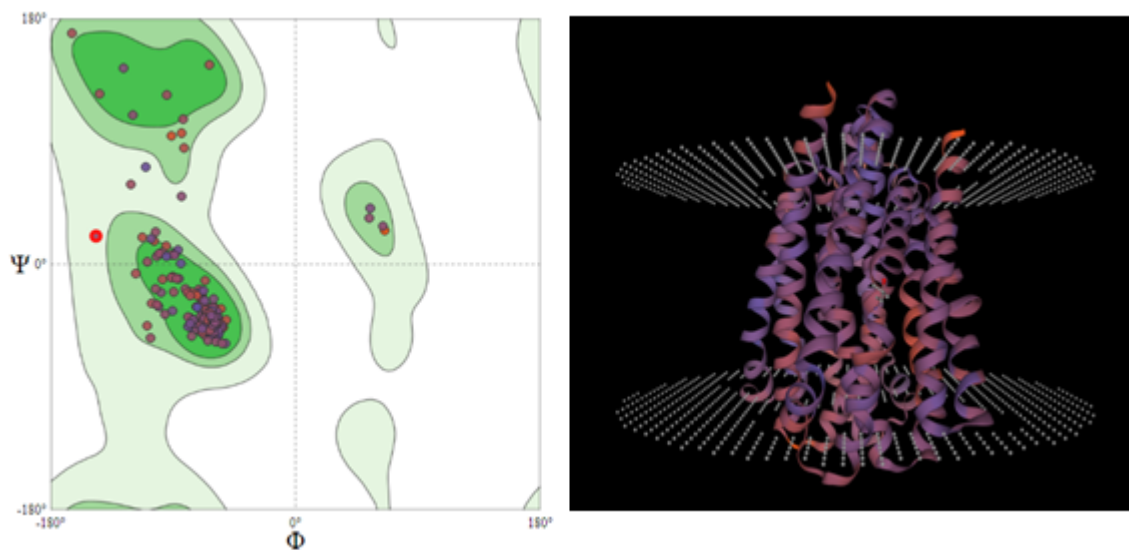


**Figure 3.24:** Ramachandran plot and 3D protein predictions of MatE protein of *Pseudomonas syringae* pv. *tomato* DC3000 (WP\_005782319.1).

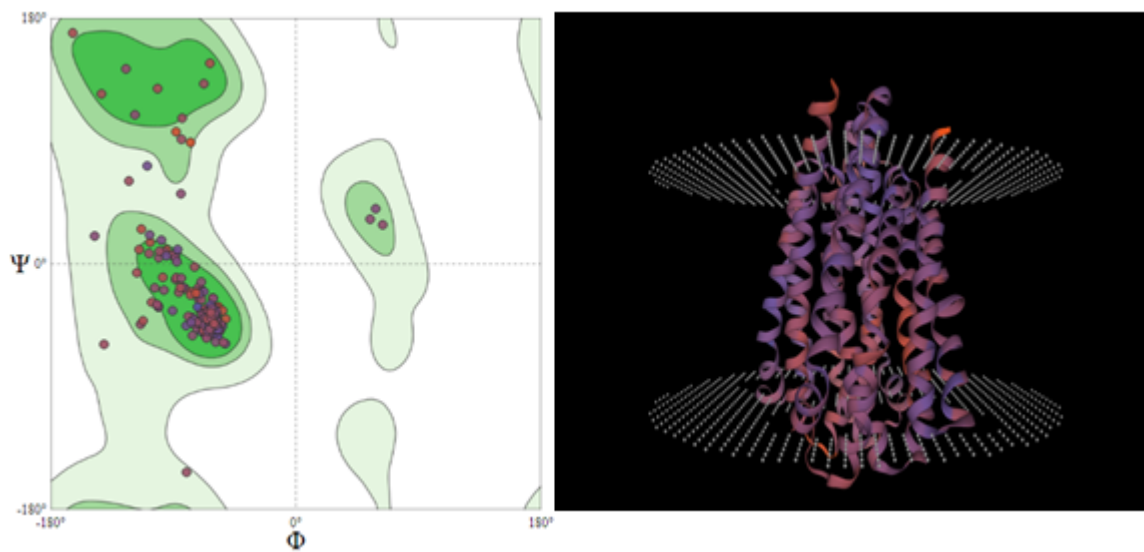


**Figure 3.25:** Ramachandran plots and 3D protein predictions of MatE protein of *Pseudomonas savastanoi* pv. *savastanoi* (WP\_019742037.1). A comparison of three proteins from different organisms from the same species shows a similar pattern of the protein folding and Ramachandran plot distribution. Minor differences are not visible in these figures due to the specific structure of the proteins. Linear analysis shows the sequence polymorphisms and mutations in different organisms which might have influenced the folding and function of the proteins.

The three-dimensional structure of the proteins helped us understand the changes in the sequence structures and more characteristics of each specific structure. For the additional analysis, we did protein domain analysis, since domains can function separately from the protein and alter its function.

### 3.3.6. Predicted Protein Domains

Protein domains are often described as structural domains which are self-stabilizing and fold independently of the rest of the protein chain. The importance of protein domains, especially if they are conserved, is in their functionality. If a conserved protein domain with a known function is present in the sequence, then the protein function may be determined. We can also determine the protein classification, its biological function, follow the evolutionary mechanisms and design a protein more accurately.

The majority of domains annotated in UniProtKB/Swiss-Prot are defined by the InterPro resources PROSITE, Pfam, and SMART.

The MatE protein sequence of *Pseudomonas syringae* pv. *actinidiae* was not found in the PROSITE database.

Due to the importance of the protein domains and determination of the function of the MatE protein of our interest, we performed additional analysis using the InterPro server. MatE protein of interest has been found and the results we got are similar to the results of the

SMART tool, with the addition of several more details and domains mentioned. Important notice is that our protein of interest is recognized as a transmembrane protein with few protein domains. Its localization is predicted to be a membrane, biological functions to be antiporter activity, and in biological processes to be part of transmembrane transport, as we will see in the discussion part.

For further analysis, SMART analysis was performed on all the species of interest with the corresponding MatE proteins. The results we got are as follows in 3.29.

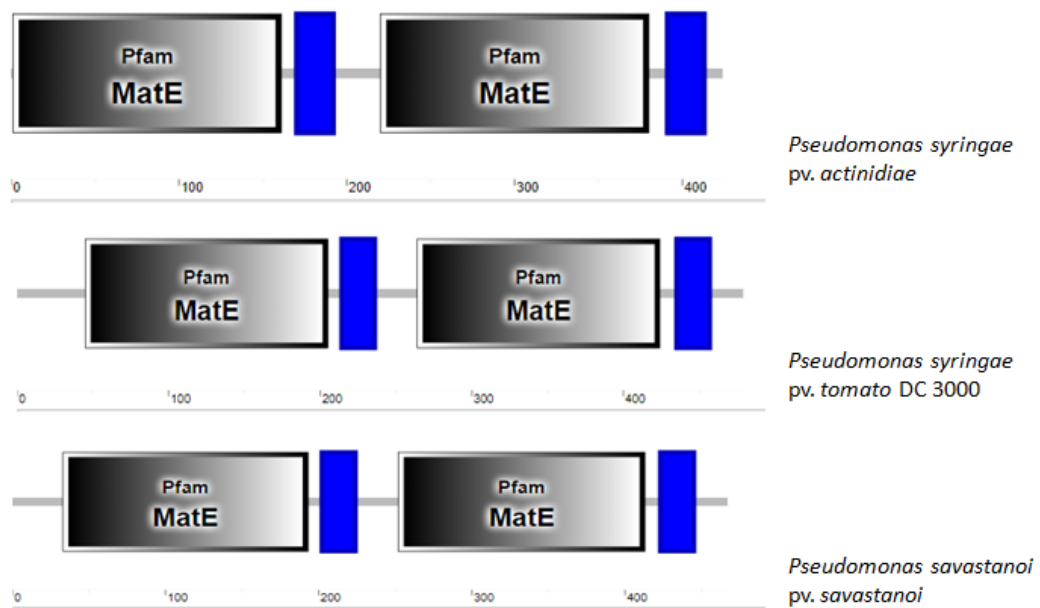


**Figure 3.29:** SMART analysis of protein domains of 3 MatE query proteins of interest with a graphical representation of different domains and regions they are found at. All three of the MatE proteins are found at a similar region of the sequence having a similar length and with two additional transmembrane domains next to them. *Pseudomonas syringae* pv. *actinidiae* (WP\_003380687.1); *Pseudomonas syringae* pv. *tomato* DC3000 (WP\_005782319.1); *Pseudomonas savastanoi* pv. *savastanoi* (WP\_019742037.1).

SMART server table representation of the protein domains of three species of interest with the name of domains and regions they are found at, as well as the E-value for that domain, is available in the Appendix 7. The focus is on the MatE domains. As it can be seen on the 3.29, the total length of the domains is the same, even though the start and end are at different positions.

Our next step was to check the docking sites, so we can have an additional look into the protein interaction.

### 3.3.7. Prediction of Docking Sites

By predicting the docking sites it is possible to see the kinds and differences between small substrate molecules interacting with our protein of interest. Ligands or substrate molecules that interact with the protein can increase the protein efficiency, affect the interaction of the protein with other molecules or influence it in other ways. One specific ligand was found to be interacting with our protein of interest. The RCSB PDB has provided us with a detailed 3D representation of the ligand and protein interaction.

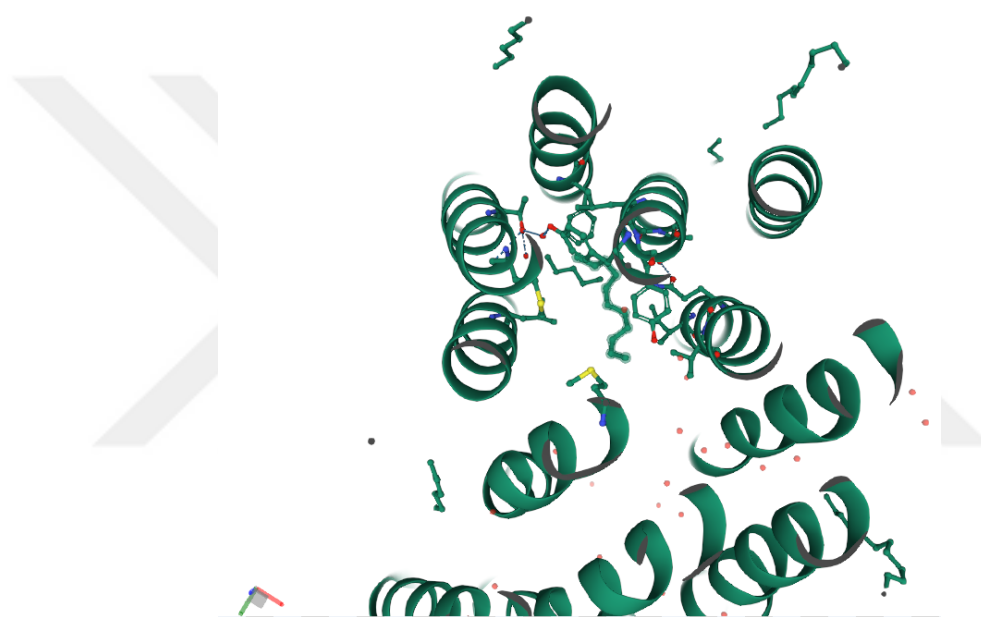


**Figure 3.30:** Interaction of the ligand “OLC” and the MatE protein of *P. syringae* pv. *actinidiae* (WP\_003380687.1). Ligand is in the inner part of the protein tube-like formation, a small molecule with no folds, as it can be seen in the 3.31.

## 3.4. PREDICTION OF THE INTERACTOME

Interactome prediction of the novel proteins is not always easy. For the MatE protein, several different servers and databases were used. The best results, shown in 3.33, were obtained using the STRING database.

The interactome analysis of *Pseudomonas syringae* pv. *tomato* and *Pseudomonas savastanoi* pv. *savastanoi* was also performed, but since the only available template in the STRING database has been a sequence of the *Pseudomonas syringae* pv. *tomato* DC3000, all of the obtained results were the same. Thus, interactions of the query sequence have all matching and highly similar interactome groups according to the available bioinformatics tool.

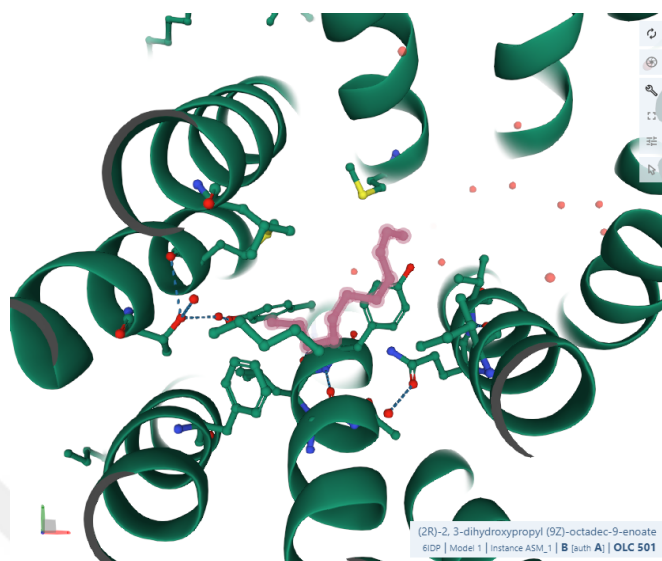


**Figure 3.31:** Closer look at the protein and ligand interaction of the MatE protein of the *Pseudomonas syringae* pv. *actinidiae* with the ligand “OLC” (WP\_003380687.1 protein). Ligand attaches to the inner part of the tube-like formation of the protein. Also, the ligand has several chains but is still simple and only one found to interact with the MatE protein.

| Small Molecules                                                                                                                                                           |                                                                                                                                 |                                                                                                           |            |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|------------|
| Ligands 1 Unique                                                                                                                                                          |                                                                                                                                 |                                                                                                           |            |
| ID                                                                                                                                                                        | Chains                                                                                                                          | Name / Formula / InChI Key                                                                                | 2D Diagram |
| OLC<br><a href="#">Query on OLC</a><br><input type="button" value="Download Ideal Coordinates CCD File"/><br><input type="button" value="Download Instance Coordinates"/> | B [auth A]<br>, C [auth A]<br>, D [auth A]<br>, E [auth A]<br>, F [auth A]<br>, G [auth A]<br>, H [auth A]<br>, I [auth A]<br>+ | <b>(2R)-2,3-dihydroxypropyl (9Z)-octadec-9-enoate</b><br>$C_{21}H_{40}O_4$<br>RZRNAYUHWVFMIP-GDCKJWNLSA-N |            |

**Figure 3.32:** Tabular representation of the ligand details which interact with the MatE protein of the *P. syringae* pv. *actinidiae* (WP\_003380687.1)\*. There is only one ligand found to be interacting with our protein of interest, but it has several chains.

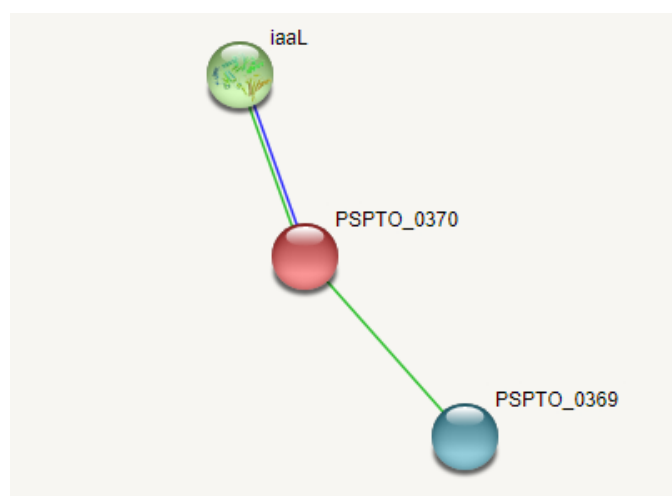


**Figure 3.33:** Interactome analysis of the MatE protein (PSPTO\_0370) of *P. syringae* pv. *actinidiae* (WP\_003380687.1) by STRING. The analysis was done at the confidence level of 0.900. IaaL protein is from the *P. syringae* pv. *actinidiae* as well (WP\_003380685.1). There are only two interacting proteins with the MatE protein: IaaL and outer membrane porin protein from *P. syringae* pv. *tomato*.

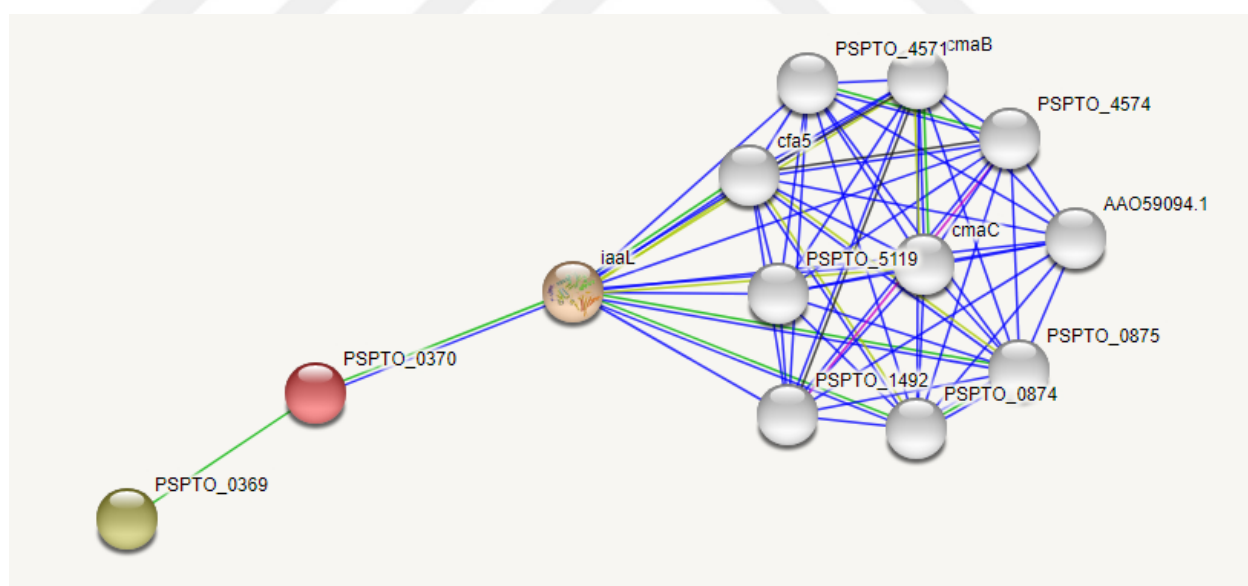


**Figure 3.34:** Wider interactome analysis of the MATE efflux family protein (PSPTO\_0370) of *P. syringae* pv. *actinidiae* (WP\_003380687.1) by STRING. All of the interacting proteins are from the *Pseudomonas syringae* pv. *tomato* DC3000 family. Not all of the interacting proteins are characterized. Out of those which are characterized, they have various functions. It is important to note the difference between the number of MATE and IaaL protein interactions. MATE only interacts with two proteins. IaaL protein is from the *P. syringae* pv. *actinidiae* as well (WP\_003380685.1).

### 3.5. SUBCELLULAR LOCALIZATION OF MATE AND IAA PROTEINS

The localization of the proteins of interest has been done using different servers. The two main servers which gave the best results are “pLoc-mGneg” and “LocTree3”. Obtained results are summarized as follows in 3.12.

**Table 3.12:** Subcellular localization of the proteins of interest based on two-server analysis and available scores and accuracy for better comparison and analysis. The server helped us to determine the localization of the proteins as inner membranous for the MatE, and cytoplasmic for the IaaL proteins. Two different servers gave us the same results.

| Organism                                                           | Prot | “pLoc-mGneg”<br>analysis | “LocTree3”<br>analysis | Score | Accuracy |
|--------------------------------------------------------------------|------|--------------------------|------------------------|-------|----------|
| <i>Pseudomonas syringae</i> pv. <i>actinidiae</i> WP_003380687.1   | MatE | Cell inner<br>membrane   | Inner<br>membrane      | 100   | 0.99     |
| <i>Pseudomonas syringae</i> pv. <i>tomato</i> WP_005782319.1       | MatE | Cell inner<br>membrane   | Inner<br>membrane      | 100   | 0.99     |
| <i>Pseudomonas savastanoi</i> pv. <i>savastanoi</i> WP_019742037.1 | MatE | Cell inner<br>membrane   | Inner<br>membrane      | 100   | 0.99     |
| <i>Glycine max</i> (Soybean) NP_001238722.1                        | MatE | Cell inner<br>membrane   | Inner<br>membrane      | 6     | 0.88     |
| <i>Pseudomonas syringae</i> pv. <i>actinidiae</i> WP_003380685.1   | IaaL | Cytoplasm                | Cytoplasm              | 100   |          |
| <i>Pseudomonas syringae</i> pv. <i>tomato</i> WP_007247550.1       | IaaL | Cytoplasm                | Cytoplasm              | 99    |          |
| <i>Pseudomonas savastanoi</i> pv. <i>savastanoi</i> WP_057456329.1 | IaaL | Cytoplasm                | Cytoplasm              | 83    |          |
| <i>Glycine max</i> (Soybean) N/A                                   | IaaL | N/A                      | N/A                    | N/A   |          |

All of the MatE proteins are located in the cell’s inner membrane, while the IaaL proteins have been located as cytoplasmic proteins, except for the *Glycine max*. LocTree3 analysis provided us with additional details on the score and accuracy of the analysis. In all of the cases obtained results have a high level of the accuracy of the protein location. The results obtained are the same in both cases. Based on the LocTree3 analysis, MatE proteins are inner membranous, while the IaaL proteins are cytoplasmic. A more detailed explanation is in the

discussion part.

### **3.6. MSA AND PHYLOGENETIC TREE COMPARISON BETWEEN MORE DIFFERENT BACTERIAL STRAINS**

We have the MSA for separate IaaL and MatE proteins with outgroup organisms included in 3.1 and 3.8. For the additional analysis of the proteins we wanted to see the level to which MatE and IaaL proteins are different, so we did the multiple sequence alignment for various species having our proteins of interest. Resulting MSA is presented in the 3.35. No significant similarities between the proteins are found. There are no conserved residues between IaaL and MatE proteins, which is expected due to their different functional role, subcellular localization, and protein family.

### **3.7. INTERGENIC AND PROMOTER SEQUENCES CONSERVATION ANALYSIS**

Analysis of the intergenic sequences from multiple organisms containing the *matE* and *iaaL* genes to see the conservation level was performed by the ClustalW program. It was important to check the level to which intergenic sequence is conserved among different organisms. The presence of different nucleotides in an intergenic region could indicate different functions of a *matE* and *iaaL* genes, their regulation, and their interactome with other molecules in a genome. Descriptions of the genomes from which the sequences analyzed were taken are available in the Appendix 8. The length of the sequence analyzed is 119 nucleotides since the end of the *matE* gene and the beginning of the *iaaL* gene are separated by the intergenic sequence of this length. The results we got are presented in the 3.36.

The same analysis is done for the promoter sequence of the *matE* and *iaaL* genes due to the fact they are both transcribed from the same promoter. The length of the sequence analyzed is 100 nucleotides. The results we got are present in the 3.37.

Conservation analysis of the intergenic region of *matE* and *iaaL* genes, and promoter region showed slight differences between the level to which they are conserved. The intergenic region is completely conserved between many different organisms from the *Pseudomonas* species. No gap or mismatch is found, the same sequence is in between the genes, suggesting the same function across the species. The promoter region showed polymorphism in the sequence between different organisms. This difference in the sequence might be due to the



**Figure 3.35:** MSA of the various MatE efflux proteins and Indoleacetic-lysine ligase proteins by ClustalW server. The figure shows only one part of the sequence alignment due to the high matching score of the part of the sequence. Included species are different bacteria containing proteins of interest. The graphical representation of multiple sequence alignment indicates different information for our homologous proteins. Firstly the color of each letter represents the protein's physicochemical properties. The red color is indicating hydrophobic (including aromatic) residues, the blue one is indicating acidic residues, the pink is indicating basic residues, the green one is indicating hydroxyl, sulfhydryl, and amine groups and lastly, the gray one is indicating unusual amino acids. Secondly, each symbol of graphical representation of MSA is telling us how the latter is similar to the query protein. An asterisk (\*) indicates that the residue is the same across all query proteins; a colon (:) indicates that the residues are different but highly similar across query proteins, and a dot (.) indicates that residues are different but similar to some degree, across query proteins.

|                                |                                                             |     |
|--------------------------------|-------------------------------------------------------------|-----|
| NZ_MTHQ01000062.1:27568-27686  | TTAGGGAACCGATCTGTCCATCGGGGTAACATAGCCC6CGTTACCGCCAATCAGCCGC  | 60  |
| NZ_MTYU01000005.1:59768-59886  | TTAGGGAACCGATCTGTCCATCGGGGTAACATAGCCC6CGTTACCGCCAATCAGCCGC  | 60  |
| NZ_MSBX01000060.1:58984-59102  | TTAGGGAACCGATCTGTCCATCGGGGTAACATAGCCC6CGTTACCGCCAATCAGCCGC  | 60  |
| NZ_MOMK01000007.1:c43954-43836 | TTAGGGAACCGATCTGTCCATCGGGGTAACATAGCCC6CGTTACCGCCAATCAGCCGC  | 60  |
| NZ_MTH001000010.1:59560-59678  | TTAGGGAACCGATCTGTCCATCGGGGTAACATAGCCC6CGTTACCGCCAATCAGCCGC  | 60  |
| NZ_MSBW01000009.1:c4482-4364   | TTAGGGAACCGATCTGTCCATCGGGGTAACATAGCCC6CGTTACCGCCAATCAGCCGC  | 60  |
| NZ_MOMN01000008.1:c43133-43015 | TTAGGGAACCGATCTGTCCATCGGGGTAACATAGCCC6CGTTACCGCCAATCAGCCGC  | 60  |
| NZ_MTYK01000006.1:59312-59430  | TTAGGGAACCGATCTGTCCATCGGGGTAACATAGCCC6CGTTACCGCCAATCAGCCGC  | 60  |
| *****                          |                                                             |     |
| NZ_MTHQ01000062.1:27568-27686  | TGCTGATGGGCGGATACGTCATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA | 119 |
| NZ_MTYU01000005.1:59768-59886  | TGCTGATGGGCGGATACGTCATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA | 119 |
| NZ_MSBX01000060.1:58984-59102  | TGCTGATGGGCGGATACGTCATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA | 119 |
| NZ_MOMK01000007.1:c43954-43836 | TGCTGATGGGCGGATACGTCATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA | 119 |
| NZ_MTH001000010.1:59560-59678  | TGCTGATGGGCGGATACGTCATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA | 119 |
| NZ_MSBW01000009.1:c4482-4364   | TGCTGATGGGCGGATACGTCATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA | 119 |
| NZ_MOMN01000008.1:c43133-43015 | TGCTGATGGGCGGATACGTCATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA | 119 |
| NZ_MTYK01000006.1:59312-59430  | TGCTGATGGGCGGATACGTCATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA | 119 |
| *****                          |                                                             |     |

**Figure 3.36:** ClustalW analysis of conservation level of the intergenic sequences between *matE* and *iaaL* genes in different organisms. Descriptions of the genomes used in the analysis and full data are available in the appendix. An asterisk sign (\*) is used to describe the highest conservation level for the nucleotides analyzed. Sequences analyzed showed to be 100% conserved.

|                                   |                                                             |     |
|-----------------------------------|-------------------------------------------------------------|-----|
| NZ_SNQV01000018.1:54003-54103     | TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCATGCCGATGCTCATC | 60  |
| NZ_JABEQW010000007.1:c44523-44423 | TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCATGCCGATGCTCATC | 60  |
| NZ_JAESNH010000014.1:57592-57692  | TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCATGCCGATGCTCATC | 60  |
| NZ_SNQU01000019.1:54003-54103     | TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCATGCCGATGCTCATC | 60  |
| NZ_JAESNF010000017.1:53993-54093  | TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCATGCCGATGCTCATC | 60  |
| NZ_SNQS01000018.1:54003-54103     | TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCATGCCGATGCTCATC | 60  |
| NZ_RBSG01000878.1:c4985-4885      | TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCATGCCGATGCTCATC | 60  |
| NZ_JAESNG010000017.1:54003-54103  | TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCATGCCGATGCTCATC | 60  |
| *****                             |                                                             |     |
| NZ_SNQV01000018.1:54003-54103     | TGAAAATCGGAGATGGCGTTCAACGGGAAGAGAAGGGTCAA                   | 101 |
| NZ_JABEQW010000007.1:c44523-44423 | TGAAAATCGGAGATGGCGTTCAACGGGAAGAGAAGGGTCAA                   | 101 |
| NZ_JAESNH010000014.1:57592-57692  | TGAAAATCGGAGATGGCGTTCAACGGGAAGAGAAGGGTCAA                   | 101 |
| NZ_SNQU01000019.1:54003-54103     | TGAAAATCGGAGATGGCGTTCAACGGGAAGAGAAGGGTCAA                   | 101 |
| NZ_JAESNF010000017.1:53993-54093  | TGAAAATCGGAGATGGCGTTCAACGGGAAGAGAAGGGTCAA                   | 101 |
| NZ_SNQS01000018.1:54003-54103     | TGAAAATCGGAGATGGCGTTCAACGGGAAGAGAAGGGTCAA                   | 101 |
| NZ_RBSG01000878.1:c4985-4885      | TGAAAATCGGAGATGGCGTTCAACGGGAAGAGAAGGGTCAA                   | 101 |
| NZ_JAESNG010000017.1:54003-54103  | TGAAAATCGGAGATGGCGTTCAACGGGAAGAGAAGGGTCAA                   | 101 |
| *****, *****                      |                                                             |     |

**Figure 3.37:** ClustalW conservation analysis of the promoter sequence of *matE* and *iaaL* genes in different organisms. Full descriptions of the genomes analyzed are available in the appendix. The length of the promoter sequence analyzed is 100 nucleotides. The conservation level of the promoter region is not 100%. Noticeable is the polymorphism in the *Pseudomonas syringae* pv. *actinidiae* (NZ\_RBSG01000878.1) noted by the dot in the alignment.

broad strains of proteins from organisms included, even though they belong to the same species. Also, it might be the reason for different protein structures when folded in the tertiary structure. Polymorphisms found to belong to the *Pseudomonas syringae* group, and different strains of *Pseudomonas syringae* pv. *actinidiae* (strain 19071, K6 BV351\_148, 1404, K4 BV349\_147, and others).

Further analyzes of these results are explained in detail in the discussion part. Full details of the analyzes are available in the appendix 9 since we cannot present here the complete sequences used, only their representative parts.

After a detailed analysis of the various sequences of MatE and IaaL proteins from *Pseudomonas* species, we got many results showing us the importance and novel research areas in the field of protein influence on the toxicity or disease occurrence in the host organism. Minor changes in the sequence such as several residues in the structure of MatE protein of *Pseudomonas syringae* pv. *actinidiae* gives us a way to analyze the protein interactions, the significance of the loop region residues, and overall protein folding machinery. Discussions and conclusions are in the following chapters.



## 4. DISCUSSION

Our thesis's main question was deciphering the role of a MatE protein of *Pseudomonas syringae* pv. *actinidiae* on the bacterial toxicity and analysis of protein structural changes which might influence its overall function. Analysis of the MatE protein from different organisms from the *Pseudomonas* species helped us understand the protein interactions, structure, and possible reasons for its virulence effects in tomatoes and kiwi.

Analyzing all of the data and results, we got an insight into different perspectives and details over the bacterial species of *Pseudomonas syringae* pv. *actinidiae* and two closely related species, *Pseudomonas syringae* pv. *tomato* DC3000, and *Pseudomonas savastanoi* pv. *savastanoi*. Even though the focus was on these three species, various other species from the *Pseudomonas* family that have a MatE protein as part of their sequence are analyzed. Correlations and dependence on various factors contributing to virulence and disease development were questioned and analyzed following the available software. The main focus of the study was to obtain results that would show the protein-protein and gene-gene interaction during the infection process. Along with this, it was crucial to investigate virulence factors correlation and cell systems included in the disease development process, as well as 3D protein model creation so further analysis could be done. Alongside the 3D protein modeling, analysis of intergenomic and promoter sequences was crucial for the complete understanding of the protein structure. For all of the analyses, 16 different tools/servers have been used, 2 genes (*matE* and *iaaL*), and corresponding proteins (MatE and IaaL) for each of the genes in 3 species (6 proteins of focus in total and additional organisms from the same and different species containing the same protein of interest). Tables and figures were created in a manner to show a comparison of the analyzed subjects and to better understand their influence on virulence in tomatoes and kiwi.

### 4.1. BLAST TOOL SEARCH

By the BLAST analysis of the incomplete *iaaL* gene sequences and corresponding proteins of *Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* DC3000, *Pseudomonas savastanoi* pv. *savastanoi* are presented in 3.1 and 3.2. The *iaaL* and *matE* genes and corresponding proteins were analyzed in the terms of the maximum score achievement,

query cover, percentage identity, and the accessed length for the analysis.

Through the analyzes we conducted, we intended to search for the correlation of our protein of interest with the same proteins from different species. By making correlations and deeper analyses, structural and functional connections and dependence on other molecules can be fully understood, as well as interactions between similar proteins. In our research, it was the starting point for all other analyzes.

Table 3.1 shows the first three results of the search for homologs of each partial protein AXK59710.1, AAO53915.1, and RMS74473.1 (indoleacetic acid lysine ligase - IaaL) from *Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* str. DC3000 and *Pseudomonas savastanoi* pv. *savastanoi*. We confirmed that all of them are highly similar to AMP-binding proteins in other organisms in the same genus. There are no gaps in any of those alignments. Query coverage is near 100% with few mismatches and a high percentage of identity (more than 99.5%). It means that the number of identical residues (matches) relative to the length of the alignment is high, and is expressed in percentages of similarity between the aligned sequences. This makes our conclusion that the proteins among homologs of proteins being highly similar to the AMP-binding proteins are relevant. More specifically, all of the proteins used in the analysis are classified as AMP-binding proteins, having the MatE protein sequence and belonging to the *Pseudomonas* group. For the *iaaL* gene, since the query cover is 100%, our sequence of interest was completely analyzed, and the predicted maximum score was achieved. Percentage identity for gene and protein analysis differs in a term of higher identity for the nucleotide sequence. It is so due to the amino-acid complexity, for some amino acids different nucleotide base arrangement codes for the same amino acid. The query cover for the *matE* gene is significantly different due to the length difference between the query and comparison sequence and is only 39%. BLAST could compare our sequences of interest only for the length of the shorter sequence, even though the results we got to show more than 90% similarity between sequence parts compared. This shows that the sequence of our interest, the query sequence has the matching sequence part included in the bigger sequence. Accessed length is the maximum possible as well, giving us another insight into the high level of the detailed analysis performed by the BLAST. In the case of the *iaaL* gene and protein, accessed length is following the sequence to which comparison is done. In other words, we compared the sequence of *Pseudomonas syringae* pv. *actinidiae* with *Pseudomonas syringae* pv. *tomato* DC3000, and *Pseudomonas savastanoi* pv. *savastanoi*.

Table 3.2 shows the first three results of the search for homologs of each partial protein AXK59714.1, AAO53914.1, and RMS74471.1 (multidrug transporter MatE protein, also described as membrane protein) from *Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* str. DC3000 and *Pseudomonas savastanoi* pv. *savastanoi*. There are no gaps in any of these alignments. Query coverage and percentage identity are high (over 98% for IaaL from *P. syringae* group, 91% for MatE protein), and the mismatch number is low. This is again a strong indication of the protein conservation amongst the homolog protein sequences as well as of the structure and function. Sequences were incomplete so the length of the sequences compared is different. The protein sequences we firstly had were PCR products from the studies conducted by (Flores *et al.*, 2018), so further steps required full-length proteins to get reliable data.

Sequence identity is higher for the IaaL proteins than for the MatE, showing the slightly different structure of the MatE protein in various species, despite belonging to the same genus. As *P. s.* pv. *actinidiae* species are classified as highly active and aggressive bacterial strains, it is possible to relate its unique pattern of sequence with the adaptation and species-specific development. This implies the high need for the full analysis and understanding of the mechanism of action of the bacterial species.

BLAST analysis in general shows high similarity in sequence composition of both proteins for the *iaaL* and *matE* gene sequences. As previously mentioned, the sequences of both genes have been found in the cluster of genes, so the interconnection has similarities with the genes and proteins of other species which was expected. Furthermore, a slight difference in the sequence analysis of the *matE* gene showed a unique pattern in the sequence organization, but still a too high similarity with other sequences, so the parallel of the similarity in the process of virulence factor production and infection occurrence can be drawn.

The similarity of the sequence of *matE* and *iaaL* genes and protein variances in different bacterial strains implies the high correlation in the toxicity influence in the tomato and kiwi. More importantly, it is significant to note that there are still differences in their structure, indicating the possibly unique mechanism of action and influence on the plants. Sequence similarity indicates the overall structural and functional similarity, which is important for us because it gives us a good comparison example for the MatE protein of the *P. syringae* pv. *actinidiae* with the MatE proteins from other species. The same applies to the IaaL protein.

The first step of the key analysis of the *matE* gene and protein influence on the toxicity in tomato and kiwi started with the sequence similarity of bacterial species which are being analyzed. Since there was an initial high match between the sequences analyzed, further analyses could be done to fulfill our aim of identifying the toxicity influence on plant species. To confirm the BLAST results we got, we did a homology search for the MatE and IaaL proteins and found the complete protein sequences.

Table 3.3 shows the BLAST results of the protein sequence homology search for the IaaL protein of *Pseudomonas syringae* pv. *actinidiae* with two other species of interest: *Pseudomonas syringae* pv. *tomato* DC3000, and *Pseudomonas savastanoi* pv. *savastanoi*. The goal of the BLAST analysis was to find the complete protein sequences of the proteins and by so, to check how similar the sequences of the given proteins are in a term of the base-pairing and amino-acid sequence comparison, so their structural, functional, and evolutionary relationship can be determined. The same analysis is done for the MatE proteins and the results are presented in the table 3.8.

We found the sequences with the results close to 100% match and query cover, with a low mismatch number (max 3 mismatch). All of the proteins belong to the *Pseudomonas* group. Further on, we continued our work with the complete sequences for the proteins of interest with the 100% sequence identity. Complete BLAST results can be found in the Appendix 11. The partial and full sequence accession numbers we focused on are available in the 2.1.

#### 4.2. MULTIPLE SEQUENCE ALIGNMENT

Multiple sequence alignment (MSA) is used to compare more strains of the MatE protein at the same time, so the similarity, differences, evolutionary development, and conservation between them could be determined. Observing MatE protein sequences from the different organisms from *Pseudomonas* species and outgroups at the same time helped us to do a more detailed analysis of the common functional features of the MatE protein and thus influence the toxicity in plants. MSA was performed using ClustalW and T-coffee servers. Additional COBALT server from the NCBI was used to double-check the protein analysis. Sequences used in MSA analysis were retrieved from NCBI (PubMed) and UniProt.

The focus of the analysis is on the conserved, identical, or highly similar amino acid residues of the MatE proteins in *Pseudomonas* species, especially in the *Pseudomonas syringae* pv.

*actinidiae*. An asterisk (\*) under the results in the figures indicates that the residue is the same across all query proteins; a colon (:) indicates that the residues are different but highly similar across query proteins, and a dot (.) indicates that residues are different but similar to some degree, across query proteins.

After we got the complete protein sequences, we were able to do the multiple sequence alignments for the MatE proteins and IaaL proteins, separately.

The first MSA analysis we did was ClustalW multiple sequence alignment of IaaL (indoleacetate-lysine ligase) proteins from *P. syringae* pv. *actinidiae*, *P. syringae* pv. *tomato* DC3000, *P. savastanoi* pv. *savastanoi*, and outgroup organisms. 3.1 shows the sequences of *Salmonella enterica*, *Escherichia coli*, *Arsenophonus nasoniae*, and *Erwinia mallotivora* used as an outgroup organism for better insight into the differences between the total length sequences. The starting part of the protein sequence has a less similar sequence and no highly conserved amino acids. Comparison with the outgroup organisms resulted in many gaps. We showed three parts of the sequences aligned. The first one is from the beginning of the sequence and contains many gaps and low conservation levels. Gaps are present at the beginning of the alignment, not having the same start position and amino acid Methionine. There are only three positions marked by a colon (:) in the first alignment part, indicating that the residues are different but highly similar across query proteins. All other positions are without any conservation indicator due to the big difference and gaps in the alignment. The remaining two parts of the sequence aligned had a different conservation levels. It was much higher, the matching score was significant, and having the middle part of the sequence is an indicator of the same conserved domain existing among the species. Gaps are present in the second part of the alignment presented in the figure as well, referring to the non-existence of the residues in the *Pseudomonas* species aligned with the *Salmonella*, *Escherichia*, and *Arsenophonus* species, but still might have a functional common domain with the outgroup organisms due to the partial sequence conservation. The third part of the alignment presented shows a higher conservation level between different species. It is the indicator of the unique protein structures across the species which are characterized by the same function.

When the same analysis is performed without the outgroup organisms, the conservation level of the alignment was much higher. Sequences aligned had no gaps. Again, three parts from the alignment were taken so that better insight into the sequence can be made. There is a

difference in residues of the aligned sequences marked by the colon (:). This means that our sequences have different residues but are still highly similar across the species. Still, the sequences are almost identical, and it is no surprise since we showed that IaaL proteins are also denoted as the AMP proteins which are conserved.

COBALT results for the IaaL proteins MSA analysis with the outgroup species were no different from the ClustalW results. The alignments we got were the same 3.3, with the gray color in COBALT results marking the part of a sequence that has a gap and thus differs from the other sequences. The red color marks the identical and conserved parts, which in our case were the rest of the aligned sequences. Also, the blue parts in the alignment mark the polymorphism residues across the species and organisms. It shows us the regions where mutations inside the sequence occurred and led us to investigate sequences in more detail once we wanted to make a 3D model of a protein.

T-coffee MSA analysis for the IaaL proteins of different organisms 3.4 showed results as we got in a ClustalW MSA analysis. There is a low conservation level between the outgroup organisms and the *Pseudomonas* species, with many gaps present especially at the beginning of the sequence. This MSA did not show us polymorphism residue positions, but only the conservation levels and gaps.

The same tools were used for the MSA analysis of the MatE proteins.

ClustalW MSA analysis was performed for the MatE protein from the *Pseudomonas* species and outgroup species. The results we got are slightly different from those for the IaaL proteins. Highly identical and conserved residues or sequence parts were noted by the asterisk (\*) and more residues that are different but highly similar across query proteins were noted by the colon (:), as for the IaaL proteins. These are indications of the sequence's unique pattern and amino acid conformation, which are still slightly different in terms of their protein structural configuration, even though their function is so far believed to be the same. The analysis was performed with the outgroup organism of *Glycine max*, in which we found a protein with the same description but a partially different sequence. 3.8 contains three parts of the MatE protein sequences analyzed. The first part of the sequence has many gaps, with no matching parts between *Glycine max* and *Pseudomonas*. The second and third parts do show a significant conservation level, still containing the gaps. But, the middle part of the alignment shows a highly similar and conserved sequence, indicating the existence of a functional

domain that is conserved and might be the key factor in the virulence function of the bacteria. We have to note that the *Glycine max* is a very different organism, and the MatE protein it contains might be identical to our proteins of interest by description name but does not have the completely same function as a MatE protein from the *Pseudomonas*. This could also be one of the explanations for many gaps in the sequence comparison result. Another explanation for the gaps in the different functions and structures, but common specific functional domains, for which we found a match by analysis. The second and third parts of the alignment results presented in the 3.8 have a higher conservation level, even though there are still gaps present. This also matches with our implication that MatE proteins of *Pseudomonas* and *Glycine max* have common functional domains, but the rest of the sequence is different. Currently, the aim was to see the sequence comparison between *Pseudomonas* and outgroup species. The sequences we analyzed are full-length sequences from different *Pseudomonas* groups, so the data are more relevant and wide.

Additionally, the COBALT analysis was performed for the MatE proteins as well 3.9. The results we got were following the ClustalW results. Gray parts indicate the gap present in the sequence part which is absent from other species except in the *P. syringae* pv. *actinidiae* MatE efflux protein. Also, COBALT did not show us all of the residue polymorphisms present in the MatE protein sequence, but according to the T-coffee MSA analysis presented in the 3.10, conserved residues are marked and show us that not the complete sequence is conserved across the species, with many gaps present and residue polymorphisms. Differences in the residue structure of the sequence might influence the overall protein activity, as well as its folding pattern. Parts of the sequences analyzed by a T-coffee server are classified as „bad“, others as „average“ and some as „good“. A most central part of the sequence is „good“ alignment with many conserved residues among the different or inside the same species. This result is following our previous statement that the central conserved domain in the MatE proteins is present among the *Pseudomonas* species. There is a significant amount of „average“ quality alignment with the conserved residues inside the species. A „bad“ quality alignment is assigned for the *Glycine max* sequence since it is the „outer group“ and is used for the comparison of the gene conservation with the other species. It has many different residues in its protein sequence and thus is classified as a “bad” part of the alignment.

Multiple sequence alignment (MSA) as a continuation of the analysis for three species of interest and influence on toxicity in tomato and kiwi brought us valuable insight into the

structure and function of the MatE and IaaL proteins. As it can be seen, sequence structure and similarity between the species of interest are unique and partially intersected with other species, so the specific protein function and influence on the toxicity levels in the *Pseudomonas* species can be questioned. The conservation level, percent identity, and presence of gaps at specific positions help us conclude the common structure and function in the biological processes, including the influence on the toxicity of MatE proteins in the kiwi and tomato plants, as well as their differences in sequence structure which might influence the function. Every analysis that follows is based on and guided by the results we got in the previous analysis.

#### 4.3. PERCENT IDENTITY MATRIX

Percent identity matrix (3.5 for the IaaL proteins and 3.9 for the MatE proteins) gives us the percentage of the identity across the species and proteins analyzed, enabling us to see the percentage similarities between aligned sequences. The identity is in correspondence with the results obtained by the ClustalW software, and shows high identity for *iaaL* gene-encoded proteins, while identity is lower for the *matE* gene-encoded proteins.

IaaL proteins are more than 97% identical, but less than 100%, showing the polymorphisms present in the protein sequences. Even in the extended matrix table 3.6, there is no 100% match between the sequences. In the extended matrix are available results for the outgroup organisms as well, but with the low identity with the *Pseudomonas* species, up to 40%.

For the MatE proteins, the least similarity is with the MatE protein from *Glycine max* (Soybean), being only around 23%, while *Pseudomonas* species are all higher than 87% identity. Once again, the structural identity is shown, so the functional analysis could be continued on the protein interactions in the toxicity of the *Pseudomonas* species. Percentage identity of MatE proteins of *Pseudomonas savastanoi* pv. *savastanoi* and *Pseudomonas syringae* pv. *tomato* DC3000 is higher than 90%, while other queries of interest were around 87% and 88%. The higher similarity between those two species gave us a reason for the more detailed protein structure and function analysis of the MatE protein and its role in the toxicity and disease influence in plant species. When we see the extended percent identity matrix with more organisms from the *Pseudomonas* species, we can see that there are less than 100% identical sequences, confirming the MSA results which showed us the differences in the residue structure of the protein sequence (3.10).

#### 4.4. PHYLOGENETIC TREE

Phylogenetic tree analysis was performed using the Phylogeny.fr server. The analysis aimed to check for the length of the branches which represent the genetic distance between the organisms. A distance scale is several nucleotides per site in the alignment, so it gives a measure/scale of a genetic distance calculated as a sum of all the base pair differences between the sequences analyzed. It is presented statistically and 0.1 in length equals 1% statistically. Number 1 means a high level of confidence and positioning of the node in the topology of a tree (100% confidence level and genetic match between the sequences analyzed). Lengths of the branches are proximal to the length of time since the evolutionary divergence, too. Phylogenetic tree and genetic distancing enable us to follow the evolutionary connections between the organisms and sequences analyzed, their correlation, and possible functional or structural similarities.

The first phylogenetic tree is made for the IaaL proteins of *Pseudomonas* species including the outgroup organisms of *Arsenophonus* and *Erwinia* 3.5. The *Arsenophonus* is phylogenetically more distant than the *Pseudomonas* and *Erwinia*. We used 3 sequences of *Arsenophonus* and they show a distance between each other. More significant is the distance between *Pseudomonas* species which are evolutionary close, but still, have a genomic distance when phylogenetically compared. The values of the distance vary, from 0.99 to 0.27. The branch length is proportional to the number of substitutions per site. The number of substitutions we identified are not big but are significant for the further analyzes we wanted to perform. Even though the organisms from *Pseudomonas* species have a common origin and are closely related, the differences between them indicate unique sequences which might play a role in the protein function they carry.

Three IaaL proteins of interest for which we paid special attention to the evolutionary relationship are WP\_003380685.1 from *Pseudomonas syringae* pv. *actinidiae*, described in the phylotree as “P.syringae”, WP\_007247550.1 from *P. syringae* pv. *tomato* DC3000, described as a “P.syringae\_tomatoDC3000”, and WP\_057456329.1 from *P. savastanoi* pv. *savastanoi* is described as “P.savastanoi” 3.6. A key point is that our proteins of interest show a phylogenetic distance, as we showed in the MSA by the presence of polymorphisms.

Even though the *Erwinia* is an outgroup organism, it has an IaaL protein and is phylogenetically distant from *Pseudomonas*. It was an interesting way to see the correlation of organisms

in terms of genomic distance. The extended phylogenetic tree in 3.5 with the added proteins from the same family for the visualization and analysis together with the outgroup organisms show a significant phylogenetic difference. Confidence level and the genetic distance between the species of interest are similar to the closely related proteins from the same species. The confidence level and genetic distance are close to 0.99 in some cases, 0.85 in other cases, and 1 for outgroup species. Between the *Pseudomonas* species, it is 0.99, 0.85, and more. Additional proteins in the phylogenetic tree were used for the comparison and better visualization of the protein correlation with other proteins from the same species.

The phylogenetic tree for the MatE proteins with the outgroup organism *Glycine max* shows a long phylogenetic distance 3.11. The confidence level is more than 0.6, and the genomic distance is high. Even so, the extended phylogenetic trees give us a more broad representation of the distance between our species of interest and other species. The unique position and distance of our proteins help us to focus on more related species which are evolutionarily less divergent. By doing so, we could focus on the *Pseudomonas* species more.

Soybean MatE protein as shown in 3.11 is correlated with the other proteins analyzed, due to its specific origin. But, as the sequence of the protein is shown to be highly similar, the protein is used as an additional comparison and outgroup example. The MatE proteins are found in different species, and their phylogenetic correlation is inevitable. An interesting part to notice is that different bacterial species have more distant MatE proteins than *Pseudomonas*, but are close to each other. Even though *Erwinia* has the same protein, its evolutionary distance is much greater (distance is 0.9) from the *Pseudomonas* species than other included species. Small phylogenetic differences between the *Pseudomonas* species could mean that the proteins have a common functional domain, similar function, or interactions with other proteins. Thus, they are a valuable source of data for the comparison between them and our proteins of interest.

From the 3.12 we can see that the phylogenetic distance between MatE proteins of *Pseudomonas* species is small, but it is still present. Visually, we can divide the organisms analyzed into three groups according to the phylogenetic relatedness. From each of the groups, we took the one specific organism for which we made a 3D protein prediction, so we can see the differences between them in terms of the sequence and protein model.

Three MatE proteins to which we paid special attention in the phylogenetic tree

are WP\_003380687.1 from the *P. syringae* pv. *actinidiae*, described in a phylotree as a “P.syringae\_MT1”, WP\_054995848.1 from *P. syringae* group, described as the “P.syringae\_group:MT4”, and WP\_005782319.1 from *P. savastanoi* pv. *savastanoi*, described as the “P.syringae\_group\_MT2”.

The phylogenetic tree for the IaaL and MatE proteins showed us the distance and relation of our proteins of interest and proved to us that they are not the same proteins with the different accession numbers but real proteins found in the *Pseudomonas* species. By knowing this, we were able to continue our analysis with the proteins of interest and draw conclusions.

After BLAST, MSA, and phylogenetic tree construction, it was possible to build the 3D model of the proteins of interest. The first model was the MatE efflux family protein of *Pseudomonas syringae* pv. *actinidiae*. Models were made by the usage of Swiss-model and Phyre-2 servers.

#### 4.5. 3D PROTEIN PREDICTION

This step is significant for the overall analysis of the MatE proteins and structural changes they might have in different organisms from the same species. Sidechains interactions and sequence polymorphisms can lead to a different 3D folding pattern, so the function of the protein might be influenced. This was also the starting point of the analysis of our proteins of interest since the differences in the residue alignment are shown through the MSA analysis and phylogenetic tree distancing.

We made three analyses of three different proteins from *Pseudomonas* species, with the accent on the MatE protein from the *P. syringae* pv. *actinidiae*. Later on, secondary structure sequences were analyzed in detail to see the loop and alpha-helix structures and differences, as well as the Ramachandran plot analysis.

According to the Swiss-model protein prediction tool, our 3D model of the first protein analyzed was MatE protein from *Pseudomonas syringae* pv. *actinidiae*. It is shown to be a monomer, consisting of one single unit which is an alpha-helix structure interconnected with loops. More specifically, out of 10 alpha-helices interconnected by loops. Alpha-helices are organized in the tube-like formation, creating the specific form of a protein that is transmembrane and has a function in the transport of molecules through the cell membrane. This is another indicator of the *matE* gene-encoded protein having a role in the toxicity and disease occurrence in tomato and kiwi by transporting the virulence factors inside the cell, as

suggested by the analyses by other researchers and explained in the introduction part.

The Swiss-model works based on one template which is the most similar and reliable with the protein of interest from the data basis. In our case, it was the MATE transporter VcmN template protein from the *Vibrio Cholerae*.

In the upper right corner of 3.13 representing the 3D model prediction of MATE efflux protein is located the QMEANDisCo Global value of the protein analysis. The QMEANDisCo tool can derive both global (entire structure) and local (per residue) absolute quality estimates based on one single model. It uses QMEAN as a basis. QMEAN in general is a composite scoring function describing the major geometrical aspects of protein structures. In short, together with the GMQE value which is next to the QMEAN value, it estimates the accuracy of the tertiary structure of the resulting model (Benkert, Tosatto and Schomburg, 2008).

In general, the “higher” the value is, the better the score is, even though the values above 0.7 are considered reliable for the final quaternary prediction. The value we obtained is 0.67 with the  $\pm 0.05$  value, being on the edge of the expected value, so we conducted additional analysis by other servers to additionally check tertiary and quaternary structure.

The sequence identity of the template and query sequence is only 20.93%, while the sequence similarity is 31%. Values obtained are low, meaning that the template used is not the perfect match, but according to the other available templates, this one had the highest identity and coverage score. This also shows the need to find better templates and analyze the proteins more by other means and methods. The coverage score of the sequence is 96%, and the figure depicts the alpha-helices of the sequence as well. The next figure shows the model-template alignment and the identical residues marked by the black bold residue letters, as well as the predicted helix position according to the sequence. Not many identical residues were found between the template and the model, but it was expected since the sequence identity is only 21% 3.14.

3.15 gives us the representation of the helical 3D shape of the MatE efflux family protein analyzed by the Swiss model. In total 10 alpha-helices are interconnected by loops into the “tube” -like shape. The model is built following the model-template alignment, and the prediction is similar to the model built by the PHYRE2 server. The tube-like model is favorable for the transport proteins which are part of the transmembrane transport system. In

the case of our protein of interest, that is exactly the structure we expected as we are analyzing the toxicity effect of *matE* gene-encoded protein variants. Alpha helices are interconnected by loops and visually the helices are of closely uniform length.

Further analysis included a Phyre2 server for the control and better understanding of the MatE protein structure and possible function. By having more detailed controls, the reliability of the obtained results increases. According to Gosto (Gosto, 2018), Phyre2 software showed a low deviation from the X-ray crystallography protein structure prediction when used for the 3D protein prediction analysis.

3.16 presents the MatE 3D protein model from *Pseudomonas syringae* pv. *actinidiae* by Phyre2. Sequence identity is the same as in the Swiss model, 21%, and coverage is even bigger than in the Swiss model, being 100%, respectively. The number of aligned residues is also highly similar, while the template used for the analysis is identical. We can see that the protein is again having a tube-like formation and is made of alpha helices connected by loops. Several helices are approximately the same, visually counted since no automatic data is assigned to them. An additional detail we can see in the Phyre2 protein model is the red-colored active site of the protein, which we could not see in the model made by the Swiss-model tool. For us, it is interesting and important to see, because that is the region of a protein where the substrate or some other interacting molecule binds. By binding, it can influence the activity or the function of the protein. We will analyze the active sites later on, but for now, it is useful to see that we have them, and their approximate location.

Since the 3D models of the protein we made were highly similar, we wanted to see a more detailed structure of the MatE proteins from different organisms of *Pseudomonas* species. Differences in structure, especially in the loop region would indicate the differences in folding patterns, so before we continued further, 3D models of the proteins from *P. syringae* pv. *tomato* DC3000 and *P. savastanoi* pv. *savastanoi* were made. Their secondary sequence structures are first displayed and analyzed. We downloaded the secondary structure of the protein from the Phyre2 3D protein prediction results. The model is built according to the template sequence, and the protein sequence of interest is in the first row of the block in the 3.17. This figure only gives us insight into our MatE sequence of interest which we compared with two other MatE proteins from different organisms of the *Pseudomonas* species.

The first secondary structure figure 3.18 is the alignment taken from the Phyre2 server for

the Mate protein of *P. syringae* pv. *actinidiae*. Red parts encircled mark the difference in the structure with two other MatE proteins with which it is compared, MatE proteins from *P. syringae* pv. *tomato* DC3000 and *P. savastanoi* pv. *savastanoi*.

In total, there are 14 differences found in the structure. Some parts of the sequence are at the end of the alignment line and cross to the next line, so it seems to be more, but in total there are 14 different parts in the complete protein alignment. Some differences are only residue differences, others are missing the residue totally, or having a residue in a place where two other proteins do not have it. It is important to notice that differences in the residue structure are found in the loop regions of the proteins. Since loop regions are less stable than helix regions, it might be a strong indication of the influence on total protein folding pattern or influence on the protein function. We can not be sure since the model we have is made according to the 21% same template, which is not accurate enough. The Following descriptions are related to the figures 3.18, 3.19, and 3.20.

The first difference that we can see is at the beginning of the sequence. In the *actinidiae*, the first residues are EGS, while for the *tomato* and *savastanoi* it is KGS. A second difference is in the second block of the alignment. Only *actinidiae* has GSAA residues, while *tomato* and *savastanoi* have GSGV residues. The next difference is the presence of the residues in *actinidiae* which the other two organisms do not have. In the second block of the alignment and beginning of the third alignment, *actinidiae* has ML and E residues, missing only GG part (MLGGE), while the other two are completely missing these residues MLGAE, with the additional difference in the residues, having A residue instead of G as in the case of *actinidiae*. Another difference is in the GK residues, the third block of the alignment. *Actinidiae* and *tomato* do have these residues, while *savastanoi* is lacking them. The same case is with the I residue of the same block, *savastanoi* not having it while the other two have them. There are more differences like this. The next one is the fourth block of the alignment. *Actinidiae* has IVVFDLGLV missing and different residues, *tomato* IVVPDLGLV, while *savastanoi* has ILVFDLGLV. As it can be seen, some of them are the same and some are different, while all of them have the last V residue, not all of them have the I residue at the beginning. It is easier to visualize directly from the figures, but an important note to be made is that our proteins have differences between them in the sequence structures and loop regions, even though they are from the same *Pseudomonas* species. Similar polymorphisms occurred in the fourth block and next loop where the *actinidiae* sequence has different residues from two other

organisms. In this region, *tomato* and *savastanoi* have the same residues, but in the previous loop comparisons, they also had differences between them. At the end of the same block is the next difference, and the situation is the same as in the previous one. *Actinidiae* sequence has AVP residues while the other two organisms have ALPA residues and do not miss the last A residue, while *actinidiae* does not have it. The sixth block has the same way of differentiation between the organisms as in the previous case. *Actinidae* sequence has an AMNTTR residue sequence in the short loop region, while *tomato* and *savastanoi* have AMNTAR residues. In the same block, *actinidiae* and *tomato* have an AE residue, while *savastanoi* lacks them. This is another interesting difference showing the unique distribution and polymorphisms between the organisms inside the same species. It also might be an indication of the specific function of the proteins. Once again, in the sixth block, *actinidiae* has a unique polymorphism in the HVSG region, while two other proteins have HVAG residues. In the seventh block, there is a loop in which two residues are different in the *actinidiae* than in two other proteins. MGRPM loop in *actinidiae* has two M residues and is present in the structure, while the other two proteins have MGRPA residues, and the A residue, in the end, is not missing. In the seventh block, we have another difference in the structure. In the *actinidiae*, there is a G..A structure, while in *tomato* and *savastanoi* we have a GAHK residue with no missing parts. And lastly, in the eighth block, the last difference is at the end of the sequence. Reasons for this might be various, such as the isolation process or the different protein isolation and characterization techniques.

Overall analysis by both models resulted in significant results and enabled us insight into the 3D structure of the protein of interest. Knowing its 3D structure 3.17 and the presence of an active site can further determine its possible role and influence on the overall function and toxicity. Specific alpha-helices shapes interconnected by loops and organized into a tube-like structure gave valuable data to this work. Moreover, secondary structure analysis and differences in residues in the loop regions showed us that there are differences between our proteins of interest which might be the key part of the protein function and interaction. Thus, it might be the reason how the MatE proteins play a role in the toxicity of the bacteria. By creating a 3D protein model structure with all of the data available, including QMEAN values, Ramachandran plot, and sequence similarity analysis we could see that variants of the MatE protein for the structurally highly similar proteins between different species, leading to the conclusion of a functionally same role in the process of disease occurrence in tomato and

kiwi, with the possible difference in the level of this functionality and overall role.

Our next step was an analysis of the Ramachandran plot of the MatE proteins and their comparison. Firstly, we made a Phyre2 3D model of the MatE protein from *Pseudomonas syringae* pv. *actinidiae* and two other proteins, since the Ramachandran plot results are automatically assigned by the same server.

Conservation analysis is based on the Jensen-Shannon divergence approach for conservation analysis. At the end of the analysis, conservation scores are calculated. The calculation is done per base, indicating how many species in a given multiple alignment match at each locus.

On the right side of 3.22 is an indication scale of how conserved the sequence structure is. As shown, the structures analyzed are colored in blue/purple indicating a low conservation level, with a few parts colored light green/light blue, indicating a medium level of sequence conservation. The results can be explained by the software's inability to find a better matching template sequence according to which our model of the protein of interest could be built since there was a low identity level amongst the templates we found and query sequence. Even though BLAST analysis resulted in a high conservation description of a sequence, it was based on the specific primary sequence-structure analysis when compared with the genes and proteins which are highly similar in origin. In this case, the template used for our query sequence is different from the previously used comparison sequences so the results obtained in terms of a conservation level are different. It is important to mention that 3D structure prediction is as accurate as possible according to the available data, and the template we used now has been automatically assigned from the database according to the input data. There was no opportunity to choose a template based on the *matE* gene or protein sequences such as *Pseudomonas syringae* pv. *tomato* DC3000 or *Pseudomonas savastanoi* pv. *savastanoi* for the query sequence of *Pseudomonas syringae* pv. *actinidiae* analysis, as we did for the BLAST analysis. Even though our template is only 21% identical to the MatE protein of interest, the conservation level has a low or medium conservation scale, which in turn shows the need for a better and wider analysis of the *actinidiae* strain.

More detailed analysis by Ramachandran plot and related 3D protein structures reveals the significance of the study and accuracy of the models predicted. As we previously mentioned, we made 3D models for each MatE protein of interest from three organisms of *Pseudomonas*

species. By Ramachandran plot, we analyze the statistical distribution of the combinations of the backbone dihedral angles  $\phi$  (phi) and  $\psi$  (psi). The plot shows us which values of the Phi/Psi angles are possible for the amino acids of interest, their orientation in the model, and favorable combinations of those amino acids. Phi defines rotation about the CA-N bond of the residue, and psi defines rotation about the CA-C bond of the same residue. Dark areas presented in the plot of our model are “core” regions where the most favorable combinations are located. Small dots represent the residues at the positions corresponding to the residue’s phi and psi. According to the Ramachandran plots we got, there is a high concentration of the residues in the “dark” areas which is the most favorable orientation of the amino acids. Scores of the Ramachandran plots are considered to be good if they have values obtained at a resolution between 2-2.5Å, and our models are created at the resolution of 2.21Å. The resolution we got is exactly in the quality range. Since it was an automatic analysis, we could not choose different resolutions for our analysis, and automatically did analysis gave us valuable results. It is interesting to note that dots which are outside the dark regions and present the residues in the less favorable areas and combinations are present in all three species, but the general  $\phi$  (phi) and  $\psi$  (psi) angle orientation is good. Slight differences in the number of such residues are present as well, but they are following the differences in the sequences of the proteins. Together with the Ramachandran plot analysis are the 3D structures of the proteins we used. All three proteins have a tube-like structure, visually similar shape, consisting of alpha-helices interconnected by loops 3.23, 3.24 and 3.25.

To have better insight and reliable data, the Swiss model was built for the 3D Indoleacetic acid lysine ligase protein (IaaL) encoded by the *iaaL* gene of *Pseudomonas syringae* pv. *actinidiae*. This analysis is performed to check for the *iaaL* gene-encoded protein variants models and structural relatedness, since both genes, *matE*, and *iaaL* are found to be interconnected and in the same gene cluster. We wanted to see the difference in the 3D shape of the proteins we were working with.

The template sequence used in this analysis is from the enzyme from the ligase family, not an organism as in the previous example. The best template is given automatically by the Swiss-model tool, with the highest available identity and similarity ratio.

QMEANDisCo Global value is close to 0.54, which is on the edge of reliable data”, but provides us with the valuable info and the best possible match that could be found globally

for the tertiary structure prediction (when the sequence is globally analyzed). For the local (QMQE) analysis of the Indoleacetic acid lysine ligase protein (IaaL), the value is a bit lower than the global, less accurate, but the best possible we could get.

The structure of the IaaL protein encoded by the *iaaL* gene is homo-dimer, composed of alpha-sheets and beta strands interconnected by loops. The analysis is done on the 1.9 Å, which is close to the optimum value of 2.0 Å. Sequence similarity is 29%, while the sequence identity is only 20.60%. Coverage of the sequence during the analysis is 85%, almost full range of a sequence.

Ramachandran plot shows a similar distribution of the residues as for the MatE protein, with the highest density of residue distribution in the “dark” area where it is the most favorable position. There are also several residues in the less favorable area, but the focus is on the majority of residues in the dark area.

IaaL protein of *Pseudomonas syringae* pv. *actinidiae* is more complex than the MatE protein of the *Pseudomonas syringae* pv. *actinidiae*, having two types of structural units in monomer, and more residues oriented in the Psi/phi formation. Such complexity and difference are high due to the different species analyzed, as well as different additional roles and functions in cooperation with other genes. It is important for us in this research due to the fact it is a part of the same cluster and interconnected with the MatE protein family, as well as in the structure of the transport machinery for the bacteria, which makes it a helpful protein for the comparison and overall protein family analysis. Another reason which is even more important, as previously mentioned in the section of *iaaL* and *matE* gene explanation, functions of these two genes are interconnected. The presence/absence of one gene influences the activity of the other. The difference in the structure and their interconnected functions are an important aspect of the disease treatment methods in the future. The function of the IaaL proteins is different, but protein activity might have a common regulatory part with the MatE protein. The 3D structure difference and Ramachandran plot analysis of both proteins of the same species were important for the sake of the comparison of interconnection analysis of those two. They are structurally different but functionally might be connected, belong to the same cluster, and comparing them with other variants can be drawn a parallel between them, as well as the more detailed functional differences and similarities.

#### 4.6. PROTEIN DOMAINS IDENTIFICATION

Identification of protein domains has a significant role in the overall function analysis of the protein. The importance of the protein domains, especially if they show to be conserved, is in their functionality. If a conserved protein domain with a known function is found to be present in the query sequence, then the function of the whole sequence and the protein can be determined. Three different servers from InterPro resources were used for the protein domain analysis: PROSITE, Pfam, and SMART. In further analysis, the PROSITE server gave no results, so the other two were used to obtain data.

Pfam is the first server to give us the results. It is a database of curated protein families each of which is defined by two alignments and a profile hidden Markov model (HMM)(Finn *et al.*, 2013).

3.26 and 3.27 shows us the significant Pfam-A match for our query. It is a MatE domain, and the HMM length is 161, analyzed from 5 - 153 residues with low E-value, meaning analysis provided us with good results.

Green-colored sequences tell us how well the query sequence matches the HMM definition for our protein family. In our case, the query sequence 100% matches the HMM definition of the protein family according to which the analysis is done. Pfam shows two MatE domains confirmed by the SMART analysis as well. These two conserved domains found across the species are high indicators of the functional importance of our protein of interest and good structure prediction results.

SMART analysis showed the same domains as Pfam, two MatE domains, and two additional transmembrane proteins which were not the focus of this research. Analysis ranged residues 1-380 and had a significant E-value ( $4.8e-15$  and  $2.8e-15$ ), giving us reliable results. Our protein of interest was recognized as a transmembrane protein with many protein domains. Its localization was predicted to be a membrane, biological functions to be antiporter activity, and in biological processes, it is part of transmembrane transport. InterPro analysis was used as an additional analysis for the MatE protein of *Pseudomonas syringae* pv. *actinidiae*, while IaaL protein was not found in the database of this server. The results we got are found in the Appendix 7.

Sequences of MatE proteins from *Pseudomonas syringae* pv. *tomato* DC3000 and *Pseu-*

*domonas savastanoi* pv. *savastanoi* were also analyzed since the aim of this study is to check for the different *matE* gene variants' protein effect on the toxicity of bacterial strains. Protein domains analysis showed us that there is no significant variation in protein domains of our three species of interest 3.29. Domain functions and structures are highly similar, all three are transmembrane domains of the same length and residue position. According to these results, we can see that all three variants have the same or highly similar role in the disease development of tomato and kiwi species since their transmembrane activity enables them to inject the virulent proteins inside the plant cell and induce disease. Different variants of the MatE protein across three bacterial species are shown to have slight differences in their structure, but overall to be highly similar in structure, and residue positioning, enabling further analysis of the functional similarities as well.

#### 4.7. INTERACTOME ANALYSIS

Interactome analysis gives us the total molecular interaction scheme of our protein of interest with other proteins. There are different levels of confidence on which it is possible to perform the analysis. Those offered by the server from the STRING database are 0.400 and 0.900.

For the MatE protein of *Pseudomonas syringae* pv. *actinidiae* results we got on the confidence level of 0.900 are three interacting proteins-“*iaaL*” protein, Indoleacetate-lysine ligase from *Pseudomonas syringae* pv. *tomato* DC3000, Outer membrane porin, OprD family from *Pseudomonas syringae* pv. *tomato* DC3000, and Mate efflux family protein from *Pseudomonas syringae* pv. *tomato*. Wider interactome analysis showed broad interacting proteins with the IaaL. This indicates that IaaL protein could have various functions and functional interactions, depending on the process it is involved in. The MatE protein has a much different scheme of interaction. Its network is limited to two interacting proteins. Another reason could be an insufficient amount of available data on the MatE protein research. There are three colors of the lines by which proteins are connected. Green color marks gene neighborhood, blue color marks gene co-occurrence, and light green color mark text mining.

Interaction of the MatE protein with the “*iaaL*” protein is shown to be the strongest, depicted by three “colors” - connections for gene neighborhood, gene co-occurrence, and text mining reference. Second interaction from *P. s.* pv. *tomato* DC3000 which was significant was with the MATE efflux family protein that had gene co-occurrence interaction. The third interaction was with the Outer membrane porin, OprD protein family, and was indicated by the gene

neighborhood interaction. Gene neighborhood indicates gene pairs that encode interacting or functionally related proteins (*Gene neighbourhood*, no date).

The interactome analysis of *Pseudomonas syringae* pv. *tomato* DC3000 and *Pseudomonas savastanoi* pv. *savastanoi* was also performed, but since the only available template in the STRING database was a sequence of the *Pseudomonas syringae* pv. *tomato* DC3000, all of the obtained results were the same. Thus, interactions of the query sequence have all matching and highly similar interactome groups according to the available bioinformatics tool.

Summarizing the interactome analysis, we can see that the strongest correlations are between the proteins encoding for the functionally similar proteins, giving us a great clue about the high correlation of different *matE* gene variants and corresponding proteins on the toxicity of *Pseudomonas* strains.

Functional, interactome, and 3D protein prediction analysis showed a direct structural and functional correlation of the proteins, so their influence on the disease development in plants can be conclusive as well.

#### 4.8. SUBCELLULAR LOCALIZATION ANALYSIS

According to the protein localization, it is possible to further analyze and draw more conclusions on the protein function and interaction. For the MatE and IaaL proteins, two different servers were used, and the results we got from both servers were the same 3.12.

According to the analysis, all of the MatE proteins are located in the cell's inner membrane, while the IaaL proteins were recognized as cytoplasmic proteins, except for the *Glycine max* protein for which data were not available. *Glycine max* protein was used as an additional protein in the analysis due to the presence of the *matE* gene and relevant proteins. LocTree3 analysis provided us with additional details on the score and accuracy of the analysis. In all of the cases obtained results have a high level of the accuracy of the protein location. The results we got are the same in both analyses which imply the reliability of the analysis. Based on the LocTree3 analysis, MatE proteins are inner membranous, while the IaaL proteins are cytoplasmic. All of the accuracy and score levels are high, ranging from 96% to 100% accuracy and score, which gives us the valuable and reliable results of the analysis.

Being located in the inner membrane and cytoplasm is following the main function analysis

and mechanism of action of the proteins in their role in toxicity and disease development in the plant. As they are proven to be located in such a position to influence the action of other proteins and virulence factors, it is visible that genes of interest and consecutive proteins are playing an important role in the disease occurrence. High similarity, structural correlation, and domain analysis resulted in data that led us to believe that function of different variants might be the same, including their influence on toxicity in tomatoes and kiwi, with a slight difference in the intensity of this influence and spectrum of the influence range the MatE protein from *P. syringae* pv. *actinidiae* might have.

#### 4.9. DOCKING SITE PREDICTION

Docking site prediction represents an important aspect of protein function analysis. In the case of *Pseudomonas syringae* pv. *actinidiae* interactions, "OLC" ligand is interacting with the MatE protein of *Pseudomonas syringae* pv. *actinidiae*. The biochemical function of the OLC ligand is binding identical proteins, and its biological function is ion transmembrane transport. Also, the OLC ligand is found to be located in the cell membrane. According to the ligand properties and function, it is possible to draw a connection between the protein function and ligand binding. When bound, ligand enables MatE protein to transfer the virulence factors into the plant and thus enable the bacterial influence on the plant disease development. Without ligands present, protein interaction and binding would be difficult, and the biological activity of protein altered. In this way, when bound to a ligand, proteins are easily activated and functionally play the main role in virulence factors transmission.

3.30, 3.31, and 3.32 show OLC ligand details and the number of protein chains it interacts with. Since it biologically activates the protein by transferring the signal to and from the cell, the higher the number of chains it interacts with, the better the binding affinity and signal transmission. OLC binds to 8 chains of the MatE protein, enabling strong signal transmission, more specifically – enabling the transmission of the virulence factors and bacterial influence on the cell. This analysis was one more indicator of the functional role of the MatE efflux protein. By having the same ligands as *matE* gene-encoded proteins from other variants and related species, we were able to conclude the variants of the *matE* gene's possible effect on the toxicity of *Pseudomonas syringae* pv. *actinidiae* in tomato and kiwi.

#### 4.10. MSA BETWEEN MORE DIFFERENT SPECIES

Multiple Sequence Alignment between more different bacterial strains shows us the diversity of species, even if they belong to the same family. Several different sequences containing MatE protein and IaaL protein were compared. The highest matching score is between our species of interest, IaaL proteins of *Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* DC3000, and *Pseudomonas savastanoi* pv. *savastanoi*. The same applies to the MatE protein comparison. Even though many compared sequences belong to the same species, there was a significant difference in their sequence matching due to the difference in the protein function. Not all of the proteins had the same protein localization and interaction, so the sequence match is different as well. 3.35 depicts the aforementioned sequence comparison and matching between different species. Only part of the compared sequence is shown with a high matching score and gaps available. Different bacteria species containing the MatE protein were compared. We have several parts of the sequences assigned by the asterisk (\*) meaning we have conserved parts of the sequence across the species, and several assigned by the colon (:) for the residues which are different but highly similar across query proteins. Another important note is that we have some sequences which are assigned by a dot (.), meaning they are different but similar to some degree. All of those help us to build an image on the sequences we compare, to see that our sequences of interest for the MatE and IaaL proteins are similar in structure to other proteins, but still unique, indicating their specific function. Even though their function and sequence structure is not the same, their functions might be influenced by each other, since we saw the same example for the genes. If one of the genes, *matE* or *iaaL* is silenced, then the other gene is dysfunctional as well. The same rule might apply to the proteins.

#### 4.11. INTERGENIC AND PROMOTER SEQUENCES CONSERVATION ANALYSIS

The final step in the analysis process of the *matE* and *iaaL* genes and their possible role in the toxicity of *Pseudomonas syringae* pv. *actinidiae* on tomato and kiwi was the analysis of intergenic and promoter sequences of these two proteins. The *matE* and *iaaL* genes are part of the same operon and transcribed from the same promoter. Changes in the intergenic and promoter sequences between different organisms in the same species might be the reason for a different influence and range of the influence of these proteins on other proteins in the bacteria. Also, differences in these regions might influence the slight changes in the 3D folding pattern

of the protein, so the complete interaction of the protein might be changed. We firstly analyzed intergenic sequences by the ClustalW server. The conservation level of the sequences was 100%, and no differences, gaps, or polymorphisms were found. High conservation between organisms indicates strong evolutionary relation, as well as the interspecies relation of the protein function. The intergenic sequence is 120 nucleotides long, and 84 sequences were analyzed.

For the promoter region, the same number of sequences was analyzed. The length of the promoter is 100 nucleotides. In this region, we found some changes in the structure. Polymorphisms were found on some spaces but still marked only by a dot ( . ), meaning the nucleotides are not identical but still highly similar across the species and organisms analyzed. Even the slight changes in the promoter could influence the transcription level and overall toxicity influence.

#### 4.12. CONCLUSION

Detailed analysis of the MatE proteins led us to different ways of research as we were progressing. Our initial assumption was that proteins from different organisms have different structures and functions, so the nucleotide and amino acid structures are different as well. By identifying their differences we would be able to conclude the MatE protein function of *Pseudomonas syringae* pv. *actinidiae* in the disease development in tomato and kiwi.

We used 14 different tools/servers for the analysis of *matE* and *iaaL* genes, and corresponding proteins for the *Pseudomonas syringae* pv. *actinidiae*, *Pseudomonas syringae* pv. *tomato* DC3000, and *Pseudomonas savastanoi* pv. *savastanoi* focuses on organisms from the *Pseudomonas* genus. We also used additional organisms that contain MatE and IaaL proteins from the same and outgroups species. By the tools we used we followed different data banks and high evaluation ratings of tools, so the results we got are as accurate as possible.

BLAST analysis resulted in a detailed check on all of the sequence local alignment details and internal properties of the sequences analyzed. The query coverage was high during the analysis, even though the percentage identity of the sequences was not always 100%. This is due to the different species from which the genes and proteins originate, as well as the difference in the sequence length analyzed. Yet, multiple sequence alignment resulted in a similar outcome, showing more sequence similarity between the analyzed species at once.

We could see the highly conserved regions of the sequence between all three species, as well as different amino acids with physicochemical properties whose arrangement is important for the function and ligand interaction. Additionally, the percent identity matrix showed the high identity between the genes and proteins of interest across the species, indicating high homology and conservation, which enabled us the 3D protein prediction and comparison of different variants in their role in the induction of toxicity. ClustalW, COBALT and T-coffee servers used for the MSA analysis resulted in the identical outcome from all three organisms of interest, with the only difference in the additional details that the specific server provides. With all of the results from the BLAST tool analysis, we could see that the same proteins from different organisms have different descriptions. IaaL protein is described as the “AMP binding protein” across the species, and we showed it by MSA and detailed BLAST search of the PDB. Also, the IaaL protein is sometimes described as the “indoleacetic acid lysine” protein in the *Pseudomonas syringae* pv. *actinidiae* while in the *P. syringae* pv. *tomato* and *P. savastanoi* is described as an “indoleacetic acid ligase” protein. Important notice is that both of them are the same protein having the same properties, but due to the various entries and misclassification in the PDB and Uniprot, are saved under the different descriptions.

For the MatE protein, descriptions were uniform across the databases given as the MatE efflux or detoxification protein.

With the results of the BLAST and MSA tools, the homologous sequences between the organisms of the same protein are characterized as highly similar and conserved. We found out that MatE and IaaL proteins are not only present in the *Pseudomonas* species, but the other species such as *Archaea* and *Eukarya*. Also, a high conservation level indicates the similar function of the proteins as well. But, even though highly similar, MatE protein sequences between the three organisms from the *Pseudomonas* species were not identical. Variants of the proteins according to their secondary structure might be a reason for the different influences on the toxicity in kiwi and tomato. The initial idea that protein structure might be more different was changed due to the high percent identity matrix and MSA. Even so, slight differences were an indicator of the unique physicochemical properties of the proteins.

Phylogenetic tree construction relies on the protein sequences of the species, so the tree branching is reliable in making a connection between the genes and proteins we analyzed. After we got the MSA results to be highly conserved, we had to be sure that we are working

with the different proteins, not the same proteins with the different descriptions and accession numbers. According to the phylogenetic tree, the smallest evolutionary distance is amongst the same proteins from different species (IaaL proteins are more closely related in between, as well as the MatE proteins). An exception was the *Glycine max* MatE proteins that had more distance since it is from the outgroup, not a bacterial species, but was used for the comparison. Differences in the branching at the phylogenetic tree showed us that there are variations between the proteins we analyzed, and have an evolutionary distance which indicated the unique structure between the organisms of the same species. Branching distance being small indicates a high similarity in the protein structure, but even the slight differences can influence the overall function of the protein.

According to the sequence comparison analysis, it was possible to build a 3D protein model of MatE protein of *Pseudomonas syringae* pv. *actinidiae*. SWISS-MODEL search resulted in a relatively low sequence identity of a query and a template sequence, but with high coverage. The reason for the low sequence identity could be the different lengths of the sequences compared, as well as the template source organism and lack of data in a database. Even so, the QMEAN value was good enough to proceed and build the model. The resulting model was in an alpha-helical structure oriented around the axis. An additional tool used, Phyre2 showed the same results and structure, with a percent identity of 21% as well. Thus, the structure of the protein could be depicted as an alpha-helical, 10 alpha-helices interconnected by loops. Additionally, the 3D protein prediction is made for MatE proteins of all three aforementioned species which have the *matE* gene.

Important to notice was the visually significant structural similarity and sequence coverage between the proteins, even though there was a low conservation level according to the template we built the model from. This is the implication of the protein sequence change and specific polymorphisms for each of the species of interest. Every protein has a unique sequence with polymorphisms in its structure. For this reason, visualization of the sequence and 3D models are highly important. The cross-species analysis shows that our protein of interest conserves its structural formation across the species, indicating the similar or same function and influence on the toxicity in the species we analyzed. The tube-like formation is characteristic of the membrane proteins, so the predicted model we got could be close to the real form of the protein. a 3D model of the protein might show us the result of the changes in the structure visible in the three-dimensional form of the protein.

A very important aspect of the 3D model prediction was the additional proteins' secondary structure control. Since the predicted models were similar, there was a need to see to which degree is the similarity implied. We found specific changes in the secondary sequence structure. Polymorphisms present in the loop regions of the three proteins of interest were the key discovery in the protein analysis. It was important because differences in the structure when performed by the MSA tools were present, but we did not know the location of the polymorphisms- if it was in the helix or the loop region. When we did a Phyre2 3D model prediction and compared secondary structures, the loop regions contained polymorphisms. Such differences might influence the specific 3D folding of proteins, and we might not see the differences in the predicted models due to the small query-template identity according to which the model is made. Thus, not only structural changes might be present between the proteins of interest, but specific differences in the MatE protein of *P. syringae* pv. *actinidiae* might be the reason for the higher level of toxins in the tomato and kiwi.

By creating a 3D protein model structure with all of the data available, including QMEAN values, Ramachandran plot, and sequence similarity analysis we could see that variants of the *matE* gene code for the structurally highly similar proteins between different species, leading to the conclusion of a functionally same or highly similar role in the process of disease occurrence in tomato and kiwi. Different variants having the same or highly similar way of functioning cause plants to be unable to use their defense mechanisms and sustain toxin resistance. Instead, tomato and kiwi (in our case) are being defeated by this complex mechanism of the *Pseudomonas* genus, which leads to disease development and total plant death, eventually.

Pfam shows two MatE domains existing in the MatE protein and are confirmed by the SMART tool analysis as well. These two conserved domains found across the species are high indicators of the functional importance of our protein of interest and good structure prediction results.

Protein domains analysis showed us that there is no significant variation in protein domains of our three species of interest. Domain functions and structures are highly similar, all three are transmembrane domains of the same length and residue position. Out of this data, we can see that all three variants have the same or highly similar role in the disease development of tomato and kiwi species since their transmembrane activity enables them to inject the

virulent proteins inside the plant cell and induce disease. Different variants of the *matE* gene across three bacterial species are shown to have slight differences in their structure, but overall structure to be highly similar, enabling further analysis of the functional similarities as well.

Interactome analysis showed that the strongest correlations are between the proteins encoded for the functionally similar role, giving us a great clue about the high correlation of different *matE* gene variants and corresponding proteins on the toxicity of *Pseudomonas* strains. Being located in the inner membrane space, *matE* gene-encoded proteins are shown to influence the action of other proteins and virulence factors, enabling their transfer across the membrane so that bacterial infection occurs. Also, the OLC ligand that interacts with the protein is located in the cell membrane, and without it, the full mechanism of action would not be possible. Knowing the mechanism of action and function of the ligand, it is possible to connect it with the protein's function in disease development. It is a complex interconnected mechanism that controls the overall toxicity of the bacteria. Since *matE* and *iaaL* genes are part of the same operon, dysfunction of one gene also causes dysfunction of another gene, so the proteins might be influenced by each other. More specifically, the function of the complete mechanism might be influenced. IaaL protein is cytoplasmic while the MatE is inner-membranous. If one of them is not functional, thus the whole process of disease induction is interrupted. Overall, disabling the function of one protein might be the key to solving the disease problem. Interrupting the process by changing one of the aforementioned proteins, especially the MatE protein, would disable the bacteria to deliver virulence factors inside the cell, so the disease process could be terminated.

The final aspect of the research on the MatE protein is the promoter and protein intergenic region analysis. Intergenic and promoter sequence polymorphisms between different organisms in the *Pseudomonas* species might be one more part that is different and unique, such as the changes we found in the loop regions. The 100% conservation level of the intergenic region did not give us any new clue about the possible difference in protein regulation and function. But, the difference in the promoter region suggests the different regulation levels of the MatE protein between the organisms. Even though the polymorphisms are not numerous, they might be the reason for a specific protein binding and regulation of the overall function of the protein.

Differences in the promoter region and loop region of the protein might be the reason for

specific protein activity, function, and influence on the interacting molecules and toxicity role of the bacteria. We found our protein of interest, MatE of *Pseudomonas syringae* pv. *actinidiae* has specific differences and so might have a different mechanism of action from other organisms in the same species, thus having a different influence on the disease development in the tomato and kiwi.

Sequence structure differences, phylogenetic distance, 3D model similarities and loop differences of proteins, similar functional domains and ligand binding activities, differences in the promoter residues, and the conserved intergenic region between two proteins are unique properties of each analyzed protein. Characterization of a MatE protein gave us a clue about the importance of the polymorphisms on the protein function. As we have a reference frame for the *Pseudomonas syringae* pv. *actinidiae* to be a *Pseudomonas syringae* pv. *tomato* DC3000 and *Pseudomonas savastanoi* pv. *savastanoi*, the possible functional similarities, and differences might be projected. This reference frame means available reliable comparison data from the species which are already investigated. Thus, comparison parallels can be drawn with the results we obtain and can be conclusive in the research.

The MatE protein variants encoded by the *matE* gene in the *Pseudomonas syringae* and *Pseudomonas savastanoi* organisms might have a different influence on the process of disease development in tomato and kiwi. The different influences might be due to the specific polymorphisms found in the secondary structure loop regions as well as in the promoter region of the *matE* gene. The overall mechanism of action, sequence structure, 3D protein structure, protein domains, ligands with which our proteins of interest interact, as well as their phylogeny, have unique properties, and the results we got indicate a uniformity in the function, but with the difference in the level of the influence on the disease caused by the *Pseudomonas* species. When the overall analysis is taken into the correlation, we can conclude that *Pseudomonas syringae* pv. *actinidiae*, even with a high conservation level of the sequence has slight differences in its sequence structure, promoter region, and protein loop residues, and thus the influence and role in the toxicity induction and disease development in tomato and kiwi could be unique and stronger than in other organisms from the same species.

The complete mechanism of action of the transport of the virulent factors across the cell membrane and alteration of the defense mechanisms of the plant remains unknown due to the many processes which are not well understood in the bacteria and plant and their interaction.

This research is another step closer to the full understanding of this complex bacteria and resolving the exact mechanism and reasons for the disease occurrence, even though further experimental analysis is necessary to get the in vivo results. The research aimed to understand the bacteria better so novel and effective ways for plant protection can be further developed. By understanding the MatE structure, strong relation with the IaaL protein, and unique properties, novel plant protection techniques might be developed, and future crops saved from the disease.



## REFERENCES

- Arnold, D. L., Pitman, A. and Jackson, R. W. (2003) 'Pathogenicity and other genomic islands in plant pathogenic bacteria', *Molecular Plant Pathology*. Wiley, 4(5), pp. 407–420. doi: 10.1046/j.1364-3703.2003.00187.x.
- Benkert, P., Tosatto, S. C. E. and Schomburg, D. (2008) 'QMEAN: A comprehensive scoring function for model quality assessment', *Proteins: Structure, Function, and Bioinformatics*. Wiley, 71(1), pp. 261–277. doi: 10.1002/prot.21715.
- Boch, J. *et al.* (2002) 'Identification of pseudomonas syringae pv. Tomato genes induced during infection of arabidopsis thaliana', *Molecular Microbiology*. Wiley, 44(1), pp. 73–88. doi: 10.1046/j.1365-2958.2002.02877.x.
- Brooks, D. M. *et al.* (2004) 'Identification and characterization of a well-defined series of coronatine biosynthetic mutants of pseudomonas syringae pv. Tomato DC3000', *Molecular Plant-Microbe Interactions®*. Scientific Societies, 17(2), pp. 162–174. doi: 10.1094/mpmi.2004.17.2.162.
- Castillo-Lizardo, M. G. *et al.* (2015) 'Contribution of the non-effector members of the HrpL regulon, iaaL and matE, to the virulence of pseudomonas syringae pv. Tomato DC3000 in tomato plants', *BMC Microbiology*. Springer Science; Business Media LLC, 15(1). doi: 10.1186/s12866-015-0503-8.
- Cerboneschi, M. *et al.* (2016) 'Indole-3-acetic acid in plant–pathogen interactions: A key molecule for in planta bacterial virulence and fitness', *Research in Microbiology*. Elsevier BV, 167(9-10), pp. 774–787. doi: 10.1016/j.resmic.2016.09.002.
- Chang, G. S. *et al.* (2008) 'Phylogenetic profiles reveal evolutionary relationships within the “twilight zone” of sequence similarity', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 105(36), pp. 13474–13479. doi: 10.1073/pnas.0803860105.
- Coburn, B., Sekirov, I. and Finlay, B. B. (2007) 'Type III secretion systems and disease', *Clinical Microbiology Reviews*. American Society for Microbiology, 20(4), pp. 535–549. doi: 10.1128/cmr.00013-07.
- Comai, L. and Kosuge, T. (1982) 'Cloning characterization of iaaM, a virulence determinant of pseudomonas savastanoi.', *Journal of Bacteriology*. American Society for Microbiology, 149(1), pp. 40–46. doi: 10.1128/jb.149.1.40-46.1982.

- Comeau, S. R. *et al.* (2004) 'ClusPro: A fully automated algorithm for protein-protein docking', *Nucleic Acids Research*. Oxford University Press (OUP), 32(Web Server), pp. W96–W99. doi: 10.1093/nar/gkh354.
- Dixon, R. A. (2001) 'Natural products and plant disease resistance', *Nature*. Springer Science; Business Media LLC, 411(6839), pp. 843–847. doi: 10.1038/35081178.
- Djami-Tchatchou, A. T. *et al.* (2019) 'Dual role of auxin in regulating plant defense and bacterial virulence gene expression during *Pseudomonas syringae* ptoDC3000 pathogenesis'. Cold Spring Harbor Laboratory. doi: 10.1101/2019.12.29.881581.
- Domain (2018). Available at: <https://www.uniprot.org/help/domain> (Accessed: 11 December 2018).
- Farhadi, T. (2018) 'Advances in protein tertiary structure prediction', *Biomedical and Biotechnology Research Journal (BBRJ)*. Medknow, 2(1), p. 20. doi: 10.4103/bbrj.bbrj\_94\_17.
- Finn, R. D. *et al.* (2013) 'Pfam: The protein families database', *Nucleic Acids Research*. Oxford University Press (OUP), 42(D1), pp. D222–D230. doi: 10.1093/nar/gkt1223.
- Flores, O. *et al.* (2018) 'Genetic and phenotypic characterization of indole-producing isolates of *Pseudomonas syringae* pv. *Actinidiae* obtained from Chilean kiwifruit orchards', *Frontiers in Microbiology*. Frontiers Media SA, 9. doi: 10.3389/fmicb.2018.01907.
- Fu, S.-F. *et al.* (2015) 'Indole-3-acetic acid: A widespread physiological code in interactions of fungi with other organisms', *Plant Signaling & Behavior*. Informa UK Limited, 10(8), p. e1048052. doi: 10.1080/15592324.2015.1048052.
- Galán, J. E. *et al.* (2014) 'Bacterial type III secretion systems: Specialized nanomachines for protein delivery into target cells', *Annual Review of Microbiology*. Annual Reviews, 68(1), pp. 415–438. doi: 10.1146/annurev-micro-092412-155725.
- Gene neighbourhood (no date). Available at: [https://www.ebi.ac.uk/ols/ontologies/mi/terms?obo\\_id=MI%3A0057](https://www.ebi.ac.uk/ols/ontologies/mi/terms?obo_id=MI%3A0057).
- Glass, N. L. and Kosuge, T. (1988) 'Role of indoleacetic acid-lysine synthetase in regulation of indoleacetic acid pool size and virulence of *Pseudomonas syringae* subsp. *savastanoi*', *Journal of Bacteriology*. American Society for Microbiology, 170(5), pp. 2367–2373. doi: 10.1128/jb.170.5.2367-2373.1988.
- Gosto, R. (2018) 'Assessment of accuracies of protein 3-dimensional prediction software', *Southeast Europe Journal of Soft Computing*. International University of Sarajevo, 7(2). doi: 10.21533/scjournal.v7i2.161.
- Green, E. R. and Mecsas, J. (2016) 'Bacterial secretion systems: An overview', *Microbi-*

- ology Spectrum*. Edited by I. T. Kudva. American Society for Microbiology, 4(1). doi: 10.1128/microbiolspec.vmbf-0012-2015.
- Hall, B. G. (2013) 'Building phylogenetic trees from molecular data with MEGA', *Molecular Biology and Evolution*. Oxford University Press (OUP), 30(5), pp. 1229–1235. doi: 10.1093/molbev/mst012.
- Henry, E. *et al.* (2017) 'Direct and indirect visualization of bacterial effector delivery into diverse plant cell types during infection', *The Plant Cell*. Oxford University Press (OUP), 29(7), pp. 1555–1570. doi: 10.1105/tpc.17.00027.
- Hu, M. *et al.* (2010) 'Indole affects biofilm formation in bacteria', *Indian Journal of Microbiology*. Springer Science; Business Media LLC, 50(4), pp. 362–368. doi: 10.1007/s12088-011-0142-1.
- Katagiri, F. and Tsuda, K. (2010) 'Understanding the plant immune system', *Molecular Plant-Microbe Interactions®*. Scientific Societies, 23(12), pp. 1531–1536. doi: 10.1094/mpmi-04-10-0099.
- Kelley, L. A. *et al.* (2015) 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nature Protocols*. Springer Science; Business Media LLC, 10(6), pp. 845–858. doi: 10.1038/nprot.2015.053.
- Kunkel, B. N. and Harper, C. P. (2017) 'The roles of auxin during interactions between bacterial plant pathogens and their hosts', *Journal of Experimental Botany*. Oxford University Press (OUP), 69(2), pp. 245–254. doi: 10.1093/jxb/erx447.
- Lam, H. N. *et al.* (2014) 'Global analysis of the HrpL regulon in the plant pathogen *Pseudomonas syringae* pv. *Tomato* DC3000 reveals new regulon members with diverse functions', *PLoS ONE*. Edited by D. Arnold. Public Library of Science (PLOS), 9(8), p. e106115. doi: 10.1371/journal.pone.0106115.
- Li, X.-Z. and Nikaido, H. (2009) 'Efflux-mediated drug resistance in bacteria', *Drugs*. Springer Science; Business Media LLC, 69(12), pp. 1555–1623. doi: 10.2165/11317030-000000000-00000.
- Lim, H. N., Lee, Y. and Hussein, R. (2011) 'Fundamental relationship between operon organization and gene expression', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 108(26), pp. 10626–10631. doi: 10.1073/pnas.1105692108.
- Lindeberg, M. *et al.* (2006) 'Closing the circle on the discovery of genes encoding hrp regulon members and type III secretion system effectors in the genomes of three model pseu-

- domonas syringae strains', *Molecular Plant-Microbe Interactions*®. Scientific Societies, 19(11), pp. 1151–1158. doi: 10.1094/mpmi-19-1151.
- Lindeberg, M., Cunnac, S. and Collmer, A. (2012) 'Pseudomonas syringae type III effector repertoires: Last words in endless arguments', *Trends in Microbiology*. Elsevier BV, 20(4), pp. 199–208. doi: 10.1016/j.tim.2012.01.003.
- Masuda, S. *et al.* (2006) 'Identification and functional characterization of a new human kidney—specific h+/organic cation antiporter, kidney-specific multidrug and toxin extrusion 2', *Journal of the American Society of Nephrology*. American Society of Nephrology (ASN), 17(8), pp. 2127–2135. doi: 10.1681/asn.2006030205.
- McAtee, P. A. *et al.* (2018) 'Re-programming of pseudomonas syringae pv. Actinidiae gene expression during early stages of infection of kiwifruit', *BMC Genomics*. Springer Science; Business Media LLC, 19(1). doi: 10.1186/s12864-018-5197-5.
- McCann, H. C. *et al.* (2013) 'Genomic analysis of the kiwifruit pathogen pseudomonas syringae pv. Actinidiae provides insight into the origins of an emergent plant disease', *PLoS Pathogens*. Edited by J. L. Dangl. Public Library of Science (PLOS), 9(7), p. e1003503. doi: 10.1371/journal.ppat.1003503.
- Mugnai, L. *et al.* (1994) 'The grouping of strains of Pseudomonas syringae subsp. savastanoi by DNA restriction fingerprinting', *Journal of Phytopathology*. Wiley, 142(4), pp. 209–218. doi: 10.1111/j.1439-0434.1994.tb00014.x.
- Otsuka, M. *et al.* (2005) 'A human transporter protein that mediates the final excretion step for toxic organic cations', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 102(50), pp. 17923–17928. doi: 10.1073/pnas.0506483102.
- Plewczyński, D. and Ginalska, K. (2009) 'The interactome: Predicting the protein-protein interactions in cells', *Cellular and Molecular Biology Letters*. Walter de Gruyter GmbH, 14(1), pp. 1–22. doi: 10.2478/s11658-008-0024-7.
- Preiter, K. *et al.* (2005) 'Novel virulence gene of pseudomonas syringae pv. Tomato strain DC3000', *Journal of Bacteriology*. American Society for Microbiology, 187(22), pp. 7805–7814. doi: 10.1128/jb.187.22.7805-7814.2005.
- Preston, G. M. (2000) 'Pseudomonas syringae pv. tomato: The right pathogen, of the right plant, at the right time', *Molecular Plant Pathology*. Wiley, 1(5), pp. 263–275. doi: 10.1046/j.1364-3703.2000.00036.x.
- Pseudomonas syringae* pv. *Actinidiae* (bacterial canker of kiwifruit). *Invasive species com-*

- pendium* (2021). Available at: <https://www.cabi.org/isc/datasheet/45002> (Accessed: 9 December 2021).
- Pseudomonas syringae* pv. *Tomato DC3000 bacterial spotting on a plant leaf*. (2021). Available at: <https://wustljcubes.wordpress.com/2015/06/22/> (Accessed: 22 June 2015).
- Pseudomonas syringae shown using SEM* (2021). Available at: [https://microbewiki.kenyon.edu/index.php/Bacterial\\_nucleation\\_in\\_pseudomonas\\_syringae](https://microbewiki.kenyon.edu/index.php/Bacterial_nucleation_in_pseudomonas_syringae) (Accessed: 13 December 2021).
- Puhar, A. and Sansonetti, P. J. (2014) 'Type III secretion system', *Current Biology*. Elsevier BV, 24(17), pp. R784–R791. doi: 10.1016/j.cub.2014.07.016.
- Rico, A. and Preston, G. M. (2008) 'Pseudomonas syringae pv. Tomato DC3000 uses constitutive and apoplast-induced nutrient assimilation pathways to catabolize nutrients that are abundant in the tomato apoplast', *Molecular Plant-Microbe Interactions®*. Scientific Societies, 21(2), pp. 269–282. doi: 10.1094/mpmi-21-2-0269.
- Roy, A., Kucukural, A. and Zhang, Y. (2010) 'I-TASSER: A unified platform for automated protein structure and function prediction', *Nature Protocols*. Springer Science; Business Media LLC, 5(4), pp. 725–738. doi: 10.1038/nprot.2010.5.
- Ruan, J. *et al.* (2019) 'Jasmonic acid signaling pathway in plants', *International Journal of Molecular Sciences*. MDPI AG, 20(10), p. 2479. doi: 10.3390/ijms20102479.
- Scortichini, M. *et al.* (2012) 'Pseudomonas syringae pv. Actinidiae: A re-emerging, multifaceted, pandemic pathogen', *Molecular Plant Pathology*. Wiley, 13(7), pp. 631–640. doi: 10.1111/j.1364-3703.2012.00788.x.
- Serizawa, S. *et al.* (1989) 'Occurrence of bacterial canker of kiwifruit in japan: Description of symptoms, isolation of the pathogen and screening of bactericides.', *Japanese Journal of Phytopathology*. The Phytopathological Society of Japan, 55(4), pp. 427–436. doi: 10.3186/jjphytopath.55.427.
- Shen, Y., Tang, J. and Guo, F. (2019) 'Identification of protein subcellular localization via integrating evolutionary and physicochemical information into chou's general PseAAC', *Journal of Theoretical Biology*. Elsevier BV, 462, pp. 230–239. doi: 10.1016/j.jtbi.2018.11.012.
- Sievers, F. *et al.* (2011) 'Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega', *Molecular Systems Biology*. EMBO, 7(1), p. 539. doi: 10.1038/msb.2011.75.
- Sisto, A. *et al.* (1999) 'Isolation and characterization of pseudomonas syringae subsp. Savas-

- tanoi mutants defective in hypersensitive response elicitation and pathogenicity', *Journal of Phytopathology*. Wiley, 147(6), pp. 321–330. doi: 10.1046/j.1439-0434.1999.00398.x.
- Spaepen, S. and Vanderleyden, J. (2010) 'Auxin and plant-microbe interactions', *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory, 3(4), pp. a001438–a001438. doi: 10.1101/cshperspect.a001438.
- Spaepen, S., Vanderleyden, J. and Remans, R. (2007) 'Indole-3-acetic acid in microbial and microorganism-plant signaling', *FEMS Microbiology Reviews*. Oxford University Press (OUP), 31(4), pp. 425–448. doi: 10.1111/j.1574-6976.2007.00072.x.
- Structure assesment* (no date). Available at: <https://swissmodel.expasy.org/assess/help>.
- Talapatra, S. K. and Talapatra, B. (2014) 'Shikimic acid pathway', in *Chemistry of plant natural products*. Springer Berlin Heidelberg, pp. 625–678. doi: 10.1007/978-3-642-45410-3\_13.
- Wang, B. *et al.* (2015) 'Tryptophan-independent auxin biosynthesis contributes to early embryogenesis in arabidopsis', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 112(15), pp. 4821–4826. doi: 10.1073/pnas.1503998112.
- Weyens, N. *et al.* (2009) 'Exploiting plant-microbe partnerships to improve biomass production and remediation', *Trends in Biotechnology*. Elsevier BV, 27(10), pp. 591–598. doi: 10.1016/j.tibtech.2009.07.006.
- Worley, J. N. *et al.* (2013) '*Pseudomonas syringae* pv. Tomato DC3000 CmaL (PSPTO4723), a DUF1330 family member, is needed to produce l-allo-isoleucine, a precursor for the phytotoxin coronatine', *Journal of Bacteriology*. American Society for Microbiology, 195(2), pp. 287–296. doi: 10.1128/jb.01352-12.
- Xin, X.-F. *et al.* (2016) 'Bacteria establish an aqueous living space in plants crucial for virulence', *Nature*. Springer Science; Business Media LLC, 539(7630), pp. 524–529. doi: 10.1038/nature20166.
- Xin, X.-F., Kvitko, B. and He, S. Y. (2018) '*Pseudomonas syringae*: What it takes to be a pathogen', *Nature Reviews Microbiology*. Springer Science; Business Media LLC, 16(5), pp. 316–328. doi: 10.1038/nrmicro.2018.17.
- Zedalis, J. and Eggebrecht, J. (2018) *Biology for AP® courses*. Available at: <https://openstax.org/books/biology-ap-courses/pages/3-4-proteins> (Accessed: 9 December 2021).
- Zhang, X. *et al.* (2012) 'Twelve transmembrane helices form the functional core of mammalian MATE1 (multidrug and toxin extruder 1) protein', *Journal of Biological Chemistry*.

Elsevier BV, 287(33), pp. 27971–27982. doi: 10.1074/jbc.m112.386979.

Zhao, Y. (2010) 'Auxin biosynthesis and its role in plant development', *Annual Review of Plant Biology*. Annual Reviews, 61(1), pp. 49–64. doi: 10.1146/annurev-arplant-042809-112308.

Zhao, Y. (2012) 'Auxin biosynthesis: A simple two-step pathway converts tryptophan to indole-3-acetic acid in plants', *Molecular Plant*. Elsevier BV, 5(2), pp. 334–338. doi: 10.1093/mp/ssr104.

Zhao, Y. (2014) 'Auxin biosynthesis', *The Arabidopsis Book*. BioOne, 12, p. e0173. doi: 10.1199/tab.0173.





|                  |                                                               |             |     |
|------------------|---------------------------------------------------------------|-------------|-----|
| Salmonella       |                                                               | -----MTQRDP | 6   |
| Escherichia      | -----MQPRITAITGCQAFDYATFRDSVRISSHVYSTTGDSGSLTQRDP             |             | 44  |
| Arsenophonus     | MLSIYKEEMVWNTVKDKIATLFLQLPAISQWTYYQQFSQSIPVMEKK-----GYLAKNDQ  |             | 53  |
| Erwinia          | -MSAYDVEKEWLRVAGIAAETHQDADLRGFYTQDFRKWPVMDKD-----GFTAQTER     |             | 52  |
| P.savastanoi     | -MEDSHDCLRYSRISSTAAKIHQNDFEGFTYQDFRTHVPIMDKD-----GFAAQTER     |             | 52  |
| P.syringae_group | -MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRTHVPIMDK-----GFAAQTER      |             | 52  |
| P.syringae_NR    | -MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRKHVPIMDK-----GFAAQTER      |             | 52  |
| P.syringae       | -MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRKHVPIMDK-----GFAAQTER      |             | 52  |
| P.avellanae      | -MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQDFRKHVPIMDK-----GFAAQTER      |             | 52  |
| P.savastanoi_NR  | -----THQNDFEGFTYQDFRTHVPIMDK-----GFTAQTER                     |             | 33  |
| P.amygdali       | -MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQNFRRKHVPIMDK-----GFTAQTER     |             | 52  |
| P.savastanoi_NR1 | -MTAYDVEKEWSRISNTAAKTHQNDFEGFTYQNFRRKHVPIMDK-----GFTAQTER     |             | 52  |
|                  | :                                                             | :           |     |
| Salmonella       | A----AEQIPMLLGATSGTTGAMKMISVRLKSSSKDAVSASEHNLHNLSRQVFSAGDVV   |             | 62  |
| Escherichia      | A----AEQIPMLLGATSGTTGAMKMISVRLKSSSKDAVSASEHNLHNLSRQVFSAGDVV   |             | 100 |
| Arsenophonus     | SLHLNLSKNLYIGFTSGTSGTLKRCLYKMSGDVYNVKG-----CEEYNIFCSTDTC      |             | 105 |
| Erwinia          | CVA--ENERNCLIGFTSGTSGNLKRCYYYYDCEVEDSSR----SNVFNSNGFIQPGDR    |             | 106 |
| P.savastanoi     | CLE--RNERNCLIGFTSGTSGNLKRCYYYYDCEVEDSSR----SNVFNSNGFIQPGDR    |             | 106 |
| P.syringae_group | CLE--RNERNCLIGFTSGTSGNLKRCYYYYDCEVEDSSR----SNVFNSNGFIQPGDR    |             | 106 |
| P.syringae_NR    | CLE--RNERNCLIGFTSGTSGNLKRCYYYYDCEVEDSSR----SNVFNSNGFIQPGDR    |             | 106 |
| P.syringae       | CLE--RNERNCLIGFTSGTSGNLKRCYYYYDCEVEDSSR----SNVFNSNGFIQPGDR    |             | 106 |
| P.avellanae      | CLE--RNERNCLIGFTSGTSGNLKRCYYYYDCEVEDSSR----SNVFNSNGFIQPGDR    |             | 106 |
| P.savastanoi_NR  | CLE--RNERNCLIGFTSGTSGNLKRCYYYYDCEVEDSSR----SNVFNSNGFIQPGDR    |             | 87  |
| P.amygdali       | CLE--RNERNCLIGFTSGTSGNLKRCYYYYDCEVEDSSR----SNVFNSNGFIQPGDR    |             | 106 |
| P.savastanoi_NR1 | CLE--RNERNCLIGFTSGTSGNLKRCYYYYDCEVEDSSR----SNVFNSNGFIQPGDR    |             | 106 |
|                  | . : *:* ***** *                                               |             |     |
| Salmonella       | GNLFTINLFSSLHYSAEILKHCRAHVVPVGDIALLRKSHFDLADIQLSTLFGVPATLV    |             | 122 |
| Escherichia      | GNLFTINLFSSLHYSAEILKHCRAHVVPVGDIALLRKSHFDLADIQLSTLFGVPATLV    |             | 160 |
| Arsenophonus     | ANLFTINLLFSAHCLFSNVASACGANIIISGDVFNVLASEHFDALIDVNTNVLIGIPCEII |             | 165 |
| Erwinia          | ANLFTINLFNALNNITMMAGNCGAHVSVSGDITLLTRSHFEALKEIKLNVLGVPSTIL    |             | 166 |
| P.savastanoi     | ANLFTINLFNALNNITMMAGNCGAHVSVSGDITLVTKSHFEALNSIKLNVLGVPSTIL    |             | 166 |
| P.syringae_group | ANLFTINLFNALNNITMMAGNCGAHVSVSGDITLLTKSHFEALNSIKLNVLGVPSTIL    |             | 166 |
| P.syringae_NR    | ANLFTINLFNALNNITMMAGNCGAHVSVSGDITLLTKSHFEALNSIKLNVLGVPSTIL    |             | 166 |
| P.syringae       | ANLFTINLFNALNNITMMAGNCGAHVSVSGDITLLTKSHFEALNSIKLNVLGVPSTIL    |             | 166 |
| P.avellanae      | ANLFTINLFNALNNITMMAGNCGAHVSVSGDITLLTKSHFEALNSIKLNVLGVPSTIL    |             | 166 |
| P.savastanoi_NR  | ANLFTINLFNALNNITMIAGNCGAHVSVSGDITLLTKSHFEALNSIKLNVLGVPSTIL    |             | 147 |
| P.amygdali       | ANLFTINLFNALNNITMMAGNCGAHVSVSGDITLLTKSHFEALNSIKLNVLGVPSTIL    |             | 166 |
| P.savastanoi_NR1 | ANLFTINLFNALNNITMMAGNCGAHVSVSGDITLLTKSHFEALNSIKLNVLGVPSTIL    |             | 166 |
|                  | .*****: : : * *::: **::: .**:* :.: .*:**. ::                  |             |     |
| Salmonella       | QFVETMIGLGVPLVIKKVVFTGEPMLAAHENYLREVLGGDLRIIGLYGLSECGFIGMSVE  |             | 182 |
| Escherichia      | QFVETMIGLGVPLVIKKVVFTGEPMLAAHENYLREVLGGDLRIIGLYGLSECGFIGMSVE  |             | 220 |
| Arsenophonus     | QFIQVMQKKEKLAINKVIFAGEALKAYQREIINKTFGNVDVHIIGAYACSEAGIIAFLSG  |             | 225 |
| Erwinia          | QFIDAMQQHGWHVPVEKIVFNGEGLKTFOKKIIRDAFGEQVSLVGVYGSSEGILGFTAN   |             | 226 |
| P.savastanoi     | QFINAMQHNGVHINIEKVVFTEGESLKTFOKKIIRQAFFGEQVSIVGVYGSSEGILGFTNS |             | 226 |
| P.syringae_group | QFIDAMQQHGWHIDIEKVVFNGEGLKTFOKKIIREAFGEQVSIVGVYGSSEGILGFTNS   |             | 226 |
| P.syringae_NR    | QFIDAMQQHGWHIDIEKVVFNGEGLKTFOKKIIREAFGEQVSIVGVYGSSEGILGFTDS   |             | 226 |
| P.syringae       | QFIDAMQQHGWHIDIEKVVFNGEGLKTFOKKIIREAFGEQVSIVGVYGSSEGILGFTDS   |             | 226 |
| P.avellanae      | QFIDAMQQHGWHIDIEKVVFNGEGLKTFOKKIIREAFGEQVSIVGVYGSSEGILGFTDS   |             | 226 |
| P.savastanoi_NR  | QFIDAMQQHGWHIDIEKVVFNGEGLKTFOKKIIREAFGEQVSIVGVYGSSEGILGFTNS   |             | 207 |
| P.amygdali       | QFIDAMQQHGWHIEIEKVVFNGEGLKTFOKKIIREAFGQVQVSIVGVYGSSEGILGFTNS  |             | 226 |
| P.savastanoi_NR1 | QFIDAMQQHGWHIEIEKVVFNGEGLKTFOKKIIREAFGQVQVSIVGVYGSSEGILGFTNS  |             | 226 |
|                  | **.* * :.*:* ** :.*:* ** :.*:* ** :.*:* ** :                  |             |     |

|                  |                                                              |     |
|------------------|--------------------------------------------------------------|-----|
| Salmonella       | GHHNKYLLFSDDFFFEYEDKKGLLVTSLDPDASRKLVRYPAGDRGKLFVENNQLYFQGIK | 242 |
| Escherichia      | GHHNKYLLFSDDFFFEYEDKKGLLVTSLDPDASRKLVRYPAGDRGKLFVENNQLYFQGIK | 280 |
| Arsenophonus     | EDNNSYELLTNQFFIERD-QDRLLITSLDNMNVIPLLRYNLGDSADISLKGNIIRLTNIK | 284 |
| Erwinia          | ASHTYEFLSNKYFIEKE-GDSILITSLTRENFTPLLRYRLGDTATLSQRGDKIFVSDIQ  | 285 |
| P.savastanoi     | PCHTEYEFLSDKYFIEKE-GDSILITSLTRENFTPLLRYRLGDTATLSMKGDKLYLTDIQ | 285 |
| P.syringae_group | PCHTEYEFLSDKYFIEKE-GDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQ | 285 |
| P.syringae_NR    | PCHTEYEFLSNKYFVEKE-GDSILITSLTGENFTPLLRYRLGDTATLSLKGDKLYLTDIQ | 285 |
| P.syringae       | PCHTEYEFLSNKYFVEKE-GDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQ | 285 |
| P.avellanae      | PCHTEYEFLSNKYFIEKE-GDSIHITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQ | 285 |
| P.savastanoi_NR  | PCHTEYEFLSDKYFIEKE-GDSILITSLTRENFTPLLRYRLGDTATLSMKGDKLYLTDIQ | 266 |
| P.amygdali       | PCHTEYEFLSDKYFIEKE-GDSILITSLTRENFTPLLRYRLGDTATLSMKGDKLYLTDIQ | 285 |
| P.savastanoi_NR1 | PCHTEYEFLSDKYFIEKE-GDSILITSLTRENFTPLLRYRLGDTATLSMKGDKLYLTDIQ | 285 |
|                  | ...* :::::*. * : . : *** ** ** . : . : . *                   |     |

|                  |                                                              |     |
|------------------|--------------------------------------------------------------|-----|
| Salmonella       | RNALDFNFMGNLISHEKIARMINQKLPEAI-VQIVLSLSDERQEIMDIKICCGLSHAVDL | 301 |
| Escherichia      | RNALDFNFMGNLISHEKIARMINQKLPEAI-VQIVLSLSAERQEIMDIKICCGLSHAVDL | 339 |
| Arsenophonus     | RNNISFNFMGYIIEYETIVAVVHKFHPADISIQHLSVAKDTREVITVWVPTGIFSPQOI  | 344 |
| Erwinia          | REDMSFNFMGNLIGLSIIQQAQKQTLDRSLEIQVHLSVTDESKELVTVFVPATEVSEDER | 345 |
| P.savastanoi     | REDMSFNFMGNLIGLGIIQQTIKQTLGRSLEIQVHLSVTEERKELVTVFVQASEVDEDER | 345 |
| P.syringae_group | REDMSFNFMGNLIGLGIIQQAQKQTLGRLEIQVHLSVTDARKELVTVFVQASEVNEDER  | 345 |
| P.syringae_NR    | REDMSFNFMGNLIGLGIIQQAQKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDER | 345 |
| P.syringae       | REDMSFNFMGNLIGLGIIQQAQKQTLGRLEIQVHLSVTDARKELVTVFVQASEVDEDER  | 345 |
| P.avellanae      | REDMSFNFMGNLIGLGIIQQAQKQTLGRSLEIQVHLSVTDARKELVTVFVQASEVDEDER | 345 |
| P.savastanoi_NR  | REDMSFNFMGNLIGLGIIQQAQKQTLGRSLEMQVHLSVTEARKELVTVFVQASEVDEDER | 326 |
| P.amygdali       | REDMSFNFMGNLIGLGIIQQAQKQTLGRPLEIQVHLSVTEARKELVTVFVQASEVDEDER | 345 |
| P.savastanoi_NR1 | REDMSFNFMGNLIGLGIIQQTIKQTLGRPLEIQVHLSVTEARKELVTVFVQASEVDEEER | 345 |
|                  | *: .:***** :* * :::: : :*: **: : :*: : : :                   |     |

|                  |                                                      |     |
|------------------|------------------------------------------------------|-----|
| Salmonella       | SELEDEICHLPDIYEAYMKKRGRINLTVVEYNALLFSPRGKCLLVCDTRES  | 353 |
| Escherichia      | SELEDEICHLPDIYEAYMKKRGRINLTVVEYNALLFSPRGKCLLVCDTRES  | 391 |
| Arsenophonus     | KLLEKELTEIPDIYESKQLNQGYLVVKEVDPIQYIRSVRGKILYIVDNRD-- | 394 |
| Erwinia          | QRIEAAIAAIPDINEAYQKAQGSVAVVAKDARDYAVSERGKMLYIIDRRS-- | 395 |
| P.savastanoi     | VRIETAIADIPDIKEAYQKNQGSVSVLRKDARDYAVSERGKMLYIIDRRN-- | 395 |
| P.syringae_group | ARIETAIADIPDINEAYQKQGSVLVVRKDARDYAVSERGKMLYIIDRRN--  | 395 |
| P.syringae_NR    | TRIETAIADIPDINEAYQKQGSVVRVLRKDARDYAVSERGKMLYIIDRRN-- | 395 |
| P.syringae       | TRIETAIADIPDINEAYQKQGSVVVRKDARDYAVSERGKMLYIIDRRN--   | 395 |
| P.avellanae      | TRIETAIADIPDINEAYQKQGSVLVVRKDARDYAVSERGKMLYIIDRRN--  | 395 |
| P.savastanoi_NR  | ARIETAIADIPDINEAYQKQGSVSVLRKDARDYDVSERGKMLYIIDRRN--  | 376 |
| P.amygdali       | ARIETAIADIPDINEAYQKQGRVSVLRKDARDYAVSERGKMLYIIDRRN--  | 395 |
| P.savastanoi_NR1 | ARIETAIADIPDINEAYQKQGSVSVLRKDARDYAVSERGKMLYIIDRRN--  | 395 |
|                  | :* : :*** *: :* : : : * *** * : * *                  |     |

## Appendix 2 Multiple Sequence Alignment for MatE Proteins

CLUSTAL O(1.2.4) multiple sequence alignment for MatE proteins: 2022-05-06

|                             |                                                              |     |
|-----------------------------|--------------------------------------------------------------|-----|
| Glycine_max                 | MQDIGAPSVSLQLCITSLNKLPLFVWFLKQHNSEFGVSDMAEKESIHSFGDWRRIPICTF | 60  |
| P.savastanoi_MP3            | -----                                                        | 0   |
| P.syringae_group:MT4        | -----                                                        | 0   |
| P.syringae_group:MT2        | -----                                                        | 0   |
| P.savastanoi:MatEefflux     | -----MKSVRREGS-----                                          | 9   |
| P.syringae_gr:genomosp3.MT5 | -----                                                        | 0   |
| P.amygdali_MT               | -----                                                        | 0   |
| P.syringae_group:MT3        | -----                                                        | 0   |
| P.tomatoDC3000:MatEefflux   | -----                                                        | 0   |
| P.tomatoT1:MatEefflux       | -----                                                        | 0   |
| P.syringae_MT1              | -----                                                        | 0   |
| P.syringae_MP2              | -----                                                        | 0   |
| P.syringae_MP1              | -----                                                        | 0   |
|                             |                                                              |     |
| Glycine_max                 | FQDARLVFKADSLGREILSIALPAAMALTADPIASLVDTAFIGQIGPVELAAGVGSIALF | 120 |
| P.savastanoi_MP3            | MAVTKAVFTKGSMLRHVVFMST-----TSATSLLSIFLIEVLTV-----VYIAML      | 45  |
| P.syringae_group:MT4        | MAVTKAVFTKGSMLRHVVFMST-----TSATSLLSIFLIEVLTV-----VYIAML      | 45  |
| P.syringae_group:MT2        | MAVTKAVFTKGSMLRHVVFMST-----TSATSLLSIFLIEVLTV-----VYIAML      | 45  |
| P.savastanoi:MatEefflux     | MAVTKAVFTKGSMLRHVVFMST-----TSATSLLSIFLIEVLTV-----VYIAML      | 54  |
| P.syringae_gr:genomosp3.MT5 | MAVTKAVFTGSLVRHVVFMST-----TSATSLLSIFLIEVLTV-----VYIAML       | 45  |
| P.amygdali_MT               | MAVTKAVFTGSLVRHVVFMST-----TSATSLLSIFLIEVLTV-----VYIAML       | 45  |
| P.syringae_group:MT3        | MAVTKAVFTGSLVRHVVFMST-----TSATSLLSIFLIEVLTV-----VYIAML       | 45  |
| P.tomatoDC3000:MatEefflux   | MAVTKAVFTGSLVRHVVFMST-----TSATSLLSIFLIEVLTV-----VYIAML       | 45  |
| P.tomatoT1:MatEefflux       | MAVTKAVFTGSLVRHVVFMST-----TSATSLLSIFLIEVLTV-----VYIAML       | 45  |
| P.syringae_MT1              | MAVTKAVFTGSLVRHVVFMST-----TSATSLLSIFLIEVLTV-----VYIAML       | 45  |
| P.syringae_MP2              | MAVTKAVFTGSLVRHVVFMST-----TSATSLLSIFLIEVLTV-----VYIAML       | 45  |
| P.syringae_MP1              | MAVTKAVFTGSLVRHVVFMST-----TSATSLLSIFLIEVLTV-----VYIAML       | 45  |
|                             | : : * * . * : * : : : : * : * : * * :                        |     |
|                             |                                                              |     |
| Glycine_max                 | NQVSRIAIFPLVSVTTSFVAEEDTSGANPQTEEGRCLEAGQPTDTETKELLPKQKVTGGN | 180 |
| P.savastanoi_MP3            | RDDSLAFAFAVAKTLMFFLISMILGIAVAVSTTVSRSLGSGVPDQ-----           | 90  |
| P.syringae_group:MT4        | RDDSLAFAFAVAKTLMFFLISMILGIAVAVSTTVSRSLGSGVPDQ-----           | 90  |
| P.syringae_group:MT2        | RDDSLAFAFAVAKTLMFFLISMILGIAVAVSTTVSRSLGSGVPDQ-----           | 90  |
| P.savastanoi:MatEefflux     | RDDSLAFAFAVAKTLMFFLISMILGIAVAVSTTVSRSLGSGVPDQ-----           | 99  |
| P.syringae_gr:genomosp3.MT5 | RDDSLAFAFAVAKTLMFFIISMILGIAVAVSSTVSRALGSGVPDQ-----           | 90  |
| P.amygdali_MT               | RDDSLAFAFAVAKTLMFFIISMILGIAVAVSSTVSRALGSGVPDQ-----           | 90  |
| P.syringae_group:MT3        | RDDSLAFAFAVAKTLMFFIISMILGIAVAVSSTVSRALGSGVPDQ-----           | 90  |
| P.tomatoDC3000:MatEefflux   | RDDSLAFAFAVAKTLMFFIISMILGIAVAVSSTVSRALGSGVPDQ-----           | 90  |
| P.tomatoT1:MatEefflux       | RDDSLAFAFAVAKTLMFFIISMILGIAVAVSSTVSRALGSGVPDQ-----           | 90  |
| P.syringae_MT1              | RDDSLAFAFAVAKTLMFFLISMILGIAVAVSTAVSRSLGSAAPDQ-----           | 90  |
| P.syringae_MP2              | RDDSLAFAFAVAKTLMFFLISMILGIAVAVSTAVSRSLGSAAPDQ-----           | 90  |
| P.syringae_MP1              | RDDSLAFAFAVAKTLMFFLISMILGIAVAVSTAVSRSLGSAAPDQ-----           | 90  |
|                             | . : * : * * : . . . : . * . : * :                            |     |
|                             |                                                              |     |
| Glycine_max                 | VHNSDFVGESFNIAKEERKRHRIPSASSAIFIGGILGLIQAIFLISAAPKLLNFMGVTS  | 240 |
| P.savastanoi_MP3            | -----ARRL---ASSGVLLGSMATVCGAALELVVLDQAIRMLGAE--              | 127 |
| P.syringae_group:MT4        | -----ARRL---ASSGVLLGSMATVCGAALELVVLDQAIRMLGAE--              | 127 |
| P.syringae_group:MT2        | -----ARRL---ASSGVLLGSMATVCGAALELVVLDQAIRMLGAE--              | 127 |
| P.savastanoi:MatEefflux     | -----ARRL---ASSGVLLGSMATVCGAALELVVLDQAIRMLGAE--              | 136 |
| P.syringae_gr:genomosp3.MT5 | -----ARRL---ASSGLLLGIATVCCAVLELIFLDSAIQMLGGE--               | 127 |
| P.amygdali_MT               | -----ARRL---PSSGLLLGIATVCCAVLELIFLDSAIQMLGGE--               | 127 |
| P.syringae_group:MT3        | -----ARRL---ASSGLLLGIATVCCAVLELIFLDSAIQMLGGE--               | 127 |
| P.tomatoDC3000:MatEefflux   | -----ARRL---ASSGLLLGIATVCCAVLELIFLDSAIQMLGGE--               | 127 |
| P.tomatoT1:MatEefflux       | -----ARRL---ASSGLLLGIATVCCAVLELIFLDSAIQMLGGE--               | 127 |
| P.syringae_MT1              | -----ARRL---ASSGLLLGIATVCGAALELVVLDPAIRMLGGE--               | 127 |
| P.syringae_MP2              | -----ARRL---ASSGLLLGIATVCGAALELVVLDPAIRMLGGE--               | 127 |
| P.syringae_MP1              | -----ARRL---ASSGLLLGIATVCGAALELVVLDPAIRMLGGE--               | 127 |
|                             | ** ** : : : * : : . : : *                                    |     |

|                             |                                                               |     |
|-----------------------------|---------------------------------------------------------------|-----|
| Glycine_max                 | SPMLHPAKQYLKRLSGLAPVLLSLAMQGVFRGFKDTKTPLYATVAGDVTNIALDPLFMF   | 300 |
| P.savastanoi_MP3            | GNSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTLTATAAIAIADPLLIL   | 187 |
| P.syringae_group:MT4        | GNSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTLTATAAIAIADPLLIL   | 187 |
| P.syringae_group:MT2        | GNSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTLTATAAIAIADPLLIL   | 187 |
| P.savastanoi:MatEefflux     | GNSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTLTATAAIAIADPLLIL   | 196 |
| P.syringae_gr:genomosp3.MT5 | GSSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTLTATAAIAIADPLLII   | 187 |
| P.amygdali_MT               | GSSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTLTATAAIAIADPLLII   | 187 |
| P.syringae_group:MT3        | GSSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTLTATAAIAIADPLLII   | 187 |
| P.tomatoDC3000:MatEefflux   | GSSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTLTATAAIAIADPLLII   | 187 |
| P.tomatoT1:MatEefflux       | GSSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTLTATAAIAIADPLLII   | 187 |
| P.syringae_MT1              | GNSFTAAREYLWLTLPASVLVAVELLCVQILRSAGFGKQAMWILLTSTAIAIAIADPLLIV | 187 |
| P.syringae_MP2              | GNSFTAAREYLWLTLPASVLVAVELLCVQILRSAGFGKQAMWILLTSTAIAIAIADPLLIV | 187 |
| P.syringae_MP1              | GNSFTAAREYLWLTLPASVLVAVELLCVQILRSAGFGKQAMWILLTSTAIAIAIADPLLIV | 187 |
|                             | . : * : ** * . : * . : : * . * : : . : *** : .                |     |
| Glycine_max                 | VFRLGVSGAAIAHVISQYLISVILLWRLLEQVDLIPPS--INHLQLDR-----FL       | 348 |
| P.savastanoi_MP3            | VFDLGLVGAGIAYLISSLSIACLGFIYVH-RVAHLTCRVSLKNLSGDIRNIGRTALPAVI  | 246 |
| P.syringae_group:MT4        | VFDLGLVGAGIAYLISSLSIACLGFIYVH-RVAHLTCRVSLKNLSGDIRNIGRTALPAVI  | 246 |
| P.syringae_group:MT2        | VFDLGLVGAGIAYLISSLSIACLGFIYVH-RVAHLTCRVSLKNLSGDIRNIGRTALPAVI  | 246 |
| P.savastanoi:MatEefflux     | VFDLGLVGAGIAYLISSLSIACLGFIYVH-RVAHLTCRVSLKNLSGDIRNIGRTALPAVI  | 255 |
| P.syringae_gr:genomosp3.MT5 | VFDLGLAGAGVAYLISSLSIACLGFIYVH-RVARLTCRVSLTHLSGDISIIGRTAVPAVI  | 246 |
| P.amygdali_MT               | VFDLGLAGAGVAYLISSLSIACLGFIYVH-RVARLTCRVSLTHLSGDISIIGRTAVPAVI  | 246 |
| P.syringae_group:MT3        | VFDLGLAGAGVAYLISSLSIACLGFIYVH-RVARLTCRVSLTHLSGDISIIGRTAVPAVI  | 246 |
| P.tomatoDC3000:MatEefflux   | VFDLGLAGAGVAYLISSLSIACLGFIYVH-RVARLTCRVSLTHLSGDISIIGRTAVPAVI  | 246 |
| P.tomatoT1:MatEefflux       | VFDLGLAGAGVAYLISSLSIACLGFIYVH-RVARLTCRVSLTHLSGDISIIGRTAVPAVI  | 246 |
| P.syringae_MT1              | VFDLGLVGAGIAYLISSLSIACLGFIYVH-RVARLTSVSLTHLSGDISIISRTAVPAVI   | 246 |
| P.syringae_MP2              | VFDLGLVGAGIAYLISSLSIACLGFIYVH-RVARLTSVSLTHLSGDISIISRTAVPAVI   | 246 |
| P.syringae_MP1              | VFDLGLVGAGIAYLISSLSIACLGFIYVH-RVARLTSVSLTHLSGDISIISRTAVPAVI   | 246 |
|                             | ** ** : ** : * : ** . : : : : : * : : : * . * . : .           |     |
| Glycine_max                 | -----KNGFLLLMRVIAVTFCVTLAASLAARQGPTSMAAFQV                    | 385 |
| P.savastanoi_MP3            | GNLATPVGMAYVMAAMAPFGSQALATIGVIDRVIQVAFVVFALPGA-----           | 293 |
| P.syringae_group:MT4        | GNLATPVGMAYVMAAMAPFGSQALATIGVIDRVIQVAFVVFALPGA-----           | 293 |
| P.syringae_group:MT2        | GNLATPVGMAYVMAAMAPFGSQALATIGVIDRVIQVAFVVFALPGA-----           | 293 |
| P.savastanoi:MatEefflux     | GNLATPVGMAYVMAAMAPFGSQALATIGVIDRVIQVAFVVFALPGA-----           | 302 |
| P.syringae_gr:genomosp3.MT5 | GNLATPVGMAYVMAAMAPFGAQLAAIGVIDRVIQVVFVVFALPGA-----            | 293 |
| P.amygdali_MT               | GNLATPVGMAYVMAAMAPFGAQLAAIGVIDRVIQVAFVVFALPGA-----            | 293 |
| P.syringae_group:MT3        | GNLATPVGMAYVMAAMAPFGAQLAAIGVIDRVIQVAFVVFALPGA-----            | 293 |
| P.tomatoDC3000:MatEefflux   | GNLATPVGMAYVMAAMAPFGAQLAAIGVIDRVIQVAFVVFALPGA-----            | 293 |
| P.tomatoT1:MatEefflux       | GNLATPVGMAYVMAAMAPFGAQLAAIGVIDRVIQVAFVVFALPGA-----            | 293 |
| P.syringae_MT1              | GNLATPVGMAYVMAAMAPFGSQALAAIGVIDRVIQVAFVVFALPGA-----           | 293 |
| P.syringae_MP2              | GNLATPVGMAYVMAAMAPFGSQALAAIGVIDRVIQVAFVVFALPGA-----           | 293 |
| P.syringae_MP1              | GNLATPVGMAYVMAAMAPFGSQALAAIGVIDRVIQVAFVVFALPGA-----           | 293 |
|                             | : : : *** * . *** : * *                                       |     |
| Glycine_max                 | CLQVWLAVSLLADGLAVAGQAILAGAFANKDFNRATATASRVLMGLVLGLALAFILGTG   | 445 |
| P.savastanoi_MP3            | -----LIPILGQNL--GAM----NTARVSQAIKITYGLLI-----AYG              | 325 |
| P.syringae_group:MT4        | -----LIPILGQNL--GAM----NTARVSQAIKITYGLLI-----AYG              | 325 |
| P.syringae_group:MT2        | -----LIPILGQNL--GAM----NTARVSQAIKITYGLLI-----AYG              | 325 |
| P.savastanoi:MatEefflux     | -----LIPILGQNL--GAM----NTARVSQAIKITYGLLI-----AYG              | 334 |
| P.syringae_gr:genomosp3.MT5 | -----LIPILGQNL--GAM----NTARVSQAVKTTYGLLI-----GYG              | 325 |
| P.amygdali_MT               | -----LIPILGQNL--GAM----NTARVSQAVKTTYGLLI-----GYG              | 325 |
| P.syringae_group:MT3        | -----LIPILGQNL--GAM----NTARVSQAVKTTYGLLI-----GYG              | 325 |
| P.tomatoDC3000:MatEefflux   | -----LIPILGQNL--GAM----NTARVSQAVKTTYGLLI-----GYG              | 325 |
| P.tomatoT1:MatEefflux       | -----LIPILGQNL--GAM----NTARVSQAVKTTYGLLI-----GYG              | 325 |
| P.syringae_MT1              | -----LIPILGQNL--GAM----NTTRVSHAVKATYGLLI-----CYG              | 325 |
| P.syringae_MP2              | -----LIPILGQNL--GAM----NTTRVSHAVKATYGLLI-----CYG              | 325 |
| P.syringae_MP1              | -----LIPILGQNL--GAM----NTTRVSHAVKATYGLLI-----CYG              | 325 |
|                             | : : ** : ** : * : . : : ** : *                                |     |

|                             |                                                                  |     |
|-----------------------------|------------------------------------------------------------------|-----|
| Glycine_max                 | SHFGAKIFTQDANVLHLIQIGIPFIAATQPLNSLAFVFDGVNFGAS-DFAYSAFSMVVVA     | 504 |
| P.savastanoi_MP3            | SV-----TSLLL-----ILLAEPLASLFHVA-G--EGLVVFFAFCRWGGVLT             | 365 |
| P.syringae_group:MT4        | SV-----TSLLL-----ILLAEPLASLFHVA-G--EGLVVFFAFCRWGGVLT             | 365 |
| P.syringae_group:MT2        | SV-----TSLLL-----ILLAEPLASLFHVA-G--EGLVVFFAFCRWGGVLT             | 365 |
| P.savastanoi:MatEefflux     | SV-----TSLLL-----ILLAEPLASLFHVA-G--EGLVVFFAFCRWGGVLT             | 374 |
| P.syringae_gr:genomosp3.MT5 | AT-----TSLLL-----ILLAEPLARLFHVS-G--EGLVVFSAFCRWGGLLWT            | 365 |
| P.amygdali_MT               | AT-----TSLLL-----ILLAEPLARLFHVS-G--EGLVVFSAFCRWGGLLWT            | 365 |
| P.syringae_group:MT3        | AT-----TSLLL-----ILLAEPLARLFHVS-G--EGLVVFSAFCRWGGLLWT            | 365 |
| P.tomatoDC3000:MatEefflux   | AT-----TSLLL-----ILLAEPLARLFHVS-G--EGLVVFSAFCRWGGLLWT            | 365 |
| P.tomatoT1:MatEefflux       | AT-----TSLLL-----ILLAEPLARLFHVS-G--EGLVVFSAFCRWGGLLWT            | 365 |
| P.syringae_MT1              | AA-----TSLLL-----ILLAEPLARLFHVS-G--EGLIVFSAFCRWGGFLWT            | 365 |
| P.syringae_MP2              | AA-----TSLLL-----ILLAEPLARLFHVS-G--EGLIVFSAFCRWGGFLWT            | 365 |
| P.syringae_MP1              | AA-----TSLLL-----ILLAEPLARLFHVS-G--EGLIVFSAFCRWGGFLWT            | 365 |
|                             | :           . * *           * : ** * . * *           * : . : . : |     |
| Glycine_max                 | I--LSIISLLILLSAGFIGIWVALTIYMGLRAFAGFLRIGTGSGPWEFLRSS-----        | 555 |
| P.savastanoi_MP3            | LIGLQFIATSVFLSMGR-----PA---YVTLFGWLRASLGTMPFVWYGAHKFGSVGV        | 414 |
| P.syringae_group:MT4        | LIGLQFIATSVFLSMGR-----PA---YVTLFGWLRASLGTMPFVWYGAHKFGSVGV        | 414 |
| P.syringae_group:MT2        | LIGLQFIATSVFLSMGR-----PA---YVTLFGWLRASLGTMPFVWYGAHKFGSVGV        | 414 |
| P.savastanoi:MatEefflux     | LIGLQFIATSVFLSMGR-----PA---YVTLFGWLRASLGTMPFVWYGAHKFGSVGV        | 423 |
| P.syringae_gr:genomosp3.MT5 | LIGLQFIATSIFLSAGR-----PM---YVTLFGWLRASLGTVPFVWYGAQAFGSVGV        | 414 |
| P.amygdali_MT               | LIGLQFIATSIFLSAGR-----PM---YVTLFGWLRASLGTVPFVWYGAQAFGSVGV        | 414 |
| P.syringae_group:MT3        | LIGLQFIATSIFLSAGR-----PM---YVTLFGWLRASLGTVPFVWYGAQAFGSVGV        | 414 |
| P.tomatoDC3000:MatEefflux   | LIGLQFIATSIFLSAGR-----PM---YVTLFGWLRASLGTVPFVWYGAQAFGSVGV        | 414 |
| P.tomatoT1:MatEefflux       | LIGLQFIATSIFLSAGR-----PM---YVTLFGWLRASLGTVPFVWYGAQAFGSVGV        | 414 |
| P.syringae_MT1              | LIGLQFIATSIFLSMGR-----PM---YVTLFGWLRASLGTVPFVWYGAQAFGSVGV        | 414 |
| P.syringae_MP2              | LIGLQFIATSIFLSMGR-----PM---YVTLFGWLRASLGTVPFVWYGAQAFGSVGV        | 414 |
| P.syringae_MP1              | LIGLQFIATSIFLSMGR-----PM---YVTLFGWLRASLGTVPFVWYGAQAFGSVGV        | 414 |
|                             | : * : * : : ** *                   : : * : * . * : * : :         |     |
| Glycine_max                 | -----                                                            | 555 |
| P.savastanoi_MP3            | MLGQLMGNTIVAFACWLAHLLMKKMLDIEIHSIGNRSLHRGNS                      | 458 |
| P.syringae_group:MT4        | MLGQLMGNTIVAFACWLAHLLMKKMLDIEIHSIGNRSLHRGNS                      | 458 |
| P.syringae_group:MT2        | MLGQLMGNTIVAFACWLAHLLMKKMLDIEIHSIGNRSLHRGNS                      | 458 |
| P.savastanoi:MatEefflux     | MLGQLMGNTIVAFACWLAHLLMKKMLDIEIHSIGNRSLHRGNS                      | 467 |
| P.syringae_gr:genomosp3.MT5 | MLGQLLGNTIVAFACWVAHLLMKKRIGNKLDLTGNRSRHRGNS                      | 458 |
| P.amygdali_MT               | MLGQLLGNTIVAFACWVAHLLMKKRIGNKLDLTGNRSRHRGNS                      | 458 |
| P.syringae_group:MT3        | MLGQLLGNTIVAFACWVAHLLMKKRIGNKLDLTGNRSRHRGNS                      | 458 |
| P.tomatoDC3000:MatEefflux   | MLGQLLGNTIVAFACWVAHLLMKKRIGNKLDLTGNRSRHRGNS                      | 458 |
| P.tomatoT1:MatEefflux       | MLGQLLGNTIVAFACWVAHLLMKKRIGNKLDLTGNRSRHRGNS                      | 458 |
| P.syringae_MT1              | MLGQLLGNTVVAFCACWLAHLLMKKILGNEAII-----                           | 447 |
| P.syringae_MP2              | MLGQLLGNTVVAFCACWLAHLLMKKVLGNEAII-----                           | 447 |
| P.syringae_MP1              | MLGQLLGNTVVAFCACWLAHLLMKKVLGNKAIH-----                           | 447 |



### Appendix 3 T-coffee Multiple Sequence Alignment for IaaL Proteins

T-COFFEE, Version\_11.00 (Version\_11.00)

Cedric Notredame

SCORE=97

\*

BAD AVG GOOD

\*

|                 |   |    |
|-----------------|---|----|
| WP_195918284.1  | : | 95 |
| WP_157940369.1  | : | 93 |
| WP_051296966.1  | : | 95 |
| WP_034939628.1  | : | 98 |
| WP_057442603.1  | : | 98 |
| WP_057442603.1_ | : | 98 |
| WP_005735201.1  | : | 98 |
| WP_004662044.1  | : | 98 |
| WP_005613862.1  | : | 98 |
| WP_019331605.1  | : | 98 |
| WP_087674970.1  | : | 98 |
| WP_007247550.1  | : | 98 |
| WP_082301237.1  | : | 98 |
| cons            | : | 9  |

```

WP_195918284.1    MT-----QRPDAA-
WP_157940369.1    MQ-P-----RITAITGCAFDYTAFRDSVRISSHVYSTTGDSSGLTQRDPAA-
WP_051296966.1    MLSIYKEEMVWNTVKDKIATLFLQPAISQWTYQQFSQSIPVMEKKG---Y-LAKNDQSLHL
WP_034939628.1    MS-AYDVEKEWLRVAGIAAETHQDADLRGFTYQDFRKWVPVMDKDG---F-TAQTERCVAE-
WP_057442603.1    MT-AYDVEKEWSRISNTAAKTHQNNDFEGFTYQNF RKHVPIMDKG----F-TAQTERCLER-
WP_057442603.1    MT-AYDVEKEWSRISNTAAKTHQNNDFEGFTYQNF RKHVPIMDKG----F-TAQTERCLER-
WP_005735201.1    MT-AYDVEKEWSRISNTAAKTHQNNDFEGFTYQNF RKHVPIMDKG----F-TAQTERCLER-
WP_004662044.1    -----THQNNDFEGFTYQDFRTHVPI MDKDG---F-TAQTERCLER-
WP_005613862.1    MT-AYDVEKEWSRISNTAAKTHQNNDFEGFTYQDFRKHVPIMDKG---F-AAQTERCLER-
WP_019331605.1    MT-AYDVEKEWSRISNTAAKTHQNNDFEGFTYQDFRKHVPIMDKG---F-AAQTERCLER-
WP_087674970.1    MT-AYDVEKEWSRISNTAAKTHQNNDFEGFTYQDFRKHVPIMDKG---F-AAQTERCLER-
WP_007247550.1    MT-AYDVEKEWSRISNTAAKTHQNNDFEGFTYQDFRTHVPI MDKDG----F-AAQTERCLER-
WP_082301237.1    ME-DSHDC LYSRISSTA AKIHNNDFEGFTYQDFRTHVPI MDKDG----F-AAQTERCLER-

```

cons

|                 |                                                                 |
|-----------------|-----------------------------------------------------------------|
| WP_195918284.1  | EQ-IPMLLGATSGTTGAMKMISVRLKSSKDA--VSASEHNLIHNLRSRQVFSAGDVGVLNFTI |
| WP_157940369.1  | EQ-IPMLLGATSGTTGAMKMISVRLKSSKDA--VSASEHNLIHNLRSRQVFSAGDVGVLNFTI |
| WP_051296966.1  | NSKNSYLIGFTSGTSGTLKRCLYKMSGDVYNVKGCE-EY---NI-----FCSTDTCANLFTI  |
| WP_034939628.1  | NE-RNCLIGFTSGTSGNLKRCYYYYDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI |
| WP_057442603.1  | NE-RNCLIGFTSGTSGNLKRCYYYYDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI |
| WP_057442603.1_ | NE-RNCLIGFTSGTSGNLKRCYYYYDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI |
| WP_005735201.1  | NE-RNCLIGFTSGTSGNLKRCYYYYDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI |
| WP_004662044.1  | NE-RNCLIGFTSGTSGNLKRCYYYYDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI |
| WP_005613862.1  | NE-RNCLIGFTSGTSGNLKRCYYYYDCEVDE--DSS-RS---SVFRSNGFIQPGDRCANLFTI |
| WP_019331605.1  | NE-RNCLIGFTSGTSGNLKRCYYYYDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI |
| WP_087674970.1  | NE-RNCLIGFTSGTSGNLKRCYYYYDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI |
| WP_007247550.1  | NE-RNCLIGFTSGTSGNLKRCYYYYDCEVDE--DSS-RS---NVFRSNGFIQPGDRCANLFTI |
| WP_082301237.1  | NE-RNCLIGFTSGTSGNLKRCYYYYDCEVDE--DSS-LS---NVFRSNGFIQPGDRCANLFTI |

```
cons      :.      *: * *****: * :*
```

|                 |                                                                  |
|-----------------|------------------------------------------------------------------|
| WP_195918284.1  | NLFSSLHYSAACEILKHCRAHVVPVGDIALLRKSHDFDLADIQLSTLFGVPATLVQFVETMIGL |
| WP_157940369.1  | NLFSSLHYSAACEILKHCRAHVVPVGDIALLRKSHDFDLADIQLSTLFGVPATLVQFVETMIGL |
| WP_051296966.1  | NLLFSAHCLFSNVASACGANIISIGDFNVLASEHFDALIDVNTNVLIGIPCEIIQFIQVMQKK  |
| WP_034939628.1  | NLFSALNNITMMAGNCGAHVVSVDITLLTRSHFEALKEIKLNVLLGVPSTILQFIDAMQQH    |
| WP_057442603.1  | NLFSALNNITMMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQQH    |
| WP_057442603.1_ | NLFSALNNITMMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQQH    |
| WP_005735201.1  | NLFSALNNITMMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQQH    |
| WP_004662044.1  | NLFSALNNITMMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQQH    |
| WP_005613862.1  | NLFSALNNITMMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQQH    |
| WP_019331605.1  | NLFSALNNITMMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQQH    |
| WP_087674970.1  | NLFSALNNITMMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQQH    |
| WP_007247550.1  | NLFSALNNITMMAGNCGAHVVSVDITLLTKSHFEALNSIKLNVLLGVPSTILQFIDAMQQH    |
| WP_082301237.1  | NLFSALNNITMMAGNCGAHVVSVDITLVTKSHFEALNSIKLNVLLGVPSTILQFINAMQHN    |

|      |                                                                |
|------|----------------------------------------------------------------|
| cons | ** : : : * * : : : ** : : . ** : * . : : . * : * . : : * : * . |
|------|----------------------------------------------------------------|

|                 |                                                                  |
|-----------------|------------------------------------------------------------------|
| WP_195918284.1  | GVPLVIKKVFTGEPMLAAHENYLREVLGDDLRIIGLYGLSECGFIGMSVEGHHNKYLLFSDD   |
| WP_157940369.1  | GVPLVIKKVFTGEPMLAAHENYLREVLGDDLRIIGLYGLSECGFIGMSVEGHHNKYLLFSDD   |
| WP_051296966.1  | EKLALINKVIFAGEALKAYQREIIKNTFGNDVHIIGAYACSEAGIIAFSLGEDNNSYELLTNQ  |
| WP_034939628.1  | GVHVPVEKIVFNAGEGLKTFQKKIIRDAFGEQVSLVGVGYSSEGGILGFTANASHTYEYFLSNK |
| WP_057442603.1  | GVHIEIEKVVFNAGEGLKTFQKKIIREAFGEQVSIQVVGYSSEGGILGFTNSPCHTEYEFSLDK |
| WP_057442603.1_ | GVHIEIEKVVFNAGEGLKTFQKKIIREAFGEQVSIQVVGYSSEGGILGFTNSPCHTEYEFSLDK |
| WP_005735201.1  | GVHIEIEKVVFNAGEGLKTFQKKIIREAFGEQVSIQVVGYSSEGGILGFTNSPCHTEYEFSLDK |
| WP_004662044.1  | GVHIDIEKVVFNAGEGLKTFQKKIIREAFGEQVSIQVVGYSSEGGILGFTNSPCHTEYEFSLDK |
| WP_005613862.1  | GVHIDIEKVVFNAGEGLKTFQKKIIREAFGEQVSIQVVGYSSEGGILGFTDSPCHTEYEFSLNK |
| WP_019331605.1  | GVHIDIEKVVFNAGEGLKTFQKKIIREAFGEQVSIQVVGYSSEGGILGFTDSPCHTEYEFSLNK |
| WP_087674970.1  | GVHIDIEKVVFNAGEGLKTFQKKIIREAFGEQVSIQVVGYSSEGGILGFTDSPCHTEYEFSLNK |
| WP_007247550.1  | GVHIDIEKVVFNAGEGLKTFQKKIIREAFGEQVSIQVVGYSSEGGILGFTNSPCHTEYEFSLDK |
| WP_082301237.1  | GVHINIEKVVFTGESLKTQKKIIRQAFGEQVSIQVVGYSSEGGILGFTNSPCHTEYEFSLDK   |

|      |                                                                 |
|------|-----------------------------------------------------------------|
| cons | : : * : * ** : : : : : : * : : : * * . ** * : : : : : * : : : . |
|------|-----------------------------------------------------------------|

|                 |                                                                  |
|-----------------|------------------------------------------------------------------|
| WP_195918284.1  | FFFEYEDKKGLLVTSLDAPDASRKLVRYPAGDRGKLFVENNQLYFQGIKRNALDFNFMGNLISH |
| WP_157940369.1  | FFFEYEDKKGLLVTSLDAPDASRKLVRYPAGDRGKLFVENNQLYFQGIKRNALDFNFMGNLISH |
| WP_051296966.1  | FFIERDQ-DRLITSLDNMNVIPLLRYNLGDSADISLKGNIRLTNIKRNNISFNFMGYIEY     |
| WP_034939628.1  | YFIEKEG-DSILITSLTRENFTPLLRYRLGDTATLSQRGDKIFVSDIQREDMSFNFMGNLIGL  |
| WP_057442603.1  | YFIEKEG-DSILITSLTRENFTPLLRYRLGDTATLSMKGDKLYLTDIQREDMSFNFMGNLIGL  |
| WP_057442603.1_ | YFIEKEG-DSILITSLTRENFTPLLRYRLGDTATLSMKGDKLYLTDIQREDMSFNFMGNLIGL  |
| WP_005735201.1  | YFIEKEG-DSILITSLTRENFTPLLRYRLGDTATLSMKGDKLYLTDIQREDMSFNFMGNLIGL  |
| WP_004662044.1  | YFIEKEG-DSILITSLTRENFTPLLRYRLGDTATLSMKGDKLYLTDIQREDMSFNFMGNLIGL  |
| WP_005613862.1  | YFIEKEG-DSIHITSLTRENFTPLLRYRLGDTATLSLKGDLYLTDIQREDMSFNFMGNLIGL   |
| WP_019331605.1  | YFVEKEG-DSILITSLTRENFTPLLRYRLGDTATLSLKGDLYLTDIQREDMSFNFMGNLIGL   |
| WP_087674970.1  | YFVEKEG-DSILITSLTGENFTPLLRYRLGDTATLSLKGDLYLTDIQREDMSFNFMGNLIGL   |
| WP_007247550.1  | YFIEKEG-DSILITSLTRENFTPLLRYRLGDTATLSLKGDLYLTDIQREDMSFNFMGNLIGL   |
| WP_082301237.1  | YFIEKEG-DSILITSLTRENFTPLLRYRLGDTATLSMKGDKLYLTDIQREDMSFNFMGNLIGL  |

|      |                                                               |
|------|---------------------------------------------------------------|
| cons | : * . * : . : : *** * : ** ** . : . : : . * : * : : . ***** : |
|------|---------------------------------------------------------------|

|                 |                                                                  |
|-----------------|------------------------------------------------------------------|
| WP_195918284.1  | EKIARMINQKLPEAI-VQIVLSLSDERQEIMDIKICCGLSHAVDLSELEDEICHLDPDIYEAYM |
| WP_157940369.1  | EKIARMINQKLPEAI-VQIVLSLSAERQEIMDIKICCGLSHAVDLSELEDEICHLDPDIYEAYM |
| WP_051296966.1  | ETIVAVVHKFHPADISIQIHLVAKDTREVITVWVPTGIFSPQKIKLEKELTEIPDIYESKQ    |
| WP_034939628.1  | SIIQQAQKQTLDRSLEIQVHLSVTDESKELVTVFVQASEVDEEERARIETAIADIPDINEAYQ  |
| WP_057442603.1  | GIIQQTIKQTLGRPLEIQVHLSVTEARKELVTVFVQASEVDEEERARIETAIADIPDINEAYQ  |
| WP_057442603.1_ | GIIQQTIKQTLGRPLEIQVHLSVTEARKELVTVFVQASEVDEEERARIETAIADIPDINEAYQ  |
| WP_005735201.1  | GIIQQAQKQTLGRPLEIQVHLSVTEARKELVTVFVQASEVDEDERARIETAIADIPDINEAYQ  |
| WP_004662044.1  | GIIQQAQKQTLGRSLEMQVHLSVTEARKELVTVFVQASEVDEDERARIETAIADIPDINEAYQ  |
| WP_005613862.1  | GIIQQAQKQTLGRSLEIQVHLSVTDARKELVTVFVQASEVDEDERTRIETAIADIPDINEAYQ  |
| WP_019331605.1  | GIIQQAQKQTLGRSLEIQVHLSVTDARKELVTVFVQASEVDEDERTRIETAIADIPDINEAYQ  |
| WP_087674970.1  | GIIQQAQKQTLGRSLEIQVHLSVTDARKELVTVFVQASEVDEDERTRIETAIADIPDINEAYQ  |
| WP_007247550.1  | GIIQQAQKQTLGRSLEIQVHLSVTDARKELVTVFVQASEVDEDERARIETAIADIPDINEAYQ  |
| WP_082301237.1  | GIIQQTIKQTLGRSLEIQVHLSVTEERKELVTVFVQASEVDEDERVRIETAIADIPDIKEAYQ  |

|      |                                                     |
|------|-----------------------------------------------------|
| cons | * : : : : * : * : : : * : : : : : : : * : : : * * : |
|------|-----------------------------------------------------|

|                 |                                    |
|-----------------|------------------------------------|
| WP_195918284.1  | KKRGRINLTVVEYNALLFSPRGKCLLVCDTRES  |
| WP_157940369.1  | KKRGRINLTVVEYNALLFSPRGKCLLVCDTRES  |
| WP_051296966.1  | LNQGYLVVKEVDPIQYIRSVRGKILYIVDNRD-- |
| WP_034939628.1  | KAQGSVAVVAKDARDYAVSERGKMLYIIDRRS-- |
| WP_057442603.1  | KDQGSVSVLRKDARDYAVSERGKMLYIIDRRN-- |
| WP_057442603.1_ | KDQGSVSVLRKDARDYAVSERGKMLYIIDRRN-- |
| WP_005735201.1  | KDQGRVSVLRKDARDYAVSERGKMLYIIDRRN-- |
| WP_004662044.1  | KDQGSVSVLRKDARDYDVSERGKMLYIIDRRN-- |
| WP_005613862.1  | KDQGSVLVVRKDARDYAVSERGKMLYIIDRRN-- |
| WP_019331605.1  | KDQGSVVVVRKDARDYAVSERGKMLYIIDRRN-- |
| WP_087674970.1  | KDQGSVRVLRKDARDYAVSERGKMLYIIDRRN-- |
| WP_007247550.1  | KDQGSVLVVRKDARDYAVSERGKMLYIIDRRN-- |
| WP_082301237.1  | KNQGSVSVLRKDARDYAVSERGKMLYIIDRRN-- |

|      |                           |
|------|---------------------------|
| cons | : * : : : * * * * : * * . |
|------|---------------------------|





## Appendix 4 T-coffee MSA for MatE Proteins

T-COFFEE, Version\_11.00 (Version\_11.00)

Cedric Notredame

CPU TIME:0 sec.

SCORE=876

\*

BAD AVG GOOD

\*

P.syringae\_MT1 : 90  
 P.syringae\_MP : 90  
 P.syringae\_MP2 : 90  
 P.syringae\_grou : 90  
 P.syringae\_gr\_g : 90  
 P.amygdali\_MT4 : 90  
 P.syringae\_grou : 90  
 P.syringae\_grou : 90  
 P.savastanoi\_MP : 90  
 Glycine\_max : 61  
 cons : 87

```

P.syringae_MT1 MAV-----TKAVFTEGSVLRHV-----
P.syringae_MP MAV-----TKAVFIEGSVLRHV-----
P.syringae_MP2 MAV-----TKAVFTEGSVLRHV-----
P.syringae_grou MAV-----TKAVFTEGSVLRHV-----
P.syringae_gr_g MAV-----TKAVFTEGSVLRHV-----
P.amygdali_MT4 MAV-----TKAVFTEGSVLRHV-----
P.syringae_grou MAV-----TKAVFTKGSMLRHV-----
P.syringae_grou MAV-----TKAVFTKGSMLRHV-----
P.savastanoi_MP MAV-----TKAVFTKGSMLRHV-----
Glycine_max MQDIGAPSVSLQLCITSLNKLPLFVWFLKQHNSEFGVSDMAEKESIHSFGDWRRIPICTFFQD
  
```

```

cons * ..... *. *
  
```

```

P.syringae_MT1 -----VFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSLAFAVAKT-----
P.syringae_MP -----VFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSLAFAVAKT-----
P.syringae_MP2 -----VFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSLAFAVAKT-----
P.syringae_grou -----VFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSLAFAVAKT-----
P.syringae_gr_g -----VFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSLAFAVAKT-----
P.amygdali_MT4 -----VFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSLAFAVAKT-----
P.syringae_grou -----VFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSLAFAVAKT-----
P.syringae_grou -----VFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSLAFAVAKT-----
P.savastanoi_MP -----VFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSLAFAVAKT-----
Glycine_max ARLVFKADSLGREILSIALPAAMALTADPIASLVDTAFIGQIGPVE-LAAVGVSIALFNQVSR
  
```

```

cons : : . : * : * : : : : . : * : : . * : * : :
  
```

```

P.syringae_MT1 -LMFFLISMILGI-----AVAVSTAVS
P.syringae_MP -LMFFLISMILGI-----AVAVSTAVS
P.syringae_MP2 -LMFFLISMILGI-----AVAVSTAVS
P.syringae_grou -LMFFIISMILGI-----AVAVSSTVS
P.syringae_gr_g -LMFFIISMILGI-----AVAVSSTVS
P.amygdali_MT4 -LMFFIISMILGI-----AVAVSSTVS
P.syringae_grou -LMFFLISMILGI-----AVAVSTTVS
P.syringae_grou -LMFFLISMILGI-----AVAVSTTVS
P.savastanoi_MP -LMFFLISMILGI-----AVAVSTTVS
Glycine_max IAIFPLVSVTTSFVAEEDTLSGANPQTEEGRCLEAGQPTDTETKELLPQKVTGGNVHNSDFVG
  
```

```

cons : * : : * : : : : : : : * * * ,
  
```

|                 |                                                                |
|-----------------|----------------------------------------------------------------|
| P.syringae_MT1  | RSLGSAAPDQAR---RLASSGLLLGIATVCGAALELVLDPAIRMLGGEGNS--FTAAREYL  |
| P.syringae_MP   | RSLGSAAPDQAR---RLASSGLLLGIATVCGAALELVLDPAIRMLGGEGNS--FTAAREYL  |
| P.syringae_MP2  | RSLGSAAPDQAR---RLASSGLLLGIATVCGAALELVLDPAIRMLGGEGNS--FTAAREYL  |
| P.syringae_grou | RALGSGVPDQAR---RLASSGLLLGIATVCCAVLELIFLDSAIQMLGGEGSS--FTAAREYL |
| P.syringae_gr_g | RALGSGVPDQAR---RLASSGLLLGIATVCCAVLELIFLDSAIQMLGGEGSS--FTAAREYL |
| P.amygdali_MT4  | RALGSGVPDQAR---RLPSSGLLLGIATVCCAVLELIFLDSAIQMLGGEGSS--FTAAREYL |
| P.syringae_grou | RSLGSGVPDQAR---RLASSGVLLGSMATVCGAALELVLDQAIRMLGAEGNS--FTAAREYL |
| P.syringae_grou | RSLGSGVPDQAR---RLASSGVLLGSMATVCGAALELVLDQAIRMLGAEGNS--FTAAREYL |
| P.savastanoi_MP | RSLGSGVPDQAR---RLASSGVLLGSMATVCGAALELVLDQAIRMLGAEGNS--FTAAREYL |
| Glycine_max     | ESFNIAKEERKRHHIPASSAIFIGGILGLIQAIFLISAAKPLNFMGVTSDSPMLHPAKQYL  |

|      |                                                  |
|------|--------------------------------------------------|
| cons | .:. . : : * .**.:.* : : * : : . :.:* ..* : .*:.* |
|------|--------------------------------------------------|

|                 |                                                                |
|-----------------|----------------------------------------------------------------|
| P.syringae_MT1  | WLTLPASVLVAVELLCVQILRSAGFGKQAMWILLTSTAATAIADPLLIVFDLGLVGAGIAYL |
| P.syringae_MP   | WLTLPASVLVAVELLCVQILRSAGFGKQAMWILLTSTAATAIADPLLIVFDLGLVGAGIAYL |
| P.syringae_MP2  | WLTLPASVLVAVELLCVQILRSAGFGKQAMWILLTSTAATAIADPLLIVFDLGLVGAGIAYL |
| P.syringae_grou | WLTLPASVLVAVEQLCVQILRSAGFGKQAMWILLTATAATAIADPLLIVFDLGLAGAGVAYL |
| P.syringae_gr_g | WLTLPASVLVAVEQLCVQILRSAGFGKQAMWILLTATAATAIADPLLIVFDLGLAGAGVAYL |
| P.amygdali_MT4  | WLTLPASVLVAVEQLCVQILRSAGFGKQAMWILLTATAATAIADPLLIVFDLGLAGAGVAYL |
| P.syringae_grou | WLTLPASVLVAVEQLCVQILRSAGFGKQAMWILLTATAATAIADPLLIVFDLGLVGAGIAYL |
| P.syringae_grou | WLTLPASVLVAVEQLCVQILRSAGFGKQAMWILLTATAATAIADPLLIVFDLGLVGAGIAYL |
| P.savastanoi_MP | WLTLPASVLVAVEQLCVQILRSAGFGKQAMWILLTATAATAIADPLLIVFDLGLVGAGIAYL |
| Glycine_max     | KLRLGAPAVLLSLAMQGVFRGFKDTKTPLATVAGDVTNIALDPLFMFVRLGVSGAAIAHV   |

|      |                                          |
|------|------------------------------------------|
| cons | * .: * . :.* * .: :. . : **.:** *: **.:* |
|------|------------------------------------------|

|                 |                                                                |
|-----------------|----------------------------------------------------------------|
| P.syringae_MT1  | ISSLVSAC-LGFYVHRVARLTSGVSLTHLSGDISTISRTAVPAVIGNLATPVGMAVYMAAMA |
| P.syringae_MP   | ISSLVSAC-LGFYVHRVARLTSGVSLTHLSGDISTISRTAVPAVIGNLATPVGMAVYMAAMA |
| P.syringae_MP2  | ISSLVSAC-LGFYVHRVARLTSGVSLTHLSGDISTISRTAVPAVIGNLATPVGMAVYMAAMA |
| P.syringae_grou | ISSLVSAC-LGFYVHRVARLTSGVSLTHLSGDISIIGRTAVPAVIGNLATPVGMAVYMAAMA |
| P.syringae_gr_g | ISSLVSAC-LGFYVHRVARLTSGVSLTHLSGDISIIGRTAVPAVIGNLATPVGMAVYMAAMA |
| P.amygdali_MT4  | ISSLVSAC-LGFYVHRVARLTSGVSLTHLSGDISIIGRTAVPAVIGNLATPVGMAVYMAAMA |
| P.syringae_grou | ISSLVSAC-LGFYVHRVAHLTCRVSLKNLSGDIRNIGRTALPAVIGNLATPVGMAVYMAAMA |
| P.syringae_grou | ISSLVSAC-LGFYVHRVAHLTCRVSLKNLSGDIRNIGRTALPAVIGNLATPVGMAVYMAAMA |
| P.savastanoi_MP | ISSLVSAC-LGFYVHRVAHLTCRVSLKNLSGDIRNIGRTALPAVIGNLATPVGMAVYMAAMA |
| Glycine_max     | ISQYLISVILLWRLLLEQVDLIPPS--INHLQ-----                          |

|      |                          |
|------|--------------------------|
| cons | ** . : : * : :.* : . :.* |
|------|--------------------------|

|                 |                                                               |
|-----------------|---------------------------------------------------------------|
| P.syringae_MT1  | PFGSQALAAIGVIDRVIQVAFVCFVFP-----GALI                          |
| P.syringae_MP   | PFGSQALAAIGVIDRVIQVAFVCFVFP-----GALI                          |
| P.syringae_MP2  | PFGSQALAAIGVIDRVIQVAFVCFVFP-----GALI                          |
| P.syringae_grou | PFGAQALAAIGVIDRVIQVAFVCFVFP-----GALI                          |
| P.syringae_gr_g | PFGAQALAAIGVIDRVIQVAFVCFVFP-----GALI                          |
| P.amygdali_MT4  | PFGAQALAAIGVIDRVIQVAFVCFVFP-----GALI                          |
| P.syringae_grou | PFGSQALATIGVIDRVIQVAFVCFVFP-----GALI                          |
| P.syringae_grou | PFGSQALATIGVIDRVIQVAFVCFVFP-----GALI                          |
| P.savastanoi_MP | PFGSQALATIGVIDRVIQVAFVCFVFP-----GALI                          |
| Glycine_max     | -LDRFLKNGFLLMRVIAVTFCTLAASLAARQGPTSMAAFQVCLQVWLAVSLLADGLAVAGQ |

|      |                     |
|------|---------------------|
| cons | :. : : ** *.*.* : * |
|------|---------------------|

|                 |                                                                 |
|-----------------|-----------------------------------------------------------------|
| P.syringae_MT1  | PILGQNLGAMNTTRVSHAVKATYGLLICYGAAATSLLLILLAELARLFHVSGEGLIVFSAFCR |
| P.syringae_MP   | PILGQNLGAMNTTRVSHAVKATYGLLICYGAAATSLLLILLAELARLFHVSGEGLIVFSAFCR |
| P.syringae_MP2  | PILGQNLGAMNTTRVSHAVKATYGLLICYGAAATSLLLILLAELARLFHVSGEGLIVFSAFCR |
| P.syringae_grou | PILGQNLGAMNTARVSQAVKTTYGLLIGYGATTSLLLILLAELARLFHVSGEGLVVFSAFCR  |
| P.syringae_gr_g | PILGQNLGAMNTARVSQAVKTTYGLLIGYGATTSLLLILLAELARLFHVSGEGLVVFSAFCR  |
| P.amygdali_MT4  | PILGQNLGAMNTARVSQAVKTTYGLLIGYGATTSLLLILLAELARLFHVSGEGLVVFSAFCR  |
| P.syringae_grou | PILGQNLGAMNTARVSQAIKITYGLLIAYGSVTSLLLILLAELASLFHVAGEGLVVFSAFCR  |
| P.syringae_grou | PILGQNLGAMNTARVSQAIKITYGLLIAYGSVTSLLLILLAELASLFHVAGEGLVVFSAFCR  |
| P.savastanoi_MP | PILGQNLGAMNTARVSQAIKITYGLLIAYGSVTSLLLILLAELASLFHVAGEGLVVFSAFCR  |
| Glycine_max     | AILAGAFANKDFNRATATASRVLQMGLVLGLALAFIL-----GTGSH                 |

|      |                               |
|------|-------------------------------|
| cons | .** . : * : : . : : * . : : * |
|------|-------------------------------|

|                 |                                                                  |
|-----------------|------------------------------------------------------------------|
| P.syringae_MT1  | WGGFLWTL--IGLQFIATSI-FLSMGRPMYVTLFGWLRASLGTVPFVWYGAQAFGSVGVM LGQ |
| P.syringae_MP   | WGGFLWTL--IGLQFIATSI-FLSMGRPMYVTLFGWLRASLGTVPFVWYGAQAFGSVGVM LGQ |
| P.syringae_MP2  | WGGFLWTL--IGLQFIATSI-FLSMGRPMYVTLFGWLRASLGTVPFVWYGAQAFGSVGVM LGQ |
| P.syringae_grou | WGGLLWTL--IGLQFIATSI-FLSAGRPMYVTLFGWLRASLGTVPFVWYGAQAFGSVGVM LGQ |
| P.syringae_gr_g | WGGLLWTL--IGLQFIATSI-FLSAGRPMYVTLFGWLRASLGTVPFVWYGAQAFGSVGVM LGQ |
| P.amygdali_MT4  | WGGLLWTL--IGLQFIATSI-FLSAGRPMYVTLFGWLRASLGTVPFVWYGAQAFGSVGVM LGQ |
| P.syringae_grou | WGGVLWTL--IGLQFIATSV-FLSMGRPAYVTLFGWLRASLGTMPFVWYGAHKFGSVGVM LGQ |
| P.syringae_grou | WGGVLWTL--IGLQFIATSV-FLSMGRPAYVTLFGWLRASLGTMPFVWYGAHKFGSVGVM LGQ |
| P.savastanoi_MP | WGGVLWTL--IGLQFIATSV-FLSMGRPAYVTLFGWLRASLGTMPFVWYGAHKFGSVGVM LGQ |
| Glycine_max     | FGAKIFTQDANVLHLIQIGIPFIAATQPLNSLAFVFDGVNFGASDFAYSAFSMVVVAILSIIS  |

|      |                                                |
|------|------------------------------------------------|
| cons | :* . : : * : : * : : * : : * : : * : : * : : * |
|------|------------------------------------------------|

|                 |                                                |
|-----------------|------------------------------------------------|
| P.syringae_MT1  | -LLGNTVVAFCACWLAHLL-M--KKILGNE-----AIH----     |
| P.syringae_MP   | -LLGNTVVAFCACWLAHLL-M--KKVLGNK-----AIH----     |
| P.syringae_MP2  | -LLGNTVVAFCACWLAHLL-M--KKVLGNE-----AIH----     |
| P.syringae_grou | -LLGNTIVAFACWVAHLL-M--KKRIGNKLDL--TGNRSRHRGNS  |
| P.syringae_gr_g | -LLGNTIVAFACWVAHLL-M--KKRIGNKLDL--TGNRSRHRGNS  |
| P.amygdali_MT4  | -LLGNTIVAFACWVAHLL-M--KKRIGNKLDL--TGNRSRHRGNS  |
| P.syringae_grou | -LMGNTIVAFACWLAHLL-M--KKMLDIEIHS--IGNRSLHRGNS  |
| P.syringae_grou | -LMGNTIVAFACWLAHLL-M--KKMLDIEIHS--IGNRSLHRGNS  |
| P.savastanoi_MP | -LMGNTIVAFACWLAHLL-M--KKMLDIEIHS--IGNRSLHRGNS  |
| Glycine_max     | LLILLSAGGFIGIWVVALTIYMGLRAFAFGLRIGTGSGPWFLR-SS |

|      |                                 |
|------|---------------------------------|
| cons | * : : . * . : * : * : . . . . . |
|------|---------------------------------|

## Appendix 5 COBALT Analysis for IaaL Proteins

```

1c1|Query_10001      VEKEGDSILITSLTGENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL
1c1|Query_10002      IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL
1c1|Query_10003      IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL
1c1|Query_10004      IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL
1c1|Query_10005      IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL
1c1|Query_10006      IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL
1c1|Query_10007      IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL
1c1|Query_10008      VEKEGDSILITSLTGENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL
1c1|Query_10009      IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL
1c1|Query_10010      VEKEGDSILITSLTGENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL
1c1|Query_10011      IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL
1c1|Query_10012      IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL
1c1|Query_10013      IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL
                        *****

1c1|Query_10001      GIIQQAIKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE
1c1|Query_10002      GIIQQAIKQTLGRTLLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE
1c1|Query_10003      GIIQQAIKQTLGRTLLEIQVHLSVTDARKELVTVFVQASEVNEDERARIETAIADIPDINE
1c1|Query_10004      GIIQQAIKQTLGRPLEIQVHLSVTDEARKELVTVFVQASEVDEDERARIETAIADIPDINE
1c1|Query_10005      GIIQQAIKQTLGRPLEIQVHLSVTDEARKELVTVFVQASEVDEDERARIETAIADIPDINE
1c1|Query_10006      GIIQQAIKQTLGRPLEIQVHLSVTDEARKELVTVFVQASEVDEDERARIETAIADIPDINE
1c1|Query_10007      GIIQQAIKQTLGRPLEIQVHLSVTDEARKELVTVFVQASEVDEDERARIETAIADIPDINE
1c1|Query_10008      GIIQQAIKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE
1c1|Query_10009      GIIQQAIKQTLGRTLLEIQVHLSVTDARKELVTVFVQASEVNEDERARIETAIADIPDINE
1c1|Query_10010      GIIQQAIKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE
1c1|Query_10011      GIIQQAIKQTLGRTLLEIQVHLSVTDARKELVTVFVQASEVNEDERARIETAIADIPDINE
1c1|Query_10012      GIIQQAIKQTLGRTLLEIQVHLSVTDARKELVTVFVQASEVNEDERARIETAIADIPDINE
1c1|Query_10013      GIIQQAIKQTLGRPLEIQVHLSVTDEARKELVTVFVQASEVDEDERARIETAIADIPDINE
                        *****

1c1|Query_10001      AYQKDQGSVRVLRKDARDYAVSERGKMLYIIDRRN
1c1|Query_10002      AYQKDQGSVRVLRKDARDYAVSERGKMLYIIDRRN
1c1|Query_10003      AYQKDQGSVLVVRKDARDYAVSERGKMLYIIDRRN
1c1|Query_10004      AYQKDQGSVLVVRKDARDYAVSERGKMLYIIDRRN
1c1|Query_10005      AYQKDQGSVSVLRKDARDYAVSERGKMLYIIDRRN
1c1|Query_10006      AYQKDQGSVSVLRKDARDYAVSERGKMLYIIDRRN
1c1|Query_10007      AYQKDQGSVSVLRKDARDYAVSERGKMLYIIDRRN
1c1|Query_10008      AYQKDQGSVRVLRKDARDYAVSERGKMLYIIDRRN
1c1|Query_10009      AYQKDQGSVLVVRKDARDYAVSERGKMLYIIDRRN
1c1|Query_10010      AYQKDQGSV-----
1c1|Query_10011      AYQKDQGSVLVVRKDARDYAVSERGKMLYIIDRRN
1c1|Query_10012      AYQKDQGSVLVVRKDARDYAVSERGKMLYIIDRRN
1c1|Query_10013      AYQKDQGSVSVLRKDARDYAVSERGKMLYIIDRRN
                        *****

```

|                 |                                                              |
|-----------------|--------------------------------------------------------------|
| lc1 Query_10001 | VEKEGDSILITSLTGENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL |
| lc1 Query_10002 | IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL |
| lc1 Query_10003 | IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL |
| lc1 Query_10004 | IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL |
| lc1 Query_10005 | IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL |
| lc1 Query_10006 | IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL |
| lc1 Query_10007 | IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL |
| lc1 Query_10008 | VEKEGDSILITSLTGENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL |
| lc1 Query_10009 | IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL |
| lc1 Query_10010 | VEKEGDSILITSLTGENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL |
| lc1 Query_10011 | IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL |
| lc1 Query_10012 | IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL |
| lc1 Query_10013 | IEKEGDSILITSLTRENFTPLLRYRLGDTATLSLKGDKLYLTDIQREDMSFNFMGNLIGL |
|                 | *****                                                        |
| lc1 Query_10001 | GIIQQAQKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE |
| lc1 Query_10002 | GIIQQAQKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE |
| lc1 Query_10003 | GIIQQAQKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE |
| lc1 Query_10004 | GIIQQAQKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE |
| lc1 Query_10005 | GIIQQAQKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE |
| lc1 Query_10006 | GIIQQAQKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE |
| lc1 Query_10007 | GIIQQAQKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE |
| lc1 Query_10008 | GIIQQAQKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE |
| lc1 Query_10009 | GIIQQAQKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE |
| lc1 Query_10010 | GIIQQAQKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE |
| lc1 Query_10011 | GIIQQAQKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE |
| lc1 Query_10012 | GIIQQAQKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE |
| lc1 Query_10013 | GIIQQAQKQTLGRSLEIQVHLSVTDERKELVTVFVQASEVDEDERTRIETAIADIPDINE |
|                 | *****                                                        |
| lc1 Query_10001 | AYQKDQGSVRLRKDARDYAVSERGKMLYIIDRRN                           |
| lc1 Query_10002 | AYQKDQGSVRLRKDARDYAVSERGKMLYIIDRRN                           |
| lc1 Query_10003 | AYQKDQGSVRLRKDARDYAVSERGKMLYIIDRRN                           |
| lc1 Query_10004 | AYQKDQGSVRLRKDARDYAVSERGKMLYIIDRRN                           |
| lc1 Query_10005 | AYQKDQGSVRLRKDARDYAVSERGKMLYIIDRRN                           |
| lc1 Query_10006 | AYQKDQGSVRLRKDARDYAVSERGKMLYIIDRRN                           |
| lc1 Query_10007 | AYQKDQGSVRLRKDARDYAVSERGKMLYIIDRRN                           |
| lc1 Query_10008 | AYQKDQGSVRLRKDARDYAVSERGKMLYIIDRRN                           |
| lc1 Query_10009 | AYQKDQGSVRLRKDARDYAVSERGKMLYIIDRRN                           |
| lc1 Query_10010 | AYQKDQGSVRLRKDARDYAVSERGKMLYIIDRRN                           |
| lc1 Query_10011 | AYQKDQGSVRLRKDARDYAVSERGKMLYIIDRRN                           |
| lc1 Query_10012 | AYQKDQGSVRLRKDARDYAVSERGKMLYIIDRRN                           |
| lc1 Query_10013 | AYQKDQGSVRLRKDARDYAVSERGKMLYIIDRRN                           |
|                 | *****                                                        |



## Appendix 6 COBALT Analysis for MatE Proteins

```

lcl|Query_10001 -----MAVTKAVFTEGSVLRHVVFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSL
lcl|Query_10002 -----MAVTKAVFTEGSVLRHVVFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSL
lcl|Query_10003 -----MAVTKAVFTKGSMLRHVVFMSTGTSATSLLSIFLIEVLTVVYIAMLRDDSL
lcl|Query_10004 -----MAVTKAVFTEGSVLRHVVFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSL
lcl|Query_10005 -----MAVTKAVFIEGSVLRHVVFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSL
lcl|Query_10006 -----MAVTKAVFTEGSVLRHVVFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSL
lcl|Query_10007 -----MAVTKAVFTKGSMLRHVVFMSTGTSATSLLSIFLIEVLTVVYIAMLRDDSL
lcl|Query_10008 -----MAVTKAVFTKGSMLRHVVFMSTGTSATSLLSIFLIEVLTVVYIAMLRDDSL
lcl|Query_10009 -----MAVTKAVFTEGSVLRHVVFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSL
lcl|Query_10010 -----SATSLLSIFLIEVLTVVYIAMLRDDSL
lcl|Query_10011 -----MAVTKAVFTEGSVLRHVVFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSL
lcl|Query_10012 -----MAVTKAVFTEGSVLRHVVFMTGTSATSLLSIFLIEVLTVVYIAMLRDDSL
lcl|Query_10013 MKSVRREGSMVTKAVFTKGSMLRHVVFMSTGTSATSLLSIFLIEVLTVVYIAMLRDDSL
*****

lcl|Query_10001 AAFAVAKTLMFFLISMILGIAVAVSTAVSRSLGSAAPDQARRLASSGLLLGIATVCGAA
lcl|Query_10002 AAFAVAKTLMFFIISMILGIAVAVSSTVSRALGSGVPDQARRLPSSGLLLGIATVCCAV
lcl|Query_10003 AAFAVAKTLMFFLISMILGIAVAVSTTVSRSLGSGVPDQARRLASSGVLGSMATVCGAA
lcl|Query_10004 AAFAVAKTLMFFIISMILGIAVAVSSTVSRALGSGVPDQARRLASSGLLLGIATVCCAV
lcl|Query_10005 AAFAVAKTLMFFLISMILGIAVAVSTAVSRSLGSAAPDQARRLASSGLLLGIATVCGAA
lcl|Query_10006 AAFAVAKTLMFFLISMILGIAVAVSTAVSRSLGSAAPDQARRLASSGLLLGIATVCGAA
lcl|Query_10007 AAFAVAKTLMFFLISMILGIAVAVSTTVSRSLGSGVPDQARRLASSGVLGSMATVCGAA
lcl|Query_10008 AAFAVAKTLMFFLISMILGIAVAVSTTVSRSLGSGVPDQARRLASSGVLGSMATVCGAA
lcl|Query_10009 AAFAVAKTLMFFIISMILGIAVAVSSTVSRALGSGVPDQARRLASSGLLLGIATVCCAV
lcl|Query_10010 AAFAVAKTLMFFLISMILGIAVAVSTAVSRSLGSAAPDQARRLASSGLLLGIATVCGAA
lcl|Query_10011 AAFAVAKTLMFFIISMILGIAVAVSSTVSRALGSGVPDQARRLASSGLLLGIATVCCAV
lcl|Query_10012 AAFAVAKTLMFFIISMILGIAVAVSSTVSRALGSGVPDQARRLASSGLLLGIATVCCAV
lcl|Query_10013 AAFAVAKTLMFFLISMILGIAVAVSTTVSRSLGSGVPDQARRLASSGVLGSMATVCGAA
*****

lcl|Query_10001 LELVVLDPAIRMLGGEGNSFTAAREYLWLTLPASVLVAVELLQVQILRSAGFGKQAMWIL
lcl|Query_10002 LELIFLDSAIQMLGGEGSSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTL
lcl|Query_10003 LELVVLDPQAIRMLGAEGNSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTL
lcl|Query_10004 LELIFLDSAIQMLGGEGSSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTL
lcl|Query_10005 LELVVLDPAIRMLGGEGNSFTAAREYLWLTLPASVLVAVELLQVQILRSAGFGKQAMWIL
lcl|Query_10006 LELVVLDPAIRMLGGEGNSFTAAREYLWLTLPASVLVAVELLQVQILRSAGFGKQAMWIL
lcl|Query_10007 LELVVLDPQAIRMLGAEGNSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTL
lcl|Query_10008 LELVVLDPQAIRMLGAEGNSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTL
lcl|Query_10009 LELIFLDSAIQMLGGEGSSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTL
lcl|Query_10010 LELVVLDPAIRMLGGEGNSFTAAREYLWLTLPASVLVAVELLQVQILRSAGFGKQAMWIL
lcl|Query_10011 LELIFLDSAIQMLGGEGSSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTL
lcl|Query_10012 LELIFLDSAIQMLGGEGSSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTL
lcl|Query_10013 LELVVLDPQAIRMLGAEGNSFTAAREYLWLTLPASVLVAVEQLCVQILRSAGFGKQAMWTL
*** **

lcl|Query_10001 LTSTAATAADPILLIVFDLGLVGAGIAYLISSLVSACLGFYYVHRVARLTSGVSLTHLS
lcl|Query_10002 LTSTAATAADPILLIVFDLGLAGAGVAYLISSLVSACLGFYYVHRVARLTCTRVSLTHLS
lcl|Query_10003 LTSTAATAADPILLIVFDLGLVGAGIAYLISSLVSACLGFYYVHRVARLTCTRVSLKNLS
lcl|Query_10004 LTSTAATAADPILLIVFDLGLAGAGVAYLISSLVSACLGFYYVHRVARLTCTRVSLTHLS
lcl|Query_10005 LTSTAATAADPILLIVFDLGLVGAGIAYLISSLVSACLGFYYVHRVARLTSGVSLTHLS
lcl|Query_10006 LTSTAATAADPILLIVFDLGLVGAGIAYLISSLVSACLGFYYVHRVARLTSGVSLTHLS
lcl|Query_10007 LTSTAATAADPILLIVFDLGLVGAGIAYLISSLVSACLGFYYVHRVARLTCTRVSLKNLS
lcl|Query_10008 LTSTAATAADPILLIVFDLGLVGAGIAYLISSLVSACLGFYYVHRVARLTCTRVSLKNLS
lcl|Query_10009 LTSTAATAADPILLIVFDLGLAGAGVAYLISSLVSACLGFYYVHRVARLTCTRVSLTHLS
lcl|Query_10010 LTSTAATAADPILLIVFDLGLVGAGIAYLISSLVSACLGFYYVHRVARLTSGVSLTHLS
lcl|Query_10011 LTSTAATAADPILLIVFDLGLAGAGVAYLISSLVSACLGFYYVHRVARLTCTRVSLTHLS
lcl|Query_10012 LTSTAATAADPILLIVFDLGLAGAGVAYLISSLVSACLGFYYVHRVARLTCTRVSLTHLS
lcl|Query_10013 LTSTAATAADPILLIVFDLGLVGAGIAYLISSLVSACLGFYYVHRVARLTCTRVSLKNLS
** *****

```

```

lcl|Query_10001      GDISTISRTAVPAVIGNLATPVGMAVYMAAMAPFGSQUALAAIGVIDRVIQVAFVCFVFPALP
lcl|Query_10002      GDISIIGRTAVPAVIGNLATPVGMAVYMAAMAPFGAQUALAAIGVIDRVIQVAFVCFVFPALP
lcl|Query_10003      GDIRNIGRTALPAVIGNLATPVGMAVYMAAMAPFGSQUALATIGVIDRVIQVAFVCFVFPALP
lcl|Query_10004      GDISIIGRTAVPAVIGNLATPVGMAVYMAAMAPFGAQUALAAIGVIDRVIQVAFVCFVFPALP
lcl|Query_10005      GDISTISRTAVPAVIGNLATPVGMAVYMAAMAPFGSQUALAAIGVIDRVIQVAFVCFVFPALP
lcl|Query_10006      GDIRNIGRTALPAVIGNLATPVGMAVYMAAMAPFGSQUALATIGVIDRVIQVAFVCFVFPALP
lcl|Query_10007      GDIRNIGRTALPAVIGNLATPVGMAVYMAAMAPFGSQUALATIGVIDRVIQVAFVCFVFPALP
lcl|Query_10008      GDIRNIGRTALPAVIGNLATPVGMAVYMAAMAPFGSQUALATIGVIDRVIQVAFVCFVFPALP
lcl|Query_10009      GDISIIGRTAVPAVIGNLATPVGMAVYMAAMAPFGAQUALAAIGVIDRVIQVAFVCFVFPALP
lcl|Query_10010      GDISTISRTAVPAVIGNLATPVGMAVYMAAMAPFGSQUALAAIGVIDRVIQVAFVCFVFPALP
lcl|Query_10011      GDISIIGRTAVPAVIGNLATPVGMAVYMAAMAPFGAQUALAAIGVIDRVIQVAFVCFVFPALP
lcl|Query_10012      GDIRNIGRTALPAVIGNLATPVGMAVYMAAMAPFGAQUALAAIGVIDRVIQVAFVCFVFPALP
lcl|Query_10013      GDIRNIGRTALPAVIGNLATPVGMAVYMAAMAPFGSQUALATIGVIDRVIQVAFVCFVFPALP
***  *  ***  *****  *****  *****  *****

lcl|Query_10001      GALIPILGQNLGAMNTTRVSHAVKATYGLLICYGAAATSLLLILLAEPRLARLFHVSGEGLI
lcl|Query_10002      GALIPILGQNLGAMNTARVSQAVKTTYGLLIGYGATTSLLLILLAEPRLARLFHVSGEGLV
lcl|Query_10003      GALIPILGQNLGAMNTARVSAIKITYGLLIAYGSVTSLLLILLAEPRLARLFHVSAGEGLV
lcl|Query_10004      GALIPILGQNLGAMNTARVSQAVKTTYGLLIGYGATTSLLLILLAEPRLARLFHVSAGEGLV
lcl|Query_10005      GALIPILGQNLGAMNTTRVSHAVKATYGLLICYGAAATSLLLILLAEPRLARLFHVSAGEGLI
lcl|Query_10006      GALIPILGQNLGAMNTTRVSHAVKATYGLLICYGAAATSLLLILLAEPRLARLFHVSAGEGLI
lcl|Query_10007      GALIPILGQNLGAMNTARVSQAIKITYGLLIAYGSVTSLLLILLAEPRLARLFHVSAGEGLV
lcl|Query_10008      GALIPILGQNLGAMNTARVSQAIKITYGLLIAYGSVTSLLLILLAEPRLARLFHVSAGEGLV
lcl|Query_10009      GALIPILGQNLGAMNTARVSQAVKTTYGLLIGYGATTSLLLILLAEPRLARLFHVSAGEGLV
lcl|Query_10010      GALIPILGQNLGAMNTTRVSHAVKATYGLLICYGAAATSLLLILLAEPRLARLFHVSAGEGLI
lcl|Query_10011      GALIPILGQNLGAMNTARVSQAVKTTYGLLIGYGATTSLLLILLAEPRLARLFHVSAGEGLV
lcl|Query_10012      GALIPILGQNLGAMNTARVSQAVKTTYGLLIGYGATTSLLLILLAEPRLARLFHVSAGEGLV
lcl|Query_10013      GALIPILGQNLGAMNTARVSAIKITYGLLIAYGSVTSLLLILLAEPRLARLFHVSAGEGLV
*****  *****  ***  *  *  *****  **  *****  *****  *****

lcl|Query_10001      VVSAFCRWGGFLWTLIGLQFIATSIIFLSMGRPMYVTLFGWLRASLGTVPFVWYGAQAFGS
lcl|Query_10002      VVSAFCRWGGLLWTLIGLQFIATSIIFLSMGRPMYVTLFGWLRASLGTVPFVWYGAQAFGS
lcl|Query_10003      VVFAFCRWGGVLWTLIGLQFIATSVFLSMGRPAYVTLFGWLRASLGTMPFVWYGAHKFGS
lcl|Query_10004      VVSAFCRWGGLLWTLIGLQFIATSIIFLSMGRPMYVTLFGWLRASLGTVPFVWYGAQAFGS
lcl|Query_10005      VVSAFCRWGGFLWTLIGLQFIATSIIFLSMGRPMYVTLFGWLRASLGTVPFVWYGAQAFGS
lcl|Query_10006      VVSAFCRWGGFLWTLIGLQFIATSIIFLSMGRPMYVTLFGWLRASLGTVPFVWYGAQAFGS
lcl|Query_10007      VVFAFCRWGGVLWTLIGLQFIATSVFLSMGRPAYVTLFGWLRASLGTMPFVWYGAHKFGS
lcl|Query_10008      VVFAFCRWGGVLWTLIGLQFIATSVFLSMGRPAYVTLFGWLRASLGTMPFVWYGAHKFGS
lcl|Query_10009      VVSAFCRWGGLLWTLIGLQFIATSIIFLSMGRPMYVTLFGWLRASLGTVPFVWYGAQAFGS
lcl|Query_10010      VVSAFCRWGGFLWTLIGLQFIATSIIFLSMGRPMYVTLFGWLRASLGTVPFVWYGAQAFGS
lcl|Query_10011      VVSAFCRWGGLLWTLIGLQFIATSIIFLSMGRPMYVTLFGWLRASLGTVPFVWYGAQAFGS
lcl|Query_10012      VVSAFCRWGGLLWTLIGLQFIATSIIFLSMGRPMYVTLFGWLRASLGTVPFVWYGAQAFGS
lcl|Query_10013      VVFAFCRWGGVLWTLIGLQFIATSVFLSMGRPAYVTLFGWLRASLGTMPFVWYGAHKFGS
**  *****  *****  *****  ***  ***  *****  *****  *****

lcl|Query_10001      VGVMLGQLLGNTVVAFCACWLAHLLMKKIL-----GNE-----AIH----
lcl|Query_10002      VGVMLGQLLGNTIVAFACWVAHLLMKKRI-----GNKDLTGNRSRHRGNS
lcl|Query_10003      VGVMLGQLMGNTIVAFACWLAHLLMKKMLDIEIHSIGNR-----SLHRGNS
lcl|Query_10004      VGVMLGQLLGNTIVAFACWVAHLLMKKRI-----GNKDLTGNRSRHRGNS
lcl|Query_10005      VGVMLGQLLGNTVVAFCACWLAHLLMKKVL-----GNK-----AIH----
lcl|Query_10006      VGVMLGQLLGNTVVAFCACLLAHLLMKKVL-----GNE-----AIH----
lcl|Query_10007      VGVMLGQLMGNTIVAFACWLAHLLMKKMLDIEIHSIGNR-----SLHRGNS
lcl|Query_10008      VGVMLGQLMGNTIVAFACWLAHLLMKKMLDIEIHSIGNR-----SLHRGNS
lcl|Query_10009      VGVMLGQLLGNTIVAFACWVAHLLMKKRI-----GNKDLTGNRSRHRGNS
lcl|Query_10010      VGVMLGQLLGNTVVAFCACWLAHLLMKKIL-----GNE-----AIH----
lcl|Query_10011      VGVMLGQLLGNTIVAFACWVAHLLMKKRI-----GNKDLTGNRSRHRGNS
lcl|Query_10012      VGVMLGQLLGNTIVAFACWVAHLLMKKRI-----GNKDLTGNRSRHRGNS
lcl|Query_10013      VGVMLGQLMGNTIVAFACWLAHLLMKKMLDIEIHSIGNR-----SLHRGNS
*****  ***  *****  *****  **  *

```

## Appendix 7 SMART Table for MatE Domains

Protein domains found by the SMART server in the MatE proteins of three species of interest with the name of domains and regions they are found at, as well as the E-value for that domain, if available. The focus is on the MatE domains. As it can be seen, the total length of the domains is the same, even though the start and end are at different positions.

**Table 7.1:** *P. syringae* pv. *actinidiae*

| Domain name          | Start | End | Length | E-value |
|----------------------|-------|-----|--------|---------|
| Pfam: MatE           | 1     | 161 | 161    | 4.8E-15 |
| Transmembrane region | 170   | 192 | 23     | N/A     |
| Pfam: MatE           | 220   | 380 | 161    | 2.8E-15 |
| Transmembrane region | 391   | 413 | 23     | N/A     |

**Table 7.2:** *P. syringae* pv. *tomato*

| Domain name          | Start | End | Length | E-value |
|----------------------|-------|-----|--------|---------|
| Pfam: MatE           | 45    | 205 | 161    | 2.0E-15 |
| Transmembrane region | 214   | 236 | 23     | N/A     |
| Pfam: MatE           | 264   | 424 | 161    | 1.1E-14 |
| Transmembrane region | 435   | 457 | 23     | N/A     |

**Table 7.3:** *P. savastanoi* pv. *savastanoi*

| Domain name          | Start | End | Length | E-value |
|----------------------|-------|-----|--------|---------|
| Pfam: MatE           | 33    | 193 | 161    | 6.3E-16 |
| Transmembrane region | 202   | 224 | 23     | N/A     |
| Pfam: MatE           | 252   | 413 | 162    | 3.1E-12 |
| Transmembrane region | 423   | 445 | 23     | N/A     |

## Appendix 8 Intergenic sequence of the MatE and IaaL protein

>NZ\_RBQS01000231.1:c43018-42900 *Pseudomonas syringae* pv. *actinidiae* strain

ICMP 18708 PsalCMP18708\_Contig\_76, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_RBSG01000878.1:c3545-3427 *Pseudomonas syringae* pv. *actinidiae* strain

ICMP 19071 PsalCMP19071\_Contig\_9, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_JAESNG010000017.1:55443-55561 *Pseudomonas syringae* pv. *actinidiae* strain

Guizhou\_G25 G25\_15, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_SNQS01000018.1:55443-55561 *Pseudomonas syringae* pv. *actinidiae* strain

S2 S2\_0018, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_JAESNF010000017.1:55433-55551 *Pseudomonas syringae* pv. *actinidiae* strain

Guizhou\_G35 G35\_41, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_SNQU01000019.1:55443-55561 *Pseudomonas syringae* pv. *actinidiae* strain

S21 S21\_0019, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_JAESNH010000014.1:59032-59150 *Pseudomonas syringae* pv. *actinidiae* strain  
Guizhou\_G1 G1\_138, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_JABEQW010000007.1:c43083-42965 *Pseudomonas syringae* pv. *actinidiae* strain  
kw1606 NODE\_7\_length\_98922\_cov\_134.888897, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_SNQV01000018.1:55443-55561 *Pseudomonas syringae* pv. *actinidiae* strain  
S23 S23\_0018, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_SNQT01000019.1:55441-55559 *Pseudomonas syringae* pv. *actinidiae* strain  
S14 S14\_0019, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_JAESNE010000110.1:c9007-8889 *Pseudomonas syringae* pv. *actinidiae* strain  
Guizhou\_G40 G40\_138, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_JABEQR010000006.1:55839-55957 *Pseudomonas syringae* pv. *actinidiae* strain  
kw83 NODE\_6\_length\_98921\_cov\_143.469578, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_JABEQT010000006.1:c43083-42965 *Pseudomonas syringae* pv. *actinidiae* strain  
kw503 NODE\_6\_length\_98775\_cov\_164.623721, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC

ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_JABEQU010000006.1:55692-55810 *Pseudomonas syringae* pv. *actinidiae* strain  
kw1361 NODE\_6\_length\_98774\_cov\_180.600434, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_JABEQS010000006.1:55839-55957 *Pseudomonas syringae* pv. *actinidiae* strain  
kw98 NODE\_6\_length\_98921\_cov\_145.375137, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_JABEQV010000006.1:55692-55810 *Pseudomonas syringae* pv. *actinidiae* strain  
kw1483 NODE\_6\_length\_98774\_cov\_125.996717, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_BGJZ01000405.1:c6195-6077 *Pseudomonas syringae* pv. *actinidiae* strain  
MAFF212206, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTHZ01000005.1:59312-59430 *Pseudomonas syringae* pv. *actinidiae* strain  
C48 BV336\_5, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTHV01000017.1:59843-59961 *Pseudomonas syringae* pv. *actinidiae* strain  
C28 BV332\_17, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTCQ01000007.1:59335-59453 *Pseudomonas syringae* pv. *actinidiae* strain

C3 BV319\_7, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTIB01000007.1:59395-59513 *Pseudomonas syringae* pv. *actinidiae* strain

C62 BV338\_7, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTHR01000010.1:59707-59825 *Pseudomonas syringae* pv. *actinidiae* strain

C18 BV328\_10, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTHM01000020.1:59915-60033 *Pseudomonas syringae* pv. *actinidiae* strain

C13 BV323\_20, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTHW01000017.1:59474-59592 *Pseudomonas syringae* pv. *actinidiae* strain

C29 BV333\_17, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTZA01000008.1:59915-60033 *Pseudomonas syringae* pv. *actinidiae* strain

NZ39 BV356\_8, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTZH01000008.1:59915-60033 *Pseudomonas syringae* pv. *actinidiae* strain

NZ48 BV364\_8, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTZB01000008.1:59768-59886 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ40 BV357\_8, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTYX01000018.1:59335-59453 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ35 BV353\_18, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTYW01000023.1:59258-59376 *Pseudomonas syringae* pv. *actinidiae* strain  
K7 BV352\_23, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTYV01000148.1:c4175-4057 *Pseudomonas syringae* pv. *actinidiae* strain  
K6 BV351\_148, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_PGSS01000195.1:c43238-43120 *Pseudomonas syringae* pv. *actinidiae* strain  
MAFF 212109 212109\_contig\_195, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTHU01000006.1:59474-59592 *Pseudomonas syringae* pv. *actinidiae* strain  
C27 BV331\_6, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTHN01000018.1:59915-60033 *Pseudomonas syringae* pv. *actinidiae* strain  
C14 BV324\_18, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTHJ01000019.1:59789-59907 *Pseudomonas syringae* pv. *actinidiae* strain  
C10 BV320\_19, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTYS01000008.1:59768-59886 *Pseudomonas syringae* pv. *actinidiae* strain  
J38 BV347\_8, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTHL01000018.1:59789-59907 *Pseudomonas syringae* pv. *actinidiae* strain  
C12 BV322\_18, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_BGKA01000018.1:55467-55585 *Pseudomonas syringae* pv. *actinidiae* strain  
MAFF212211, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTYN01000018.1:59313-59431 *Pseudomonas syringae* pv. *actinidiae* strain  
C70 BV343\_18, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTYL01000007.1:59474-59592 *Pseudomonas syringae* pv. *actinidiae* strain  
C67 BV340\_7, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTZJ01000007.1:59915-60033 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ49 BV365\_7, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTZG01000007.1:59915-60033 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ46 BV362\_7, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTZE01000007.1:59915-60033 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ43 BV360\_7, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTY01000007.1:59915-60033 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ37 BV354\_7, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTYR01000022.1:59312-59430 *Pseudomonas syringae* pv. *actinidiae* strain  
C75 BV346\_22, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTYP01000019.1:59313-59431 *Pseudomonas syringae* pv. *actinidiae* strain  
C73 BV344\_19, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTYM01000019.1:59474-59592 *Pseudomonas syringae* pv. *actinidiae* strain  
C68 BV341\_19, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC

ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTZK01000006.1:59768-59886 *Pseudomonas syringae* pv. *actinidiae* strain

NZ54 BV366\_6, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC

ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTZI01000006.1:59768-59886 *Pseudomonas syringae* pv. *actinidiae* strain

NZ47 BV363\_6, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC

ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_PGSZ01000201.1:c3694-3576 *Pseudomonas syringae* pv. *actinidiae* strain

1404 1404Shizuoka-Psa3\_contig\_201, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC

ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTZD01000007.1:59417-59535 *Pseudomonas syringae* pv. *actinidiae* strain

NZ42 BV359\_7, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC

ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTYT01000147.1:c3605-3487 *Pseudomonas syringae* pv. *actinidiae* strain

K4 BV349\_147, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC

ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTIA01000007.1:59843-59961 *Pseudomonas syringae* pv. *actinidiae* strain

C54 BV337\_7, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC

ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTHY01000006.1:59474-59592 *Pseudomonas syringae* pv. *actinidiae* strain

C31 BV335\_6, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_PGSQ01000188.1:c8954-8836 *Pseudomonas syringae* pv. *actinidiae* strain

MAFF 212111 212111\_contig\_188, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_PG01000066.1:c8734-8616 *Pseudomonas syringae* pv. *actinidiae* strain

Fukuoka1 Fukuoka-Psa3\_contig\_66, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MD01000016.1:55453-55571 *Pseudomonas syringae* pv. *actinidiae* strain

M401 M401\_016, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_RBSR01000267.1:c7360-7242 *Pseudomonas syringae* pv. *actinidiae* strain

ICMP 9855 PsaICMP9855\_Contig\_148, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTHX01000006.1:59474-59592 *Pseudomonas syringae* pv. *actinidiae* strain

C30 BV334\_6, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTYQ01000004.1:59312-59430 *Pseudomonas syringae* pv. *actinidiae* strain

C74 BV345\_4, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTY001000007.1:59474-59592 *Pseudomonas syringae* pv. *actinidiae* strain  
C69 BV342\_7, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTHS01000023.1:59474-59592 *Pseudomonas syringae* pv. *actinidiae* strain  
C24 BV329\_23, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTHT01000022.1:59312-59430 *Pseudomonas syringae* pv. *actinidiae* strain  
C26 BV330\_22, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_PEJM01000208.1:c43557-43439 *Pseudomonas syringae* pv. *actinidiae* strain  
Psa 127771 SPAdes\_8\_len:103399, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_PEJL01000194.1:60051-60169 *Pseudomonas syringae* pv. *actinidiae* strain  
Psa 105743 PERGA\_2\_len:103520, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_PEIP01000001.1:c43082-42964 *Pseudomonas syringae* pv. *actinidiae* strain  
IPV-B0-9312 IPVB09312\_contig\_1, whole genome shotgun sequence  
TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_PEIO01000091.1:c43019-42901 *Pseudomonas syringae* pv. *actinidiae* strain  
IPV-B0-9286 IPVB09286\_contig\_92, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_PEJN01000228.1:c43496-43378 *Pseudomonas syringae* pv. *actinidiae* strain  
Psa JA12 PSA13\_contig\_229, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_PEJK01000092.1:59769-59887 *Pseudomonas syringae* pv. *actinidiae* strain  
Psa127784 PERGA\_39\_len:106906, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTZL01000008.1:59768-59886 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ59 BV367\_8, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_PEIN01000033.1:c43069-42951 *Pseudomonas syringae* pv. *actinidiae* strain  
IPV-B0-8893 IPVB08893\_contig\_33, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTZC01000007.1:59915-60033 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ41 BV358\_7, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MDXH01000003.1:59183-59301 *Pseudomonas syringae* pv. *actinidiae* strain  
M256 M256\_003, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTYZ01000007.1:59915-60033 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ38 BV355\_7, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_PEHG01000004.1:55356-55474 *Pseudomonas syringae* pv. *actinidiae* strain  
C1-PS-12 *Pseudomonas*2\_S2\_L001\_R1\_001\_trimmed\_contig\_4, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MDXF01000016.1:55445-55563 *Pseudomonas syringae* pv. *actinidiae* strain  
M227 M227\_016, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_PVZL01000223.1:59156-59274 *Pseudomonas syringae* pv. *actinidiae* strain  
MAFF 212138 212138-biovar6\_trimmed\_contig\_223, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTHQ01000062.1:27568-27686 *Pseudomonas syringae* pv. *actinidiae* strain  
C17 BV327\_62, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTYU01000005.1:59768-59886 *Pseudomonas syringae* pv. *actinidiae* strain  
K5 BV350\_5, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC  
ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MSBX01000060.1:58984-59102 *Pseudomonas syringae* pv. *actinidiae* strain  
MAFF 212141 Contig\_60, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC

ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MOMK01000007.1:c43954-43836 *Pseudomonas syringae* pv. *actinidiae* strain

Psa NZ-62 NZ62\_pilon\_scaffolds\_0007, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC

ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTH001000010.1:59560-59678 *Pseudomonas syringae* pv. *actinidiae* strain

C15 BV325\_10, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC

ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MSBW01000009.1:c4482-4364 *Pseudomonas syringae* pv. *actinidiae* strain

MAFF 212134 Contig\_9, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC

ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MOMN01000008.1:c43133-43015 *Pseudomonas syringae* pv. *actinidiae* strain

Psa NZ-65 NZ65\_pilon\_scaffolds\_0008, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC

ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

>NZ\_MTYK01000006.1:59312-59430 *Pseudomonas syringae* pv. *actinidiae* strain

C66 BV339\_6, whole genome shotgun sequence

TTAGGGAACCGATCTGTCCATCGGGGTAACATCATAGCCCGCGTTACCGCCAATCAGCCGCTGCTGATGGGCGGATACGTC

ATCTAGATAATCAAAGAAATGGACGATAGAGGACTCTCA

## Appendix 9 Promoter sequence of MatE and IaaL

>NZ\_RBQS01000231.1:c44458-44358 *Pseudomonas syringae* pv. *actinidiae* strain ICMP 18708 PsalCMP18708\_Contig\_76, whole genome shotgun sequence  
 TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
 CAACGGGAAGAGAAGGGTCAA

>NZ\_RBSG01000878.1:c4985-4885 *Pseudomonas syringae* pv. *actinidiae* strain ICMP 19071 PsalCMP19071\_Contig\_9, whole genome shotgun sequence  
 TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGCGATGGCGTT  
 CAACGGGAAGAGAAGGGTCAA

>NZ\_JAESNG010000017.1:54003-54103 *Pseudomonas syringae* pv. *actinidiae* strain Guizhou\_G25 G25\_15, whole genome shotgun sequence  
 TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
 CAACGGGAAGAGAAGGGTCAA

>NZ\_SNQS01000018.1:54003-54103 *Pseudomonas syringae* pv. *actinidiae* strain S2 S2\_0018, whole genome shotgun sequence  
 TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
 CAACGGGAAGAGAAGGGTCAA

>NZ\_JAESNF010000017.1:53993-54093 *Pseudomonas syringae* pv. *actinidiae* strain Guizhou\_G35 G35\_41, whole genome shotgun sequence  
 TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
 CAACGGGAAGAGAAGGGTCAA

>NZ\_SNQU01000019.1:54003-54103 *Pseudomonas syringae* pv. *actinidiae* strain S21 S21\_0019, whole genome shotgun sequence  
 TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
 CAACGGGAAGAGAAGGGTCAA

>NZ\_JAESNH010000014.1:57592-57692 *Pseudomonas syringae* pv. *actinidiae* strain  
Guizhou\_G1 G1\_138, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_JABEQW010000007.1:c44523-44423 *Pseudomonas syringae* pv. *actinidiae* strain  
kw1606 NODE\_7\_length\_98922\_cov\_134.888897, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_SNQV01000018.1:54003-54103 *Pseudomonas syringae* pv. *actinidiae* strain  
S23 S23\_0018, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_SNQT01000019.1:54001-54101 *Pseudomonas syringae* pv. *actinidiae* strain  
S14 S14\_0019, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_JAESNE010000110.1:c10447-10347 *Pseudomonas syringae* pv. *actinidiae* strain  
Guizhou\_G40 G40\_138, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_JABEQR010000006.1:54399-54499 *Pseudomonas syringae* pv. *actinidiae* strain  
kw83 NODE\_6\_length\_98921\_cov\_143.469578, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_JABEQT010000006.1:c44523-44423 *Pseudomonas syringae* pv. *actinidiae* strain  
kw503 NODE\_6\_length\_98775\_cov\_164.623721, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT

CAACGGGAAGAGAAGGGTCAA

>NZ\_JABEQU010000006.1:54252-54352 *Pseudomonas syringae* pv. *actinidiae* strain  
kw1361 NODE\_6\_length\_98774\_cov\_180.600434, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_JABEQS010000006.1:54399-54499 *Pseudomonas syringae* pv. *actinidiae* strain  
kw98 NODE\_6\_length\_98921\_cov\_145.375137, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_JABEQV010000006.1:54252-54352 *Pseudomonas syringae* pv. *actinidiae* strain  
kw1483 NODE\_6\_length\_98774\_cov\_125.996717, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_BGJZ01000405.1:c7635-7535 *Pseudomonas syringae* pv. *actinidiae* strain  
MAFF212206, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGCGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTHZ01000005.1:57872-57972 *Pseudomonas syringae* pv. *actinidiae* strain  
C48 BV336\_5, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTHV01000017.1:58403-58503 *Pseudomonas syringae* pv. *actinidiae* strain  
C28 BV332\_17, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTCQ01000007.1:57895-57995 *Pseudomonas syringae* pv. *actinidiae* strain

C3 BV319\_7, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTIB01000007.1:57955-58055 *Pseudomonas syringae* pv. *actinidiae* strain

C62 BV338\_7, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTHR01000010.1:58267-58367 *Pseudomonas syringae* pv. *actinidiae* strain

C18 BV328\_10, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTHM01000020.1:58475-58575 *Pseudomonas syringae* pv. *actinidiae* strain

C13 BV323\_20, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTHW01000017.1:58034-58134 *Pseudomonas syringae* pv. *actinidiae* strain

C29 BV333\_17, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTZA01000008.1:58475-58575 *Pseudomonas syringae* pv. *actinidiae* strain

NZ39 BV356\_8, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTZH01000008.1:58475-58575 *Pseudomonas syringae* pv. *actinidiae* strain

NZ48 BV364\_8, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTZB01000008.1:58328-58428 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ40 BV357\_8, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTYX01000018.1:57895-57995 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ35 BV353\_18, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTYW01000023.1:57818-57918 *Pseudomonas syringae* pv. *actinidiae* strain  
K7 BV352\_23, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTYV01000148.1:c5615-5515 *Pseudomonas syringae* pv. *actinidiae* strain  
K6 BV351\_148, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGCGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_PGSS01000195.1:c44678-44578 *Pseudomonas syringae* pv. *actinidiae* strain  
MAFF 212109 212109\_contig\_195, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTHU01000006.1:58034-58134 *Pseudomonas syringae* pv. *actinidiae* strain  
C27 BV331\_6, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTHN01000018.1:58475-58575 *Pseudomonas syringae* pv. *actinidiae* strain  
C14 BV324\_18, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTHJ01000019.1:58349-58449 *Pseudomonas syringae* pv. *actinidiae* strain  
C10 BV320\_19, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTYS01000008.1:58328-58428 *Pseudomonas syringae* pv. *actinidiae* strain  
J38 BV347\_8, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTHL01000018.1:58349-58449 *Pseudomonas syringae* pv. *actinidiae* strain  
C12 BV322\_18, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_BGKA01000018.1:54027-54127 *Pseudomonas syringae* pv. *actinidiae* strain  
MAFF212211, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTYN01000018.1:57873-57973 *Pseudomonas syringae* pv. *actinidiae* strain  
C70 BV343\_18, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTYL01000007.1:58034-58134 *Pseudomonas syringae* pv. *actinidiae* strain  
C67 BV340\_7, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTZJ01000007.1:58475-58575 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ49 BV365\_7, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTZG01000007.1:58475-58575 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ46 BV362\_7, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTZE01000007.1:58475-58575 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ43 BV360\_7, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTY01000007.1:58475-58575 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ37 BV354\_7, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTYR01000022.1:57872-57972 *Pseudomonas syringae* pv. *actinidiae* strain  
C75 BV346\_22, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTYP01000019.1:57873-57973 *Pseudomonas syringae* pv. *actinidiae* strain  
C73 BV344\_19, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTYM01000019.1:58034-58134 *Pseudomonas syringae* pv. *actinidiae* strain  
C68 BV341\_19, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT

CAACGGGAAGAGAAGGGTCAA

>NZ\_MTZK01000006.1:58328-58428 *Pseudomonas syringae* pv. *actinidiae* strain

NZ54 BV366\_6, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT

CAACGGGAAGAGAAGGGTCAA

>NZ\_MTZI01000006.1:58328-58428 *Pseudomonas syringae* pv. *actinidiae* strain

NZ47 BV363\_6, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT

CAACGGGAAGAGAAGGGTCAA

>NZ\_PGSZ01000201.1:c5134-5034 *Pseudomonas syringae* pv. *actinidiae* strain

1404 1404Shizuoka-Psa3\_contig\_201, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT

CAACGGGAAGAGAAGGGTCAA

>NZ\_MTZD01000007.1:57977-58077 *Pseudomonas syringae* pv. *actinidiae* strain

NZ42 BV359\_7, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT

CAACGGGAAGAGAAGGGTCAA

>NZ\_MTYT01000147.1:c5045-4945 *Pseudomonas syringae* pv. *actinidiae* strain

K4 BV349\_147, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGCGATGGCGTT

CAACGGGAAGAGAAGGGTCAA

>NZ\_MTIA01000007.1:58403-58503 *Pseudomonas syringae* pv. *actinidiae* strain

C54 BV337\_7, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT

CAACGGGAAGAGAAGGGTCAA

>NZ\_MTHY01000006.1:58034-58134 *Pseudomonas syringae* pv. *actinidiae* strain

C31 BV335\_6, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_PGSQ01000188.1:c10394-10294 *Pseudomonas syringae* pv. *actinidiae* strain

MAFF 212111 212111\_contig\_188, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_PGSU01000066.1:c10174-10074 *Pseudomonas syringae* pv. *actinidiae* strain

Fukuoka1 Fukuoka-Psa3\_contig\_66, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MDXG01000016.1:54013-54113 *Pseudomonas syringae* pv. *actinidiae* strain

M401 M401\_016, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_RBSR01000267.1:c8800-8700 *Pseudomonas syringae* pv. *actinidiae* strain

ICMP 9855 PsaICMP9855\_Contig\_148, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGCGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTHX01000006.1:58034-58134 *Pseudomonas syringae* pv. *actinidiae* strain

C30 BV334\_6, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTYQ01000004.1:57872-57972 *Pseudomonas syringae* pv. *actinidiae* strain

C74 BV345\_4, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTY001000007.1:58034-58134 *Pseudomonas syringae* pv. *actinidiae* strain  
C69 BV342\_7, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTHS01000023.1:58034-58134 *Pseudomonas syringae* pv. *actinidiae* strain  
C24 BV329\_23, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTHT01000022.1:57872-57972 *Pseudomonas syringae* pv. *actinidiae* strain  
C26 BV330\_22, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_PEJM01000208.1:c44997-44897 *Pseudomonas syringae* pv. *actinidiae* strain  
Psa 127771 SPAdes\_8\_len:103399, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_PEJL01000194.1:58611-58711 *Pseudomonas syringae* pv. *actinidiae* strain  
Psa 105743 PERGA\_2\_len:103520, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_PEIP01000001.1:c44522-44422 *Pseudomonas syringae* pv. *actinidiae* strain  
IPV-B0-9312 IPVB09312\_contig\_1, whole genome shotgun sequence  
TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_PEIO01000091.1:c44459-44359 *Pseudomonas syringae* pv. *actinidiae* strain  
IPV-B0-9286 IPVB09286\_contig\_92, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_PEJN01000228.1:c44936-44836 *Pseudomonas syringae* pv. *actinidiae* strain  
Psa JA12 PSA13\_contig\_229, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_PEJK01000092.1:58329-58429 *Pseudomonas syringae* pv. *actinidiae* strain  
Psa127784 PERGA\_39\_len:106906, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTZL01000008.1:58328-58428 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ59 BV367\_8, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_PEIN01000033.1:c44509-44409 *Pseudomonas syringae* pv. *actinidiae* strain  
IPV-B0-8893 IPVB08893\_contig\_33, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTZC01000007.1:58475-58575 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ41 BV358\_7, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MDXH01000003.1:57743-57843 *Pseudomonas syringae* pv. *actinidiae* strain  
M256 M256\_003, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTYZ01000007.1:58475-58575 *Pseudomonas syringae* pv. *actinidiae* strain  
NZ38 BV355\_7, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_PEHG01000004.1:53916-54016 *Pseudomonas syringae* pv. *actinidiae* strain  
C1-PS-12 *Pseudomonas*2\_S2\_L001\_R1\_001\_trimmed\_contig\_4, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MDXF01000016.1:54005-54105 *Pseudomonas syringae* pv. *actinidiae* strain  
M227 M227\_016, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_PVZL01000223.1:57716-57816 *Pseudomonas syringae* pv. *actinidiae* strain  
MAFF 212138 212138-biovar6\_trimmed\_contig\_223, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGCGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTHQ01000062.1:26128-26228 *Pseudomonas syringae* pv. *actinidiae* strain  
C17 BV327\_62, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MTYU01000005.1:58328-58428 *Pseudomonas syringae* pv. *actinidiae* strain  
K5 BV350\_5, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT  
CAACGGGAAGAGAAGGGTCAA

>NZ\_MSBX01000060.1:57544-57644 *Pseudomonas syringae* pv. *actinidiae* strain  
MAFF 212141 Contig\_60, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGCGATGGCGTT

CAACGGGAAGAGAAGGGTCAA

>NZ\_MOMK01000007.1:c45394-45294 *Pseudomonas syringae* pv. *actinidiae* strain

Psa NZ-62 NZ62\_pilon\_scaffolds\_0007, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT

CAACGGGAAGAGAAGGGTCAA

>NZ\_MTH001000010.1:58120-58220 *Pseudomonas syringae* pv. *actinidiae* strain

C15 BV325\_10, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT

CAACGGGAAGAGAAGGGTCAA

>NZ\_MSBW01000009.1:c5922-5822 *Pseudomonas syringae* pv. *actinidiae* strain

MAFF 212134 Contig\_9, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGCGATGGCGTT

CAACGGGAAGAGAAGGGTCAA

>NZ\_MOMN01000008.1:c44573-44473 *Pseudomonas syringae* pv. *actinidiae* strain

Psa NZ-65 NZ65\_pilon\_scaffolds\_0008, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT

CAACGGGAAGAGAAGGGTCAA

>NZ\_MTYK01000006.1:57872-57972 *Pseudomonas syringae* pv. *actinidiae* strain

C66 BV339\_6, whole genome shotgun sequence

TTGTTTCGGTCACTCAGTGAGCGCTCAATAGTCCGGGGTGCTGCACTGCCGATGCTCATCTGAAAATCGGAGATGGCGTT

CAACGGGAAGAGAAGGGTCAA

**Appendix 10 Hit table best scoring BLAST results for IaaL proteins with accession numbers and code names**

This table is too big to be included here. It can be accessed online at

<https://drive.google.com/drive/folders/1IH2KpA7VrHHCaw1BsfQodg05AH162IIj>

**Appendix 11 Hit table BLASTP for MatE and IaaL proteins**

This table is too big to be included here. It can be accessed online at

<https://drive.google.com/drive/folders/1IH2KpA7VrHHCaw1BsfQodg05AH162IIj>

**Appendix 12 Phylogenetic tree for IaaL proteins accession numbers and descriptions**

| <b>Accession</b> | <b>Taxa</b>                                                  | <b>Phylogenetic tree description</b>         |
|------------------|--------------------------------------------------------------|----------------------------------------------|
| WP_087674970.1   | <i>Pseudomonas syringae</i>                                  | P.syrngae_NR                                 |
| WP_005737399.1   | <i>Pseudomonas syringae</i> group                            | P.syringae1                                  |
| RMT63213.1       | <i>Pseudomonas syringae</i> pv. theae                        | P.syringae_theae                             |
| WP_034939628.1   | <i>Erwinia mallotivora</i>                                   | Erwinia_mallotivora                          |
| WP_180558835.1   | <i>Arsenophonus</i> endosymbiont of<br><i>Apis mellifera</i> | Arsenophonus_endosymbiont<br>_Apis mellifera |
| WP_065348662.1   | <i>Pseudomonas cerasi</i>                                    | P.cerasiNR1                                  |
| WP_003380685.1   | <i>Pseudomonas syringae</i> pv.<br><i>actinidiae</i>         | P.syringae                                   |
| WP_007247550.1   | <i>Pseudomonas syringae</i> group                            | P.syringae_tomatoDC3000                      |
| WP_122285283.1   | <i>Pseudomonas syringae</i> group<br>genomosp. 3             | P.syringae_group_genomosp.3                  |
| WP_065348662.1   | <i>Pseudomonas cerasi</i>                                    | P.cerasiNR2                                  |
| WP_024697353.1   | <i>Pseudomonas avellanae</i>                                 | P.avellanaeNR1                               |
| WP_122282667.1   | <i>Pseudomonas syringae</i> group<br>genomosp. 3             | P.syringae_genomosp.3NR1                     |
| RMT63213.1       | <i>Pseudomonas syringae</i> pv. theae                        | P.syringae_theae                             |
| WP_057433268.1   | <i>Pseudomonas syringae</i> group                            | P.savastanoiNR1                              |
| WP_057456329.1   | <i>Pseudomonas savastanoi</i>                                | P.savastanoi                                 |
| KPW91586.1       | <i>Pseudomonas syringae</i> pv.<br><i>cerasicola</i>         | P.syringae_cerasicola                        |
| WP_010217103.1   | <i>Pseudomonas amygdali</i>                                  | P.amygdaliNR2                                |
| WP_065348662.1   | <i>Pseudomonas cerasi</i>                                    | P.cerasiNR2                                  |
| WP_124232043.1   | <i>Erwinia psidii</i>                                        | Erwinia.psidiiNR2                            |
| WP_180558835.1   | <i>Arsenophonus</i> endosymbiont of<br><i>Apis mellifera</i> | Apis.melliferaNR2                            |
| WP_051296966.1   | <i>Arsenophonus nasoniae</i>                                 | Arsenophonus.nasoniaeNR                      |

**Appendix 13 Phylogenetic tree for the MatE proteins accession numbers and descriptions:**

| <b>Accessions</b> | <b>Species name</b>                              | <b>Description on the tree</b> |
|-------------------|--------------------------------------------------|--------------------------------|
| WP_005782319.1    | <i>Pseudomonas syringae</i> group                | P.syringae_group_MT2           |
| WP_005763530.1    | <i>Pseudomonas amygdali</i>                      | P.amygdali_MT4                 |
| WP_230692474.1    | <i>Pseudomonas syringae</i> group<br>genomosp. 3 | P.syringae_gr_genomosp3.MT3    |
| WP_017700567.1    | <i>Pseudomonas syringae</i>                      | P.syringae_MP                  |
| WP_017710541.1    | <i>Pseudomonas syringae</i>                      | P.syringae_MP2                 |
| WP_003380687.1    | <i>Pseudomonas syringae</i>                      | P.syringae_MT1                 |
| WP_054995848.1    | <i>Pseudomonas syringae</i> group                | P.syringae_group_MT6           |
| WP_019742037.1    | <i>Pseudomonas savastanoi</i>                    | P.savastanoi_MP3               |
| WP_230692474.1    | <i>Pseudomonas syringae</i> group<br>genomosp. 3 | P.syringae_gruop_MT5           |
| NP_001238722.1    | <i>Glycine max</i>                               | Glycine_max                    |

**Appendix 14 Phylogenetic tree for the MatE proteins accession numbers and descriptions extended tree:**

| Accessions     | Species name                                              | Description on the tree     |
|----------------|-----------------------------------------------------------|-----------------------------|
| WP_005782319.1 | <i>Pseudomonas syringae</i> group                         | P.syringae_group_MT2        |
| WP_019742037.1 | <i>Pseudomonas savastanoi</i>                             | P.savastanoi_MP3            |
| WP_007247551.1 | <i>Pseudomonas syringae</i> group                         | P.syringae_group_MT3        |
| WP_017700567.1 | <i>Pseudomonas syringae</i>                               | P.syringae_MP1              |
| WP_017710541.1 | <i>Pseudomonas syringae</i>                               | P.syringae_MP2              |
| WP_003380687.1 | <i>Pseudomonas syringae</i> pv. <i>actinidiae</i>         | P.syringae_MT1              |
| AXK59714.1     | <i>Pseudomonas syringae</i> pv. <i>actinidiae</i>         | P.actinidiae_MatEefflux     |
| WP_054995848.1 | <i>Pseudomonas syringae</i> group                         | P.syringae_group_MT4        |
| EEB56701.1     | <i>Pseudomonas syringae</i> pv. <i>tomato</i> T1          | P.tomatoT1_MatEefflux       |
| WP_005782319.1 | <i>Pseudomonas syringae</i> group                         | P.syringae_gr_genomosp3.MT5 |
| WP_005763530.1 | <i>Pseudomonas amygdali</i>                               | P.amygdali_MT               |
| AAO53914.1     | <i>Pseudomonas syringae</i> pv. <i>tomato</i> str. DC3000 | P.tomatoDC3000_MatEefflux   |
| RMS74471.1     | <i>Pseudomonas savastanoi</i>                             | P.savastanoi_MatEefflux     |

## CURRICULUM VITAE

| Personal Information |                                                                                                 |
|----------------------|-------------------------------------------------------------------------------------------------|
| Name Surname         | Fatima ALIHODZIC                                                                                |
| Place of Birth       |                                                                                                 |
| Date of Birth        |                                                                                                 |
| Nationality          | <input type="checkbox"/> T.C. <input checked="" type="checkbox"/> Other: Bosnia and Herzegovina |
| Phone Number         |                                                                                                 |
| Email                |                                                                                                 |

| Educational Information |                                             |
|-------------------------|---------------------------------------------|
| B. Sc.                  |                                             |
| University              | International University of Sarajevo        |
| Faculty                 | Faculty of Engineering and Natural Sciences |
| Department              | Genetics and Bioengineering                 |
| Graduation Year         | 2017                                        |

| M. Sc.     |                                                          |
|------------|----------------------------------------------------------|
| University | Istanbul University                                      |
| Institute  | Institute of Graduate Studies in Science and Engineering |
| Department | Molecular Biology and Genetics                           |
| Programme  | Molecular Biotechnology and Genetics                     |

| Publications                                                                                                                                                                                                                                                             |  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| Alihodzic F., Aravena A. How does MatE expression affect the toxicity of <i>Pseudomonas syringae</i> and <i>Pseudomonas savastanoi</i> in tomato and kiwi? 8 International Congress of the Molecular Biology Association of Turkey, MolBiyoKon2022 (Poster Presentation) |  |