ISADORA CRISTÓFOLI PEREIRA

EVOLUTIONARY ANALYSIS OF PATHOGENICITY GENES FROM *Erwinia psidii* INFECTING *Eucalyptus* spp.

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fitopatologia, para a obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2019

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
P436e
2019

Pereira, Isadora Cristófoli, 1994-
Evolutionary analysis of pathogenicity genes from *Erwinia psidii* infecting *Eucalyptus* spp. / Isadora Cristófoli Pereira. – Viçosa, MG, 2019.
vii, 99f. : il. (algumas color.) ; 29 cm.

Orientador: Jorge Luis Badel Pacheco.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. *Erwinia psidii*. 2. Eucalipto. 3. Goiaba. 4. Secreção. I. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-Graduação em Fitopatologia. II. Título.

CDD 22 ed. 579.34

ISADORA CRISTÓFOLI PEREIRA

EVOLUTIONARY ANALYSIS OF PATHOGENICITY GENES FROM *Erwinia psidii* INFECTING *Eucalyptus* spp.

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fitopatologia, para a obtenção do título de *Magister Scientiae*.

APROVADA: 11 de julho de 2019.

Lúcio Mauro da Silva Guimarães

Pedro Marcus Pereira Vidigal

Acelino Couto Alfenas
(Coorientador)

Jorge Luis Badel Pacheco
(Orientador)

A minha avó Benedita, meus pais, minha irmã Gabriela e a Matilde, **dedico**.

AGRADECIMENTOS

A Deus, por toda luz, apoio e por proporcionar experiências que me auxiliam rumo ao caminho do crescimento espiritual, emocional e profissional.

À minha família, por todo apoio, força, compreensão, incentivo e carinho.

À Comissão orientadora deste trabalho Jorge L. Badel Pacheco, Acelino C. Alfenas, Pedro M. Pereira Vidigal e Lúcio M. Guimarães Silva, pela orientação, correções e por toda ajuda e suporte.

Ao Samuel A. Santos por todo auxílio e grande contribuição ao projeto.

À todas as amizades maravilhosas que fiz em Viçosa, que tornaram a caminhada mais leve e feliz.

Às amizades de longa data, que mesmo de longe se fizeram sempre presentes. Em especial, Ana Flávia Miquelante e Naiara P. Persegona, duas grandes incentivadoras.

À equipe dos Laboratórios de Fitobacteriologia Molecular e Patologia Florestal, por toda ajuda e também pelos momentos agradáveis de descontração.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo fornecimento da bolsa que tornou este mestrado possível.

À UFV, que abriu portas para tanto conhecimento e oportunidades.

Ao Núcleo de Análise de Biomoléculas da Universidade Federal de Viçosa pela estrutura e suporte fornecidos para a execução do projeto.

À cada pessoa que em Viçosa conheci, que tenha me auxiliado a evoluir espiritual, emocional e profissionalmente.

Por tudo aqui vivido nesses dois anos, sou grata.

O meu mais sincero muito obrigada a todos vocês!

SUMMARY

ABSTRACT vi
RESUMO vii
GENERAL INTRODUCTION 1
REFERENCES 3
ARTICLE 1: Characterization of the *hrp/hrc* cluster of *Erwinia psidii* and prediction of its type II and III effector repertoires 5
ABSTRACT 5
INTRODUCTION 6
MATERIAL AND METHODS 8
Genome sequences of *Erwinia psidii* strains 8
Prediction of the type II secretome of *E. psidii* strains 9
Type II secretome comparison among *E. psidii* strains 9
Characterization of *hrp/hrc* gene cluster of *Erwinia psidii* 11
Prediction of the type III secretome of *E. psidii* strains 11
Prediction of the subcellular localization of the candidate type III effector proteins 12
RESULTS 13
The type II effector repertoire varies among *E. psidii* strains 13
The predicted T2SE repertoire of *E. psidii* is enriched in proteins with enzymatic activity 13
The *hrp/hrc* gene cluster of *E. psidii* is collinear with others from *Enterobacteriaceae* species 14
The type III effector repertoire varies among *E. psidii* strains 14
Multiple host subcellular localizations are predicted for *E. psidii* type III effectors 16
DISCUSSION 16
ACKNOWLEDGEMENTS 21
REFERENCES 22
SUPPLEMENTARY TABLE 42
ARTICLE 2: Homology-based prediction and sequence evolution of putative type III effectors from *E. psidii* 43
ABSTRACT 43
INTRODUCTION 44
MATERIAL AND METHODS 46
Erwinia psidii genome sequences 46
Prediction of effector genes 46

Construction of effector and housekeeping gene databases 47
Prediction of functional motifs 47
Prediction of effector proteins subcellular localization 48
Phylogenetic trees construction 48
Occurrence of mobile genetic elements 50
Selection and recombination analyses 50
RESULTS 51
Two highly conserved T3SE are present in the *E. psidii* genome 51
E. psidii AvrE proteins contain signals for retention in the endoplasmic reticulum 52
Erwinia psidii AvrE proteins are predicted to localize to the plant nucleus 52
Erwinia psidii AvrE clusters in a clade separate from most *Erwinia* species 53
Erwinia psidii clusters together with other *Erwinia* species in an MLSA tree 54
Absence of mobile genetic elements (MGEs) in the *avrE* gene 54
Selection pressures acting on the *avrE* gene 55
Erwinia psidii Eop1 retains a conserved catalytic triad 55
Erwinia psidii Eop1 is predicted to localize to different plant organelles 55
Erwinia psidii Eop1 clusters in a clade separate from those of other *Erwinia* species 56
Absence of mobile genetic elements (MGEs) in the *eop1* gene 56
Selection pressures acting on the *eop1* gene 57
DISCUSSION 57
ACKNOWLEDGEMENTS 62
REFERENCES 63
SUPPLEMENTARY INFORMATION 84
GENERAL CONCLUSION 99

ABSTRACT

PEREIRA, Isadora Cristófoli, M.Sc., Universidade Federal de Viçosa, July, 2019. **Evolutionary analysis of pathogenicity genes from *Erwinia psidii* infecting *Eucalyptus* spp.** Advisor: Jorge Luis Badel Pacheco. Coadvisor: Acelino Couto Alfenas.

Erwinia psidii is a Gram-negative bacterium that threatens both guava and eucalypt production. Despite the importance of both trees to the Brazilian economy, little is known about the mechanisms underlying this bacterium pathogenicity. Thus, in this study we used bioinformatic approaches to determine the gene composition of the *hrp/hrc* cluster of *E. psidii*, which encodes for the Type III secretion system, a well-known apparatus for its importance to the pathogenicity of many plant pathogenic Gram-negative bacteria. Computational methods were also used to predict both type II and III effector protein repertoires of *E. psidii*. In addition, evolutionary analyses were performed for the candidate type III effector proteins, in order to gain a better understanding on how this bacterium pathogenicity system evolved over time. Here, we characterized and compared the gene composition of *E. psidii hrp/hrc* cluster with those of other *Enterobacteriaceae*, and predicted 11 type III effector proteins, two of which, DspA/E and Eop1, are known important effector proteins secreted by the closely related species *E. amylovora*. Interestingly, these two candidate type III effectors seem to have been acquired by *E. psidii* through horizontal gene transfer between *Erwinia* and *Pantoea* species. Also, we identified 47 candidate type II effector proteins, most of which were annotated as enzymes, with hydrolytic or non-hydrolytic activities. These results provide important knowledge for further *in vivo* analysis, through construction of mutant strains and functional characterization of the effector candidates during plant-bacteria interactions.

RESUMO

PEREIRA, Isadora Cristófoli, M.Sc., Universidade Federal de Viçosa, julho de 2019. **Análise evolutiva dos genes de patogenicidade de *Erwinia psidii* infectando *Eucalyptus* spp.** Orientador: Jorge Luis Badel Pacheco. Co-orientador: Acelino Couto Alfenas.

Erwinia psidii é uma bactéria Gram-negativa que ameça tanto a produção de goiaba quanto a de eucalipto. Apesar da importância de ambas as culturas para a economia brasileira, pouco se conhece sobre os mecanismos de patogenicidade desta bactéria. Sendo assim, neste estudo abordagens de bioinformática foram aplicadas para a determinação da composição gênica do cluster *hrp/hrc* de *E. psidii*, o qual codifica para o sistema de secreção de tipo III, bastante reconhecido por sua importância para a patogenicidade de muitas bactérias Gram-negativas fitopatogênicas. Utilizou-se métodos de bioinformática também para a predição do repertório de efetores, tanto do tipo II quanto do tipo III, de *E. psidii*. Adicionalmente, realizou-se análises evolutivas com base na sequência destes candidatos a efetores do tipo III, a fim de determinar como a patogenicidade desta espécie bacteriana evoluiu com o decorrer do tempo. Neste estudo, caracterizou-se a composição gênica do cluster *hrp/hrc* de *E. psidii* com base na comparação com o os genes *hrp/hrc* de outras bactérias pertencentes a família *Enterobacteriaceae,* bem como identificou-se 11 proteínas candidatas a efetoras do tipo III, dentre elas DspA/E e Eop1, duas importantes proteínas efetoras de *E. amylovora*, espécie filogeneticamente próxima a *E. psidii.* Curiosamente, estas duas proteínas candidatas a efetoras, parecem ter sido adquiridas por *E. psidii* através de transferência horizontal de genes entre espécies do gênero *Erwinia* e *Pantoea*. Adicionalmente, identificou-se 47 proteínas candidatas a efetoras do tipo II, sendo a maioria descrita com ação enzimática, tanto com atividade hidrolítica como não hidrolítica. Estes resultados fornecem relevante conhecimento para futuras análises *in vivo,* podendo estas serem baseadas na construção de estirpes mutantes e caracterização funcional destes candidatos a efetores durante a interação planta-bactéria.

GENERAL INTRODUCTION

Brazil is considered one of the main producers of cellulose, paper and wood panels in the world. The majority of the national area destined to forestry is occupied by *Eucalyptus* spp. (Iba 2017), a genus native to Australia, Indonesia and other neighboring islands. Because of its high adaptability to diverse climates, *Eucalyptus* spp. are intensely cultivated worldwide by the forestry industry (Santos, Auer, and Junior 2001).

However, an emerging limiting factor was reported threatening eucalypt production. A disease of unknown etiology, characterized by wilt and dieback, appeared on eucalypt trees in the Brazilian states of São Paulo, Mato Grosso do Sul and Rio Grande do Sul. A few years later, it was demonstrated that *Erwinia psidii* was the aetiologic agent responsible for the disease (Arriel et al. 2014), which is currently considered one of the most severe diseases of eucalypt in South America (Caires et al. 2019).

E. psidii was first reported in São Paulo, causing disease in guava trees (*Psidium guajava*) (Rodrigues Neto, Robbs and Yamashiro, 1987). It has been suggested that this phytopathogenic bacterium underwent a host shift from its native host to *Eucalyptus* spp. (Coutinho et al. 2011), Nonetheless, the pathogen population exhibits a high genetic homogeneity according to phylogenetic analysis based on rep-PCR (Montoya-Estrada et al. 2018) and nothing is known about the mechanisms utilized by *E. psidii* to cause disease on its host plants and how its pathogenicity genes evolved over time.

The phenotypic characteristics of plants affected by *E. psidii* vary depending on the location, time of the year and *Eucalyptus* species. But, in general, the symptoms initiate on young eucalypt trees as water-soaked lesions around the main ribs of the leaves, evolving to necrotic lesions along the midrib and secondary veins. As disease progresses, small stems can present cankers and young green barks exhibit blisters that when punctured can exudate bacterial ooze. In the most advanced stages of infection, plants are observed with the most characteristic symptom of the disease, which is shoot and branch dieback and also wilt (Coutinho et al. 2011; Ferraz et al. 2016). Depending on the environmental condition, disease incidence can reach up to 100% of the stand, which causes loss of the apical dominance and growth reduction in the trees,

leading to decreases in production. So far, little is known about effective disease control measures (Arriel et al. 2014).

Since there is currently lack of measures to control this disease on eucalypt, development of resistant plant material could be exploited as an alternative (Montoya-Estrada et al. 2018; Ferraz et al. 2016). However, such a possibility must be supported by a good knowledge on the mechanisms utilized by *E. psidii* to cause disease on its host plant. Such knowledge is fundamental to comprehend how the plant immune system perceives the pathogen and how the pathogen suppresses the plant defense response.

One of the major pathogenicity mechanisms described for a large number of Gram-negative bacterial plant pathogens is the Type III secretion system (T3SS). This apparatus is able to translocate effector proteins across both inner and outer bacterial membranes, to a host target cell. These secreted effector proteins have the capacity to modify host cellular functions, in order to stablish a cell environment favorable to the bacterial infection (Green and Mecsas 2016).

Besides the T3SS, the Type II secretion system (T2SS) is also used by many Gram-negative bacteria to secrete effector proteins with a wide range of functions. These proteins are translocated in a folded state from the periplasmic space to the extracellular environment (Korotkov, Sandkvist, and Hol 2012). The T2SS was already reported as being encoded by a considerable number of Gram-negative plant pathogenic bacterial genera, such as *Erwinia, Pseudomonas, Xanthomonas, Xylella*, *Ralstonia*, *Dickeya* and *Pectobacterium* (Jha, Rajeshwari, and Sonti 2005; Green and Mecsas 2016). Among the latter two genera, for instance, *Dickeya dadantii* and *Pectobacterium carotovorum* rely on the secretion of cell wall degrading enzymes through the T2SS to cause soft-rot symptoms in innumerous host plants (Charkowski et al. 2012).

Thus, the aim of this study was to perform the first prediction of the virulence factors involved on the pathogenicity of *E. psidii*, focusing on the characterization of the effector repertoire secreted through the T2SS and T3SS, both relevant virulence factors for bacterial species closely related to *E. psidii*. Additionally, aiming to better understand how this bacterial species evolved to be able to infect a new host, phylogenetic and selection analyses were performed to comprehend the evolutionary history of the predicted effector genes of *E. psidii* as compared to the members of the *Enterobacteriaceae* family.

REFERENCES

Arriel, D. A. A.; Fonseca, N. R.; Guimarães, L. M. S.; Hermenegildo, P. S.; Mafia, R. G.; Borges Júnior, N.; de Souza, H. P; Alfenas, A. C. 2014. Wilt and die-back of *Eucalyptus* spp. caused by *Erwinia psidii* in Brazil. For. Pathol. 44:255–265.

Caires, N. P.; Hermenegildo, P. S.; Guimarães, L. M. S.; Mafia, R. G.; Zauza, E. A. V.; Júnior, N. B.; Badel, J. L.; Alfenas, A. C. 2019. Host range of *Erwinia psidii* and genetic resistance of *Eucalyptus* and *Corymbia* species to this pathogen. For. Pathol. 12527:1-10.

Coutinho, T. A.; Brady, C. L.; Vaart, M.; Venter, S. N.; Telechea, N.; Rolfo, M.; Perez, C.; Wingfield, M. J. 2011. A new shoot and stem disease of eucalyptus species caused by *Erwinia psidii*. Australas. Plant Pathol. 40:55–60.

Charkowski, A.; Blanco, C.; Condemine, G.; Expert, D.; Franza, T.; Hayes, C.; Hugouvieux-Cotte-Pattat, N.; Solanilla, L. E.; Low, D.; Moleleki, L.; Pirhogen, M.; Pitman, A.; Perna, N.; Reverchon, S.; Palenzuela, R. P.; Francisco, M. S.; Toth, I.; Tsuyumy, S.; Waals, V. D. J.; Wolf, V. D. J.; Gijsegem, V. F.; Yang, C. H.; Yedidia, I. 2012. The role of secretion systems and small molecules in soft-rot *Enterobacteriaceae* pathogenicity. Annu. Rev. Phytopathol. 50: 425-449.

Ferraz, H. G. M.; Demuner, G. A.; Guimarães, L. M. S.; Arriel, D. A. A.; da Silva, A. C.; Junior, N. B. 2016. Methods of inoculation and evaluation of *Erwinia psidii* in eucalypt. For. Pathol. 46:240–247.

Green, E. R.; Mecsas, J. 2016. Bacterial secretion systems – an overview. Microbiol. Spectr. 4:1–32.

Iba. 2017. Relatório 2017. Indústria Bras. de Árvores. Available at: http://iba.org/images/shared/Biblioteca/IBA_RelatorioAnual2017.pdf%0Ahttp://arxiv.org/abs/1011.1669%0Ahttp://dx.doi.org/10.1088/1751-8113/44/8/085201.

Jha, G.; Rajeshwari, R.; Sonti, R. V. 2005. Bacterial type two secretion system secreted proteins: double-edged swords for plant pathogens. Mol. Plant-

Microbe Interact. 18:891–898.

Korotkov, K. V.; Sandkvist, M.; Hol, W. G. J. 2012. The type II secretion system: biogenesis, molecular architecture and mechanism. Nat. Rev. Microbiol. 10:336–351.

Montoya-estrada, C. N.; Costa, C. R.; Badel, J. L.; Guimarães, L. M. S.; Alfenas, A. C. 2018. Root infection and aerial colonization of eucalypt host plants by *Erwinia psidii*. Trop. Plant Pathol. 44:251–257.

Rodrigues Neto, J.; Robbs, C. F.; Yamashiro, T. 1987. A bacterial disease of guava (*Psidium guajava*) caused by *Erwinia psidii* sp. nov. Fitopatol. Bras. 12:345-350.

Santos, Á. F. do; Auer, C. G.; Junior, A. G. 2001. Doenças do eucalipto no Sul do Brasil: identificação e controle. Circ. Tec. 45:1–20.

ARTICLE 1: Characterization of the *hrp/hrc* cluster of *Erwinia psidii* and prediction of its type II and III effector repertoires

PEREIRA, I. C.[1], VIDIGAL, P. M. P.[3], GUIMARÃES, L. M. S.[2], SANTOS, S. A.[2], ALFENAS, A. C.[2], BADEL, J. L.[1].

[1]Laboratory of Molecular Phytobacteriology; [2]Laboratory of Forest Pathology/Bioagro; Department of Plant Pathology, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil.
[3]NuBioMol; Center of Biological Sciences, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil.

ABSTRACT

Erwinia psidii is the causal agent of important diseases of both eucalypt and guava trees. However, little is known about the pathogenicity mechanisms utilized by the bacterium to cause disease. Many plant pathogenic Gram-negative bacteria are able to cause disease based on the secretion of effector proteins through secretion systems. Among these, the Type II and III secretion systems constitute the main important pathogenicity mechanisms for bacterial species closely related to *E. psidii.* The Type II secretion system (T2SS) transports effector proteins across bacterial membranes to the extracellular milieu while the Type III secretion system (T3SS), which is encoded by *hrp*/*hrc* genes, injects these effectors directly into host cells. In this study, we utilized bioinformatic approaches in order to determine the gene composition and organization of the *hrp/hrc* cluster of *E. psidii.* In addition, we performed the prediction of the type II and III effectors that may be delivered through T2SS and T3SS, respectively. We characterized the *hrp/hrc* cluster of *E. psidii* and found synteny with the one of *E. amylovora*. Here, we report the prediction of 12 type III effector proteins and the type II effector repertoire of *E. psidii* strains, most of which genes are annotated as enzymes, with hydrolytic or non-hydrolytic activities. Thus, we provide a gain knowledge on the interaction between *E. psidii* and its host plants. This study represents the first prediction of pathogenicity genes performed on the genome of *E. psidii*.

Keywords: Eucalypt, guava, secretion, pathogenicity.

INTRODUCTION

Erwinia psidii, a Gram-negative bacterium, is the causal agent of bacterial blight of guava (*Psidium guajava*) and dieback and wilt of eucalypt (*Eucalyptus* spp.) (Rodrigues Neto, Robbs and Yamashiro, 1987; Arriel et al. 2014). Currently, dieback of eucalypt is considered one of the most severe diseases of this host plant in South America (Caires et al. 2019). Symptoms on young eucalypt trees occur as water-soaked lesions around the main ribs of the leaves and necrosis of the petiole (Coutinho et al. 2011; Hermenegildo et al. 2019; Ferraz et al. 2016). As disease progresses, plants are observed with the most characteristic symptom, which is wilting and shoot and branch dieback (Coutinho et al. 2011; Ferraz et al. 2016).

Little is known about the molecular mechanisms required for the pathogenicity of this bacterium and how they determine the capability of this species to cause the symptoms aforementioned, during its interaction with eucalypt plants. In general, the pathogenicity of Gram-negative bacteria relies on the secretion of effector proteins, which can be transported from the bacterial cytoplasm to either the extracellular environment or directly inside host cells through six different secretion systems (types I – VI). These secretion systems have variable size, composition, architecture and can secrete a wide variety of effector proteins (Costa et al. 2015; Green and Mecsas, 2016).

It is known that some *Erwinia* species have as an important virulence factor the secretion of type II effector proteins (T2SE), through the Type II secretion system (T2SS) (Barras et al. 2014; Kazemi-Pour et al. 2004). Depending on the bacterial species, these T2SE exhibit a variety of functions, the most predominant of which are associated with enzymatic activity. Some of the effectors secreted by plant pathogenic Gram-negative bacteria are known for processing carbohydrates, but the majority are cell wall-degrading enzymes, such as cellulases and proteases (Green and Mecsas 2016).

The type II effector proteins possess a secretion signal at the N-terminus of the amino acid sequence, referred to as signal peptide (SP). The secretion of these T2SE occurs in two steps. Initially, the SP is recognized by either the Tat or Sec machinery, that are responsible for transporting them across the cytoplasmic membrane to the periplasm. This step is followed by the removal of

the SP. Then, these proteins become able to fold into a conformation that is recognizable by the Type II secretion apparatus, which transport them across the bacterium outer membrane to be delivered to the extracellular environment. Once in the extracellular milieu, the T2SE become able to interact with the host cells, thus initiating the disease development process (Voulhoux et al. 2001; Johnson et al. 2006; Green and Mecsas 2016).

With the increasing availability of bacterial genome sequences, more computational techniques have been developed for the prediction of candidate effector proteins of plant pathogenic species (Sperschneider et al. 2015; Dalio et al. 2017). Generally, predictions for T2SE are based on the presence of a SP on the N-terminal amino acid sequence, since it targets the protein to the secretion apparatus (Almagro Armenteros et al. 2019; Dalio et al. 2017), plus the absence of transmembrane (TM) domains, since proteins containing a TM usually remain anchored in the membrane where they play their biological functions, therefore, are not secreted (Dalio et al. 2017; Sperschneider et al. 2015).

Another extremely important pathogenicity mechanism utilized by many plant pathogenic Gram-negative bacteria is the delivery of Type III effector proteins (T3SE) into the host tissue, in order to promote parasitism and pathogenesis (Grant et al. 2006; Jones and Dangl 2006; Lindgren 1997). These T3SE are injected directly into the cytosol of the host plant cells by the Type III secretion system (T3SS), encoded by *hrp* (for hypersensitive reaction and pathogenicity)/ *hrc* (for HR conserved) genes (Grant et al., 2006; Jones & Dangl, 2006; Lindgren, 1997). Hence, the *hrp*/*hrc* genes are generally essential for plant pathogenic bacteria to cause hypersensitive response (HR) in resistant plants and disease (pathogenicity) on susceptible host plants (Lindgren 1997; Roine et al. 1997).

The genes of the *hrp*/*hrc* cluster are involved in the formation of the Hrp pilus, a filamentous structure required for the role of the T3SS on delivering effector proteins from the bacterial cell into the host cytosol. The Hrp pilus functions as a conduit for translocation of effector proteins (Jin et al. 2002; Cornelis and Van Gijsegem 2000; Roine et al. 1997).

It is known that the T3SS is conserved in many Gram-negative plant pathogenic bacteria, including species of the genera *Xanthomonas*, *Pseudomonas*, *Ralstonia*, *Erwinia*, *Dickeya*, *Pectobacterium* and *Pantoea* (Grant

et al. 2006; Jones and Dangl 2006). Based on gene similarity, structure and regulatory components, the *hrp/hrc* clusters previously described were separated in two groups: Group I - composed of *Pseudomonas syringae* and *Erwinia amylovora*; and Group II - *Xanthomonas campestris* and *Ralstonia solanacearum* (Alfano and Collmer 1997).

Previous studies focused on the regulation of the *hrp/hrc* genes revealed that, as for the bacteria belonging to group I, the gene cluster is positively regulated by HrpL, a σ factor member of the extracellular factor family (ECF) (Xiao and Hutcheson 1994; Wei and Beer 1995). Hence, HrpL is able to activate the expression of genes that encode for the T3SS and T3SE (Innes et al. 1993; Wei and Beer 1995). Several of these genes that are activated by HrpL have in common an upstream conserved sequence, designated *hrp* box, that is recognized by HrpL as sites for transcription initiation by RNA polymerase (Innes et al. 1993; Xiao and Hutcheson 1994). Since most of the type III effector genes are preceded by these *hrp* promoters, some bioinformatic approaches were developed to identify effector protein candidates in newly sequenced genomes through sequence similarity (Vencato et al. 2006).

Considering that both Type II and Type III secretion systems play an important role in the pathogenicity of many plant pathogenic Gram-negative bacteria, including species belonging to the genus *Erwinia*, the aim of this study was to characterize the gene composition and organization of the *hrp/hrc* cluster of *E. psidii* as well as to predict the type II and III effector protein repertoires of four newly sequenced *E. psidii* strains. Therefore, this study was conducted to contribute to the characterization of the pathogenicity mechanisms of this species, and also to propose the first model of interaction between *E. psidii* and its host plants.

MATERIAL AND METHODS

Genome sequences of *Erwinia psidii* strains

This study was conducted using the genome sequences of four *E. psidii* strains: IBSBF435 (type strain; GenBank accession number RHHM00000000) (Hermenegildo et al., 2019), LPF681, LPF534, LPF640 (Hermenegildo et al., unpublished data). These strains had their genome recently sequenced using the Illumina NovaSeq 6000 platform. The resulting reads were trimmed on AfterQC (Chen et al. 2017). Then, reads showing high quality were *de novo* assembled into scaffolds using SPAdes version 3.10.1 (Bankevich et al. 2012) and their annotation was performed using NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/). Amino acid sequences were recovered from the annotation process.

These strains are good representatives for assessing the factors underlying *E. psidii* pathogenicity, since they were isolated from its different host plants (either eucalypt or guava), and exhibited different levels of aggressiveness on two different eucalypt clones (Montoya-Estrada et al. 2018). The characteristics of these strains are listed on Table 1.

Prediction of the type II secretome of *E. psidii* strains

Computational tools were employed to predict the repertoire of type II proteins that compose the secretome of *E. psidii*. The presence of secretory signal peptides (SP) in the amino terminus of protein sequences from *E. psidii* strains was predicted using SignalP 5.0 (Almagro Armenteros et al. 2019). The parameter was set for the identification of proteins containing standard secretory signal (Sec/SPI). Proteins showing positive result for Sec/SPI were then further analyzed for the presence/absence of transmembrane (TM) domains, through TMHMM v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh et al. 2001). Parameter was set for the selection of proteins presenting no predicted transmembrane helices (PredHel = 0). Proteins showing TM domains (PredHel > 0) were removed from the set of predicted proteins containing SPs (Krogh et al. 2001). This process was executed for all the four *E. psidii* strains.

Type II secretome comparison among *E. psidii* strains

The predicted repertoire of type II secreted protein of each strain was separately submitted to CD-HIT (http://weizhongli-lab.org/cd-hit/). This server was used for filtering out highly similar proteins present in the type II secretome of each strain (Fu et al. 2012). For this, the sequence identity cut-off value was set to 0.5. Proteins that clustered together in the CDHIT analysis were assumed to have similar functions, and a single sequence was retained in order to eliminate functional redundancy. Afterwards, CD-HIT-2D (http://weizhongli-lab.org/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit-2d) was employed to compare and find protein similarities among the repertoires of the four strains. The sequence identity cut-off value was also set to 0.5. Then, the clustering pattern reported by CD-HIT-2D provided information for the construction of the type II secretome of *E. psidii* strains (Huang et al. 2010). The graphic representation of the type II secretome was built as a Venn diagram, using Jvenn (http://jvenn.toulouse.inra.fr/app/example.html), in order to display a comparison of the lists of the type II predicted effectors of the four *E. psidii* strains (Bardou et al. 2014).

The predicted proteins were submitted to GoFeat (http://computationalbiology.ufpa.br/gofeat/) for functional annotation (Araujo et al. 2018). Then, a manual curation of the data was perfomed based on the functional feature of each protein. Since the purpose was to identify proteins secreted to the extracellular space, those annotated as membrane, pilus, flagellum, fimbria proteins, or that make part of the T2SS apparatus, were not considered on the secretome repertoire. Proteins that were predicted as exclusive for only one strain were aligned against the genomes of the other *E. psidii* strains with Mauve aligner, in order to confirm that their positions in the genome did not correspond to gaps in the assemblies (Darling, Mau, and Perna 2010). Finally, for a refinement of the obtained data, the coding sequences of the type II predicted proteins were run on Phaster to predict the presence of prophage sequences (virus DNA sequences integrated on the bacterial genome) (Arndt et al. 2016). The pipeline for the prediction of the type II secretomes is shown on Figure 1A.

Characterization of *hrp/hrc* gene cluster of *Erwinia psidii*

Since little is known about the arrangement and composition of the *hrp/hrc* gene cluster of *E. psidii*, this *in silico* characterization was performed based on similarity searches against other members of the *Enterobacteriaceae* family, which had their cluster previously characterized. Further information on the species used for this analysis are described on Table 2.

Preceding this analysis, the scaffolds from the whole-genome sequence of *E. psidii* IBSBF435 (type-strain) were re-ordered on Mauve Aligner (Rissman et al. 2009), using the genome of *Erwinia amylovora* CFBP 1430 as reference, since its complete genome is assembled and deposited on NCBI databases (accession number GCA_000091565.1).

Then, in order to look for sinteny between the genome of *E. psidii* and the *hrp/hrc* cluster from the other *Enterobacteriaceae*, the CLC Sequence Viewer 8.0 (https://www.qiagenbioinformatics.com/) was used to extract the encompassing sequence between the first (*hrpN*) and the last (*hrcU*) genes of the cluster from their genome sequences. The results were exported both as graphic images to compose the *hrp/hrc* alignments and as amino acid fasta files for further similarity searches by BlastP tool version 2.9.0 (Altschul et al. 1990), using *E. psidii* IBSBF435 as query and the *hrp/hrc* cluster from the other species as subjects. Through this analysis, the amino acid sequence identity of proteins from *E. psidii* with the known *hrp/hrc* proteins from *E. amylovora, Pantoea stewartii s*ubsp. s*tewartii* and *Pectobacterium carotovorum* subsp. *carotovorum* was determined.

Prediction of the type III secretome of *E. psidii* strains

The identification of putative effector proteins secreted through the T3SS was based on searches for conserved *hrp* box sequences. For this, the initial step was a literature review, in order to verify previously predicted *hrp* box sequences. The review was done focusing on *E. amylovora* and *Pseudomonas syringae*, since the first is the type species of the genus and the latter is closely related to *E. amylovora*, regarding the conservation of *hrp* box sequences. The sequences used on this study are described on Table 3.

The following step was the execution of a customized script in Python 3.7.3 (Alves et al., unpublished data) to identify and locate *hrp* box sequences in the genome of all four strains of *E. psidii.* A GenBank file format (gbff) of each strain was used as input, whereas the script generated as an output, a fasta file containing amino acid sequences encoded by genes that had a *hrp* box sequence located upstream of them. The script unveils three genes located downstream of each putative *hrp* box sequence. The results were manually curated in order to delete identified genes that were not on the same strand as the predicted *hrp* box.

The amino acid sequences recovered with the script were submitted to BlastP tool v. 2.9.0 (Altschul et al. 1990), to look for similarity with known type III effector proteins (T3SE) deposited on NCBI database (https://www.ncbi.nlm.nih.gov/). The proteins that had no identity with previously reported T3SE, were named as Eps. To determine if the Eps were conserved among *E. psidii* strains, similarity searches were done for each Eps identified against the genome sequences of all *E. psidii* strains. Aiming to confirm and obtain a better functional annotation, the Eps amino acid sequences were submitted to GoFeat (Araujo et al. 2018) for additional sequence similarity searches.

The predicted T3SE that seemed to be exclusively present in only one strain were recovered from the corresponding gbff file and aligned against the genomes of strains in which were not identified, using Mauve aligner (Rissman et al. 2009). If the T3SE sequence was aligned with a gap position in the genome assembly of any other *E. psidii* strain, it was not considered as exclusive of one strain. Finally, a venn diagram containing all the predicted T3SE was built using Jvenn (Bardou et al. 2014). The bioinformatics workflow employed for the prediction of the type III secretome of *E. psidii* is shown on Figure 4.

Prediction of the subcellular localization of the candidate type III effector proteins

WolfPsort was used for prediction of the subcellular localization within host cells of all putative T3SE identified (Horton et al. 2007).

RESULTS

The type II effector repertoire varies among *E. psidii* strains

Based on the bioinformatics pipeline, an inventory of T2SE was predicted for the four *E. psidii* strains (Figure 1B). After obtaining an initial dataset, a manual refinement was performed in order to eliminate membrane proteins, flagellar, fimbria, pilus or other structural proteins that could not be considered as extracellular. Then, Phaster was employed to detect the presence of prophage sequences among the predicted proteins. A total number of 162, 154, 141, 150 of T2SE was predicted for IBSBF435, LPF534, LPF681, LPF640, respectively. Among these, 41 proteins were common to all strains. These were represented on the overlapped region at the center of Figure 1B. A small number of exclusive T2SE for each strain was observed, the highest of which was for the type strain IBSBF435 (Supplementary Table S1).

According to host isolation source, it was observed that the strains IBSBF435 and LPF681, both isolated from *Psidium guajava*, shared 12 type II predicted effectors while LPF534 and LPF640, both isolated from *Eucalyptus* sp., possess 8 proteins in common.

The predicted T2SE repertoire of *E. psidii* is enriched in proteins with enzymatic activity

The annotation performed by GoFeat for these common proteins among all four strains indicated the following functions: hydrolytic enzymes (14 proteins); enzymes with different activities (acetolactate decarboxylase, catalase, reductase, carbonic anhydrase, oxidoreductase, dehydrogenase, transferase) (8); uncharacterized proteins (19). Figure 2 represents the results obtained for functional annotation.

The annotation performed for proteins shared between strains isolated from the same host indicated the following results: A) IBSBF435 and LPF681 (guava) - hydrolytic enzymes (4 proteins); enzymes with different activities (oxidoreductase, transferase) (2); uncharacterized proteins (6) (Figure 1A). B)

LPF640 and LPF534 (eucalypt) - hydrolytic enzymes (3 proteins); murein transglycosylase (1); uncharacterized protein (1) (Figure 1B).

With respect to the proteins exclusive to each strain, one protein of LPF534 was annotated as an alginate lyase and the other as a hypothetical protein. The exclusive proteins of strains LPF681 and IBSBF435 were assigned diverse functions, most of which associated with enzymatic activities (Supplementary Table S1), while no exclusive protein was predicted for strain LPF640.

The *hrp/hrc* gene cluster of *E. psidii* is collinear with others from *Enterobacteriaceae* species

The *hrp/hrc* cluster from *E. psidii* IBSBF435 is likely composed of a core of 26 genes, 17 of which are *hrp* and *9 hrc* genes. The predicted arrangement of the genes appears to be collinear with those of the other three Enterobacteria used for comparison (Figure 5). Interestingly, a difference noticed between the cluster of *E. psidii, P. stewartii* subsp. *stewartii* and *P. carotovorum* subsp. *carotovorum* with that of *E. amylovora* is the presence of three additional ORFs between *hrpA* and *hrpS*.

Notably, out of 26 predicted proteins of *E. psidii*, 24 showed higher identity with *E. amylovora* proteins, ranging from 62 to 93%. In contrast, only two predicted proteins showed higher identity with the respective HrpN (66%) and HrpT (83%), from *P. stewartii* subsp. *stewartii*. Thus, despite the difference aforementioned, the sequence of the *hrp/hrc* gene cluster of *E. psidii* has more identity with that of *E. amylovora,* followed by *P. stewartii* subsp. *stewartii* and lastly by *P. carotovorum* subsp. *carotovorum.*

The predicted *E. psidii* genes were named according to their sequence similarity with the cluster of *E. amylovora* and designated (from left to right): *hrpN, hrpV, hrpT, hrcC, hrpG, hrpF, hrpE, hrpD, hrcJ, hrpB, hrpA, hrpS, hrpY, hrpX, hrpL, hrpJ, hrcV, hrpQ, hrcN, hrpO, hrpP, hrcQ, hrcR, hrcS, hrcT* and *hrcU*.

The type III effector repertoire varies among *E. psidii* strains

The total number of T3SS secreted proteins predicted was 27, 24, 25 and 25 for IBSBF435, LPF534, LPF640 and LPF681, respectively. However, the execution of a manually filtering was necessary, since the script recovered sequences from chaperones and structural genes of the *hrp/hrc* cluster, that could not be considered as secreted effector proteins. After filtering the data, the final number of predicted proteins was 13, 10, 11 and 11 for IBSBF435, LPF534, LPF640 and LPF681, respectively. The final set of type III secreted proteins of all four *E. psidii* strains from the script prediction is shown on Table 3. Results from BlastP searches, that provided insights on the putative function of these proteins are also shown in Table 4.

Proteins that did not share any similarity with previously described T3SE were then submitted to GoFeat in an attempt to unveil their functions. Those proteins were designated as Eps1, Eps2, Eps3, Eps4, Eps5, Eps6 and Eps7. According to the function enrichment analyses executed through GoFeat, these proteins were respectively designated as: mannosyl-3-phosphoglycerate phosphatase; glycosyl transferase; biphenyl 2,3-dioxygenase; uncharacterized protein; amidinotransferase; uncharacterized protein; and uncharacterized protein. Since Eps6 and Eps7 were only found in the genome of *E. psidii* IBSBF435, alignments with the genome sequences of the other three *E. psidii* strains were conducted on Mauve and demonstrated that both proteins were indeed exclusive to IBSBF435, since they did not align to gaps in their assemblies. As opposed to IBSBF435, the other three *E. psidii* strains do not possess any exclusive protein.

Among all the predicted T3SE, the following 10 proteins seem to be common to all *E. psidii* strains: Eps1, Eps2, Eps3, Eps4, Eps5, Eop1, HopW1, HsvB, HsvC and YopN. Notably, some well characterized accessory proteins that aid in effector delivery were also identified by the Python script, including HrpA, HrpK, HrpN and HrpW. Curiously, IBSBF435, LPF640 and LPF681 have the HopE1 protein in common, whilst this is absent in the strain LPF534. The absence of this protein on the genome of LPF534 was confirmed through Mauve alignment. The distribution of the predicted T3SE among *E. psidii* strains is represented in Figure 6.

Multiple host subcellular localizations are predicted for *E. psidii* type III effectors

The results from WolfPsort are on Table 5. Since this algorithm gives values for several subcellular localizations (ranging from 0 to 10), multiple host subcellular localizations were predicted for the same protein. The values given by WolfPsort were too low to accurately predict the proteins subcellular localization. In any case, in this study, a rate equal or greater than 6 was used as a criterion to predict proteins localized to a specific organelle.

DISCUSSION

In this study, we characterized the type II secretome of *Erwinia psidii*, through bioinformatic approaches. Among all predicted T2SE, the prevalent class of proteins is composed of hydrolytic enzymes. Phytopathogenic bacteria that cause soft-rot disease, such as *D. chrysanthemi* and *P. carotovorum*, depend on these enzymes for their capacity to develop soft-rot symptoms on host plants (Kazemi-Pour, Condemine, and Hugouvieux-Cotte-Pattat 2004; Charkowski et al. 2012). Enzymes, such as pectate lyases, polygalacturonases, cellulases, endoglucanases and esterases, secreted through the T2SS by these species, are required for the hydrolysis of plant cell wall components (Charkowski et al. 2012; Kazemi-Pour, Condemine, and Hugouvieux-Cotte-Pattat 2004; Barras, van Gijsegem, and Chatterjee 2014). In contrast to the proteins secreted by these two bacterial species, the hydrolytic enzymes here identified in *E. psidii* were peptidases, proteases, hydrolases, phospholipases, chitinases, glucohydrolases, glucanases, phosphatases and phospodiesterases. Thus, these results could explain the absence of soft-rot symptoms when this bacterium infects eucalypt plants (Arriel et al. 2014). Besides the hydrolytic enzymes identified, seven different enzymes activities (acetolactate decarboxylase, catalase, carbonic anhydrase, dehydrogenase, oxidoreductase, reductase, and transferase) are reported in the present study. However, no previous reports were found about the importance of these enzymatic activities for the pathogenicity of bacterial species closely related to *E. psidii*.

The T2SS is known for its capacity of transporting a wide range of proteins through the secretory pathway (Korotkov, Sandkvist, and Hol 2012). The repertoire of secreted proteins varies in number and biological functions among species (Sandkvist 2001). For instance, *P. aeruginosa* can secrete up to 15 proteins through the T2SS, in contrast to more than 20 secreted by the human pathogen *Vibrio cholerae*. Our Bioinformatics pipeline consistently predicted a considerable higher number (varying from 141 to 162) of T2SE for four different *E. psidii* strains. Whether these numbers are an overestimation of the type II secretome remains to be determined. Thus, further analyses should be performed to confirm the number of type II effector proteins predicted in this study.

In this study, we also retrieved the *hrp/hrc* cluster from the genome of *E. psidii*, compared it with those of other phytopathogenic bacteria of the *Enterobacteriaceae* family and predicted proteins that may transit through the T3SS. Comparison of these *hrp/hrc* gene clusters with respect to gene organization and composition confirms that the putative *hrp/hrc* genes of *E. psidii* and *E. amylovora* are closely related to each other (Oh and Beer 2005). The cluster from *P. stewartii* subsp. *stewartii* was also previously characterized as similar to the one of *E. amylovora*, according to gene homology and synteny (Frederick et al. 2001). Hence, the results obtained in this study are in agreement with Frederick et al. 2001, since the predicted cluster of *E. psidii* showed the second highest similarity with the one from *P. stewartii* subsp. *stewartii,* demonstrating the close relationship among these three species.

Proteins coded by homologous genes of the predicted *hrp/hrc* genes of *E. psidii* have previously been shown to play important functions for the pathogenicity of other bacterial species, such as: HrpL – an alternative sigma-factor that regulates the transcription of *hrp* and *hrc* genes of *E. amylovora* and *P. syringae*, activating the expression of the type III secretion system (Wei and Beer 1995; Gaudriault et al. 1997); in the *hrpXY* operon – HrpX functions as a sensor protein and HrpY is its response regulator partner; together they form a two-component regulatory system, that act as a modulator of *hrpL* induction (Wei, Kim, and Beer 2000); HrpS – partially regulates the expression of *hrpL* in *E. amylovora* (Wei and Beer 1995). Thus, it is an important component of the regulatory network that activates *hrpL* gene expression (Wei, Kim, and Beer

2000); *hrpA* – known to encode pilin proteins for both *E. amylovora* and *P. syringae*. The pilin proteins are required for the constitution of the pilus, which is the conduit by which the T3SS reaches host cells for the injection of effector proteins (Frederick et al. 2001; Jin et al. 2002)*; hrpN* – a gene known for encoding harpins (proteinaceous elicitors of hypersensitive responses on resistant or non-host plants) on *P. carotovorum* subsp. c*arotovorum* (Mukherjee et al. 2006) and *E. amylovora* (Wei and Beer 1993; Wei et al. 1992); HrpJ – is a component of the type III secretion system, required for efficient secretion of harpins (Kim, Wei, and Beer 1997; Bogdanove, Bauer, and Beer 1998; Nissinen et al. 2007); *hrpC* operon - composed by *hrpF, hrpG, hrcC, hrpT* and *hrpV* (Deng et al. 1998); *hrpF, hrpG*, and *hrpT* - are genes known to encode proteins that are constituents of the type III secretion apparatus of *Pseudomonas syringae* (Preston et al. 1998; Deng et al. 1998); HrpV – considered as a negative regulator of *hrp* genes transcription (Preston et al. 1998); HrpP – might be exported through the T3SS of *E. amylovora*, since it shares homology with known secreted proteins (Kim, Wei, and Beer 1997); and finally, *hrpJ* – required for the pathogenicity of *E. amylovora*, it encodes a protein that is homolog to YopN, which plays an important role on the effector secretion process of *Yersinia* spp. (Bogdanove et al. 1996).

Additionally, the nine *hrc* genes widely conserved among phytopathogenic bacterial species are also present in the genome of *E. psidii* (Alfano and Collmer 1997; Mor et al. 2001). According to the results obtained by Alfano and Collmer 1997, all these *hrc* genes are required for the constitution of the secretion system apparatus. The *hrcC* gene was previously described as responsible for encoding an outer membrane component of the T3SS (Preston et al. 1998), whereas, *hrcN, hrcR, hrcS, hrcT, hrcU* and *hrcV* encode inner membrane proteins, that compose the basal structure of the T3SS (Oh and Beer 2005).

The resemblance of the organization and composition of the predicted *hrp/hrc* cluster of *E. psidii* to that of *E. amylovora*, raises the question of whether *E. psidii* belongs to the group I of *hrp* clusters, together with *E. amylovora, P. syringae* and other species (Alfano and Collmer 1997). One particular characteristic of this group is the presence of the *hrpC* operon. Since this gene was found in the *E. psidii* cluster, a relationship of the molecular mechanisms by which *E. psidii* causes disease and those of species from group I is expected

(Deng et al. 1998). Additionally, the presence of a gene in the *E. psidii hrp/hrc* cluster sharing 86% identity with the *hrpL* gene of *E. amylovora*, indicates that *E. psidii* cluster might be regulated by HrpL, similarly to the bacterial species belonging to group I.

Here, we present the genetic organization of the *hrp/hrc* cluster of *E. psidii*, that can be responsible for coding proteins for assembly of the T3SS of this species. Considering the importance of this cluster for *E. amylovora* (Nissinen et al. 2007), *Pantoea agglomerans* pv. *gypsophilae* (Mor et al. 2001), *P. carotovorum* subsp. *carotovorum* (Lehtimäki et al. 2003) and *Pseudomonas syringae* (Deng et al. 1998), it can be assumed that the *hrp/hrc* cluster can also play an important role for the pathogenicity of *E. psidii*. However, further analysis involving functional characterization and mutant constructions will be necessary to better understand the importance of each gene for the pathogenicity of this species.

For a better understanding on how the pathogenicity of *E. psidii* can rely on the T3SS, it was fundamental to obtain an inventory of the type III secreted effectors. Thus, the putative type III secretome of *E. psidii* was predicted. Most of the putative T3SE detected in the genomes of all *E. psidii* strains included in this study were previously characterized as T3SE in other bacterial species (Guo et al. 2016; Jin et al. 2002; Barny 1995; Lewis et al. 2011; Castiblanco, Triplett, and Sundin 2018). For instance, HopW1 of *Pseudomonas syringae* is known for disrupting the actin cytoskeleton (Büttner 2016). Several cell biological plant processes depend on this structure, such as trafficking of some vacuolar proteins and endocytosis, which in turn are important for plant immune signaling. Hence, the effector HopW1 can interfere on plant immune responses (Kang et al. 2014); Eop1, a T3SE of *E. amylovora*, was described as a host limiting factor for this bacterium, due to its avirulence function on non-host plants (Asselin et al. 2011); HsvB and HsvC (*hrp*-associated systemic virulence) are effector proteins required for the systemic spread of *E. amylovora* in apple trees (Oh and Beer 2005; Shanker et al. 2017); YopN from *Yersinia* spp. is known to be required for the translocation of the effector YopE (Bogdanove et al. 1996); and finally, HopE1, a protein found in the genome of *E. psidii* strains IBSBF435, LPF640 and LPF681, was previously reported in *P. syringae* (Büttner 2016). This protein was characterized as an inhibitor of the plant cell wall-based immune mechanisms,

which was associated with callose deposition and secretion of pathogenesis-related proteins, in order to protect plant cells against pathogens (Guo et al. 2016).

Regarding the newly identified effector candidates denominated as Eps, Eps1 protein was assigned as a mannosyl-3-phosphoglycerate phosphatase, and Eps3 as a biphenyl 2,3-dioxygenase. No reports were found about the potential function of these two enzymatic activities during plant-bacteria interactions. Eps2 was characterized as a glycosyl transferase. A previous study reported that this enzyme is required for biofilm formation, as well as for extracellular polysaccharide (EPS) and lipopolysaccharide (LPS) biosynthesis in *Xanthomonas citri* subsp. *citri (Xac)* (Li and Wang 2012). The authors of that study demonstrated that the function of this enzyme is essential for the full virulence of *Xac* on host plants (Li and Wang 2012). Eps5, was annotated as an amidinotransferase enzyme. In a previous study, an enzyme with the same function was described as a virulence factor of *E. amylovora*, that contributed to the bacterium systemic infection in apple (Shanker et al. 2017). Thus, it is intriguing whether these two putative effectors, Eps2 and Eps5, could play a similar role in the pathogenicity of *E. psidii*. As for Eps4, Eps6 and Eps7, since they were annotated as uncharacterized proteins, no assumptions can be made about their putative function during the interaction of *E. psidii* with its host plants.

As a result of our genome mining for effector candidates, we found genes whose products have previously been characterized as accessory that aid in effector delivery. These included HrpA, that as mentioned above is essential for Hrp pilus formation (Kim, Wei, and Beer 1997); HrpN is a harpin secreted via the T3SS, responsible for eliciting hypersensitive responses (HR) on plants (Wei et al. 1992; Bogdanove et al. 1996), but that is also necessary for the pathogenicity of *E. amylovora* on pear (Wei et al. 1992). It is currently believed that HrpN is an accessory protein in the delivery of effectors, since it was observed that it was required for the type III translocation of DspA/E, which is an essential effector for the pathogenicity of *E. amylovora* (Bocsanczy et al. 2008; Bogdanove, Bauer, and Beer 1998); HrpK was characterized as a protein secreted through the T3SS of *E. amylovora*, but mutations in HrpK have no effect on the virulence or on the elicitation of HR responses on plants (Oh and Beer 2005; Bocsanczy et al. 2008). Thus, its function remains undetermined (Oh and Beer 2005; Nissinen et al.

2007); HrpW is a harpin secreted by the T3SS (Nissinen et al. 2007), but like HrpK, it is not required for the full virulence of *E. amylovora* (Bocsanczy et al. 2008);

The WolfPsort algorithm provided some insights into the subcellular localization of the *E. psidii* effector candidates within host cells. Some of the novel effector candidates were predicted to carry signals for multiple localization within plant cells, which is in agreement with the knowledge that some effector proteins play functions in more than one subcellular compartment (Yang et al. 2010).

Many bioinformatic approaches have been applied for the prediction of T3SE inventories of different bacterial species (Vinatzer, Jelenska, and Greenberg 2007; Schechter et al. 2007; Vencato et al. 2006; Guzmán-Guzmán et al. 2017). In this study, a *hrp*-box promoter search was used for the prediction of effector genes in the genomes of *E. psidii* strains. *hrp*-box sequences similar to the ones located upstream of several *hrp* genes in *P. syringae* and *E. amylovora* were identified on the genome of *E. psidii*, indicating that the predicted genes might as well be regulated by hrpL and are candidate effector proteins (Innes et al. 1993; Vencato et al. 2006). This prediction was extremely important to further our understanding of the mechanisms behind the pathogenicity of *E. psidii* towards its host plants. We demonstrated that the bacterium possesses an intact *hrp/hrc* gene cluster that is associated with at least twelve conserved and novel effector genes, strongly suggesting that the T3SS plays an important role during its interaction with plants.

Here, we also present the first prediction of the type II effector repertoire of *E. psidii*. Proteins that play important roles in the pathogenicity of other bacterial species, such as proteins that possess hydrolytic and non-hydrolytic activities are secreted by the T2SS. The present study represents an important advance to the purpose of understanding why *E. psidii* is capable of causing disease on its host plants (eucalypt and guava), and more specifically, what is the contribution of the Type II and III secretion system during bacteria-plant interaction.

ACKNOWLEDGEMENTS

The authors would like to thank Franciso Henrique Nunes da Silva Alves, Jorge Luis Badel Pacheco and José Cleydson Ferreira da Silva for the availability of the phyton script; and CAPES for the Master's degree scholarship provided to Isadora Cristófoli Pereira and for partially funding this research - Finance Code 001.

REFERENCES

Almagro Armenteros, J. J.; Tsirigos, K. D.; Sønderby, C. K.; Petersen, T. N.; Winther, O.; Brunak, S. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37:420–423.

Alfano, J. R.; Collmer, A. 1997. The type III (Hrp) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins and death. J. Bacteriol. 179:5655–5662.

Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J.; Pennsylvania, T. 1990. Basic Local Alignment Search Tool. J. Mol. Biol. 215:403–410.

Araujo, F. A.; Barh, D.; Silva, A.; Guimarães, L.; Ramos, R. T. J. 2018. GO FEAT: A rapid web-based functional annotation tool for genomic and transcriptomic data. Sci. Rep. 8:8–11.

Arndt, D.; Grant, J. R.; Marcu, A.; Sajed, T.; Pon, A.; Liang, Y. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44:16–21.

Arriel, D. A. A.; Fonseca, N. R.; Guimarães, L. M. S.; Hermenegildo, P. S.; Mafia, R. G.; Borges Júnior, N.; de Souza, H. P. Alfenas, A. C. 2014. Wilt and die-back of *Eucalyptus* spp. caused by *Erwinia psidii* in Brazil. For. Pathol. 44:255–265.

Asselin, J. E.; Bonasera, J. M.; Kim, J. F.; Oh, C.; Beer, S. V. 2011. Eop1 from a *Rubus* strain of *Erwinia amylovora* functions as a host-range limiting Factor. Phytopatology 101:935–944.

Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A. A.; Dvorkin, M.; Kulikov, A. S.; Lesin, V. M.; Nikolenko, S. I.; Pham, S.; Prjibelski, A. D.; Pyshkin, A. V.;

Sirotkin, A. V.; Vyahhi, N.; Tesler, G.; Alekseyev, M. A.; Pevzner, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455–477.

Bardou, P.; Mariette, J.; Escudié, F.; Djemiel, C.; Klopp, C. 2014. jvenn: an interactive Venn diagram viewer. BMC Bioinformatics. 15:1–7.

Barny, M. A. 1995. *Erwinia amylovora hrpN* mutants, blocked in harpin synthesis, express a reduced virulence on host plants and elicit variable hypersensitive reactions on tobacco. Eur. J. Plant Pathol. 101:333–340.

Barras, F.; van Gijsegem, F.; Chatterjee, A. K. 2014. Extracellular Enzymes and pathogenesis of soft-rot *Erwinia*. Annu. Rev. Phytopathol. 32:478–485.

Bocsanczy, A. M.; Nissinen, R. M.; OH, C. H.; Beer, S. V. 2008. HrpN of *Erwinia amylovora* functions in the translocation of DspA/E into plant cells. Mol. Plant. Pathol. 9:425-434.

Bogdanove, A. J.; Bauer, D. W.; Beer, S. V. 1998. *Erwinia amylovora* secretes DspE, a pathogenicity factor and functional AvrE homolog, through the Hrp (type III secretion) pathway. J. Bacteriol. 180:2244–2247.

Bogdanove, A. J.; Wei, Z. M.; Zhao, L.; Beer, S. V. 1996. *Erwinia amylovora* secretes harpin via a type III pathway and contains a homolog of *yopN* of *Yersinia spp*. J. Bacteriol. 178:1720–1730.

Büttner, D. 2016. Behind the lines-actions of bacterial type III effector proteins in plant cells. FEMS Microbiol. Rev. 40:894–937.

Castiblanco, L. F.; Triplett, L. R.; Sundin, G. W. 2018. Regulation of effector delivery by type III secretion chaperone proteins in *Erwinia amylovora*. Front. Microbiol. 9:1–10.

Charkowski, A.; Blanco, C.; Condemine, G.; Expert, D.; Franza, T.; Hayes, C.; Hugouvieux-Cotte-Pattat, N.; Solanilla, L. E.; Low, D.; Moleleki, L.; Pirhogen, M.; Pitman, A.; Perna, N.; Reverchon, S.; Palenzuela, R. P.; Francisco, M. S.; Toth, I.; Tsuyumy, S.; Waals, V. D. J.; Wolf, V. D. J.; Gijsegem, V. F.; Yang, C. H.; Yedidia, I. 2012. The role of secretion systems and small molecules in soft-rot Enterobacteriaceae pathogenicity. Annu. Rev.

Phytopathol. 50: 425-449.

Chen, S.; Huang, T.; Zhou, Y.; Han, Y.; Xu, M.; Gu, J. 2017. AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. BMC Bioinformatics 18:91-100.

Cornelis, G. R.; Gijsegem, V. F. Assembly and function of type III secretory systems. 2000. Annu. Rev. Microbiol. 54:735-774.

Costa, T. R.; Felisberto-Rodrigues, C.; Meir, A.; Prevost, M. S.; Redzej, A.; Trokter, M.; Waksman, G. 2015. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nat. Rev. Microbiol. 6:343-59.

Coutinho, T. A.; Brady, C. L.; Vaart, M.; Venter, S. N.; Telechea, N.; Rolfo, M.; Perez, C.; Wingfield, M. J. 2011. A new shoot and stem disease of eucalyptus species caused by *Erwinia psidii*. Australas. Plant Pathol. 40:55–60.

Dalio, R. J. D.; Herlihy, J., O.; McDowell, T. S.; Machado, J. M. 2017. Effector biology in focus: a primer for computational prediction and functional characterization. Mol. Plant-Microbe Interact. 31:22–33.

Darling, A. E.; Mau, B.; Perna, N. T. 2010. Progressivemauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 5.(6): e11147.

Deng, W. L.; Preston, G.; Collmer, A.; Chang, C. J.; Huang, H. C. 1998. Characterization of the *hrpC* and *hrpRS* operons of *Pseudomonas syringae* pathovars *syringae*, *tomato*, and *glycinea* and analysis of the ability of *hrpF*, *hrpG, hrcC, hrpT*, and *hrpV* mutants to elicit the hypersensitive response and disease in plants. J. Bacteriol. 180:4523–4531.

Ferraz, H. G. M.; Demuner, G. A.; Guimarães, L. M. S.; Arriel, D. A. A.; da Silva, A. C.; Junior, N. B. 2016. Methods of inoculation and evaluation of *Erwinia psidii* in eucalypt. For. Pathol. 46:240–247.

Frederick, R. D.; Ahmad, M.; Majerczak, D. R.; Arroyo-Rodríguez, A. S.; Manulis, S.; Coplin, D. L. 2001. Genetic organization of the *Pantoea stewartii* subsp. *stewartii hrp* gene cluster and sequence analysis of the *hrpA, hrpC, hrpN*, and *wtsE* operons. Mol. Plant. Microbe. Interact. 14:1213–22.

Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. 2012. CD-HIT: Accelerated for clustering

the next-generation sequencing data. Bioinformatics. 28:3150–3152.

Gaudriault, S.; Malandrin, L.; Paulin, J. P.; Barny, M. A. 1997. DspA, an essential pathogenicity factor of *Erwinia amylovora* showing homology with AvrE of *Pseudomonas syringae* is secreted via the Hrp secretion pathway in a DspB-dependent way. Mol. Microbiol. 26:1057–1069.

Grant, S. R.; Fisher, E. J.; Chang, J. H.; Mole, B. M.; Dangl, J. L. 2006. Subterfuge and manipulation: type iii effector proteins of phytopathogenic bacteria. Annu. Rev. Microbiol. 60:425–449.

Green, E. R.; Mecsas, J. 2016. Bacterial secretion systems – an overview. Microbiol. Spectr. 4:1–32.

Guo, M.; Kim, P.; Li, G.; Elowsky, C. G.; Alfano, J. R. 2016. A bacterial effector co-opts calmodulin to target the plant microtubule network. Cell Host Microbe. 19:67–78.

Guzmán-Guzmán, P.; Alemán-Duarte, M. I.; Delaye, L.; Herrera-Estrella, A.; Olmedo-Monfil, V. 2017. Identification of effector-like proteins in *Trichoderma* spp. and role of a hydrophobin in the plant-fungus interaction and mycoparasitism. BMC Genetics. 18:2–20.

Hermenegildo, S.; Santos, S. A.; Guimarães, L. M. S.; Pereira, I. C.; Vidigal, P. M. P.; Badel, J. L.; Alfenas-Zerbini, P.; Mafia, G. R.; Ferreira, A. S. V. M.; Alfenas, A. C. 2019. Draft genome sequence of *Erwinia psidii*, causal agent of Bacterial Blight of Guava (*Psidium guava*) and Dieback of Eucalypt (*Eucalyptus* spp.) Microbiol Resour Announc. 8:30–32.

Horton, P.; Park, K. J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C. J. 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 35:585–587.

Huang, Y.; Niu, B.; Gao, Y.; Fu, L.; Li, W. 2010. CD-HIT Suite: A web server for clustering and comparing biological sequences. Bioinformatics. 26:680–682.

Innes, R. W.; Bent, A. F.; Kunkel, B. N.; Bisgrove, S. R.; Staskawicz, B. J. 1993. Molecular analysis of avirulence gene *avrRpt2* and identification of a putative regulatory sequence common to all known *Pseudomonas syringae*

avirulence genes. J. Bacteriol. 175:4859–4869.

Jin, Q.; Hu, W.; McGhee, G. C.; Hart, P.; Jones, A. L.; He, S. Y. 2002. Secretion of Hrp and Dsp proteins via the Hrp pilus during type III secretion in *Erwinia amylovora*. Acta Hortic. 590:441–448.

Jha, G.; Rajeshwari, R.; Sonti, R. V. 2005. Bacterial type two secretion system secreted proteins: double-edged swords for plant pathogens. Mol. Plant-Microbe Interact. 18:891–898.

Johnson, T. L.; Abendroth, J.; Hol, W. G. J.; Sandkvist, M. 2006. Type II secretion: from structure to function. FEMS Microbiol. Lett. 255:175–186.

Jones, D. G. J.; Dangl, L. J. 2006. The plant immune system. Nature. 444:323-329.

Kang, Y.; Jelenska, J.; Cecchini, N. M.; Li, Y.; Lee, M. W.; Kovar, D. R. 2014. HopW1 from *Pseudomonas syringae* disrupts the actin cytoskeleton to promote virulence in *Arabidopsis*. PLoS Pathog. 10:1–10.

Kazemi-Pour, N.; Condemine, G.; Hugouvieux-Cotte-Pattat, N. 2004. The secretome of the plant pathogenic bacterium *Erwinia chrysanthemi*. Proteomics. 4:3177–3186.

Kim, J. F.; Wei, Z.; Beer, S. V. 1997. The *hrpA* and *hrpC* operons of *Erwinia amylovora* encode components of a type III pathway that secretes harpin. J. of Bacteriol. 179:1690–1697.

Korotkov, K. V.; Sandkvist, M.; Hol, W. G. J. 2012. The type II secretion system: biogenesis, molecular architecture and mechanism. Nat. Rev. Microbiol. 10:336–351.

Krogh, A.; Larsson, B.; Von Heijne, G.; Sonnhammer, E. L. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305:567–580.

Lehtimäki, S.; Rantakari, A.; Routtu, J.; Tuikkala, A.; Li, J.; Virtaharju, O. 2003. Characterization of the *hrp* pathogenicity cluster of *Erwinia carotovora* subsp. *carotovora*: high basal level expression in a mutant is associated with reduced virulence. Mol. Genet. Genomics. 270:263–272.

Lewis, J. D.; Lee, A.; Ma, W.; Zhou, H.; Guttman, D. S.; Desveaux, D. 2011. The YopJ superfamily in plant-associated bacteria. Mol. Plant Pathol. 12:928–937.

Li, J.; Wang, N. 2012. The *gpsX* gene encoding a glycosyltransferase is important for polysaccharide production and required for full virulence in *Xanthomonas citri* subsp. *citri*. BMC Microbiol. 12:31.

Lindgren, P. B. 1997. The role of *hrp* genes during plant-bacterial interactions. Annu. Rev. Phytopathol. 35:129–152.

Montoya-Estrada, C. N.; Costa, C. R.; Badel, J. L.; Guimarães, L. M. S.; Alfenas, A. C. 2018. Root infection and aerial colonization of eucalypt host plants by *Erwinia psidii*. Trop. Plant Pathol. 44:251–257.

Mor, H.; Manulis, S.; Zuck, M.; Nizan, R.; Coplin, D. L.; Barash, I. 2001. Genetic organization of the *hrp* gene cluster and *dspAE/BF* operon in *Erwinia herbicola* pv. *gypsophilae*. Mol. Plant-Microbe Interact. 14:431–436.

Mukherjee, S.; Keitany, G.; Li, Y.; Wang, Y.; Ball, H.; Goldsmith, E. 2006. *Yersinia* YopJ acetylates and inhibits kinase activation by blocking phosporylation. Science. 80:1211–1214.

Nissinen, R.; Ytterberg, A. J.; Institutet, K.; Wijk, K. J.; Beer, S. V. 2007. Analyses of the secretomes of *Erwinia amylovora* and selected *hrp* mutants reveal novel type III secreted proteins and an effect of HrpJ on extracellular harpin levels. Mol. Plant Pathol. 8:55–67.

Nivaskumar, M., and Francetic, O. 2014. Type II secretion system: a magic beanstalk or a protein escalator. Biochim. Biophys. Acta. 1843:1568–1577.

Oh, C. S.; Beer, S. V. 2005. Molecular genetics of *Erwinia amylovora* involved in the development of fire blight. FEMS Microbiol. Lett. 253:185–192.

Preston, G.; Deng, W. L.; Huang, H. C.; Collmer, A. 1998. Negative regulation of *hrp* genes in *Pseudomonas syringae* by HrpV. J. Bacteriol. 180:4532–4537.

Rissman, A. I.; Mau, B.; Biehl, B. S.; Darling, A. E.; Glasner, J. D.; Perna, N. T. 2009. Reordering contigs of draft genomes using the Mauve Aligner. Bioinformatics 25:2071–2073.

Rodrigues Neto, J.; Robbs, C. F.; Yamashiro, T. 1987. A bacterial disease of guava (*Psidium guajava*) caused by *Erwinia psidii* sp. nov. Fitopatol. Bras. 12:345-350.

Roine, E.; Wensheng, W.; Yuan, J.; Nurmiaho-Lassila, E. L.; Kalkkinen, N.; Romantschuk, M.; He, Y. S. 1997. Hrp pilus: an *hrp*-dependent bacterial surface appendage produced by *Pseudomonas syringae* pv. *tomato* DC300. Plant Biol. 94:3459-3464.

Sandkvist, M. 2001. Type II secretion and pathogenesis. Infect. Immun. 69:3523–3535.

Schechter, L. M.; Vencato, M.; Jordan, K. L.; Schneider, S. E.; Schneider, D. J.; Collmer, A. 2007. Multiple approaches to a complete inventory of *Pseudomonas syringae* pv. *tomato* DC3000 Type III secretion system effector proteins. Mol. Plant-Microbe Interact. 19:1180–1192.

Shen, H.; Keen, N. T. 1993. Characterization of the promoter of avirulence gene D from *Pseudomonas syringae* pv. *tomato*. J. Bacteriol. 175:5916–24.

Shanker, S.; Schaefer, G. K.; Barnhart, B. K.; Wallace-Kneale, V. L.; Chang, D.; Coyle, T. J. 2017. The virulence-associated protein HsvA from the fire blight pathogen *Erwinia amylovora* is a polyamine amidinotransferase. J. Biol. Chem. 292:21366–21380.

Sperschneider, J.; Williams, A. H.; Hane, J. K.; Singh, K. B.; Taylor, J. M. 2015. Evaluation of secretion prediction highlights differing approaches needed for oomycete and fungal effectors. Front. Plant Sci. 6:1–14.

Vencato, M.; Alfano, J. R.; Buell, C. R.; Cartinhour, S. 2006. Bioinformatics-enabled identification of the HrpL regulon and type III secretion system effector proteins of *Pseudomonas syringae* pv. *phaseolicola* 1448A. Mol. Plant-Microbe Interact. 19:1193–1206.

Vinatzer, B. A.; Jelenska, J.; Greenberg, J. T. 2007. Bioinformatics correctly identifies many type III secretion substrates in the plant pathogen *Pseudomonas syringae* and the biocontrol isolate *P. fluorescens* SBW25. Mol. Plant-Microbe Interact. 18:877–888.

Voulhoux, R.; Ball, G.; Ize, B.; Vasil, M. L.; Lazdunski, A. Â. E.; Wu, L. 2001. Involvement of the twin-arginine translocation system in protein secretion via the type II pathway. The EMBO J. 20:6735–6741.

Wei, Z.; Beer, S. V. 1993. HrpI of *Erwinia amylovora* functions in secretion of harpin and is a member of a new protein family. J. Bacteriol. 175:7958–7967.

Wei, Z.; Beer, S. V. 1995. hrpL activates *Erwinia amylovora hrp* gene transcription and is a member of the ECF subfamily of σ Factors. J. Bacteriol. 177:6201–6210.

Wei, Z.; Kim, J. F.; Beer, S. V. 2000. Regulation of *hrp* genes and Type III protein secretion in *Erwinia amylovora* by HrpX/HrpY, a novel two-component system, and HrpS. Mol. Plant-Microbe Interact. 13:1251–1262.

Wei, Z.; Laby, R. J.; Zumoff, C. H.; Bauer, D. W.; He, S. Y.; Colimer, A. 1992. Harpin, elicitor of the hypersensitive response produced by the plant pathogen *Erwinia amylovora*. Science. 257:85–88.

Xiao, Y.; Hutcheson, S. W. 1994. A single promoter sequence recognized by a newly identified alternate sigma factor directs expression of pathogenicity and host range determinants in *Pseudomonas syringae*. J. Bacteriol. 176:3089–3091.

Yang, Y.; Zhao, J.; Morgan, R. L.; Ma, W.; Jiang, T. 2010. Computational prediction of type III secreted proteins from gram-negative bacteria. BMC Bioinformatics 11:1–10.

Table 1. *Erwinia psidii* strains used in this study. Adapted from Montoya-Estrada et al. (2018).

Strain	**Host**	**Geographic origin**	**Aggressiveness on *E. urophylla* × *E. grandis* (CLR375)**	**Aggressiveness on *E. urophylla* (CLR440)**
IBSBF435*	*Psidium guajava*	Brazil: Sâo Paulo, Valinhos	Highly Aggressive	Highly aggressive
LPF681	*P. guajava*	Brazil: São Paulo, Vista Alegre do Alto	Weakly aggressive	Non virulent
LPF534	*Eucalyptus dunnii*	Brazil: Rio Grande do Sul, Guaíba	Aggressive	Aggressive
LPF640	*E. urophylla*	Brazil: São Paulo, Conceição	Aggressive	Weakly aggressive

* Type-strain

Table 2. Strains used for bioinformatic characterization of the *hrp/hrc* gene cluster of *Erwinia psidii.*

Species	Strain	GenBank accession	References
Erwinia psidii	IBSBF435	GCA_003846135.1	Hermenegildo et al. 2019
Erwinia amylovora	CFBP 1430	NC_013961.1	Alfano and Collmer 1997 Wei and Beer 1995 Wei, Kim and Beer 2000
Pectobacterium carotovorum subsp. *carotovorum*	SCC1	AY293288.1	Lehtimäki et al. 2003
Pantoea stewartii subsp. *stewartii*	DC283	AF282857.2	Frederick et al. 2001

Table 3. Conserved *hrp* box sequences used for prediction of the type III secretome of *Erwinia psidii*.

Species	**Gene**	***hrp* box sequence**	**References**
Pseudomonas syringae pv. *tomato*	*avrD*	GGAACC-$N_{15/16}$-CCAC	(Shen and Keen 1993)
P. syringae pv. *syringae*	*hrpZ*	GGAACC-N_{16}- CCACC	(Xiao and Hutcheson 1994)
P. syringae pv. *tomato*	*avrRpt2*	GGAACCNA-N_{14}-CCAC-N_2-A	(Innes et al. 1993)
Erwinia amylovora	*hrpN*	GGAACC-N_{16}-CACTC	(Wei and Beer 1995)
E. amylovora	*hrpA*	GGGAACC-N_{15}-CCACT-N_1-AAT	(Kim, Wei, and Beer 1997)
E. amylovora	*hrpC*	GGGAAC-N_{17}-CCACTCAA	(Kim, Wei, and Beer 1997)
E. amylovora	*dspE*	GGAACC-N_{15}-CAACA	(Gaudriault et al. 1997)

N: means any nucleotide.

Table 4. Predicted type III secreted proteins of *Erwinia psidii* strains.

Strain	Strand	Protein id	Annotation	Effector	Similar to proteins of	E-value
IBSBF435	Forward	WP_124232300.1	Hypothetical protein	Eps1		
	Forward	WP_124232301.1	Hypothetical protein	Eps2		
	Forward	WP_124231893.1	Hypothetical protein	Eps3		
	Forward	WP_124231894.1	Hypothetical protein	Eps4		
	Reverse	WP_124234645.1	Hypothetical protein	Eps5		
	Reverse	WP_124232053.1	Hypothetical protein	Eps6		
	Reverse	WP_124231359.1	Hypothetical protein	Eps7		
	Reverse	WP_124231866.1	Type III secretion system effector EOP1	EOP1	*Erwinia amylovora*	5E-173
	Reverse	WP_124232323.1	Type III effector HopE1	HopE1	*Erwinia tracheiphila* PSU-1	4E-130
	Reverse	WP_124234642.1	Type III effector HopW1	HopW1	*Pseudomonas syringae* group	0
	Reverse	WP_124234644.1	Hrp-associated systemic virulence protein HsvB	HsvB	*Erwinia pyrifoliae* DSM 12163	0
	Reverse	WP_124234643.1	Hrp-associated systemic virulence protein HsvC	HsvC	*Erwinia amylovora* ATCC BAA-2158	0
	Forward	WP_124231882.1	YopN family type III secretion system gatekeeper subunit	YopN	*Erwinia psidii* IBSBF435	0
LPF534	Reverse	PRJNA498492:EHW64_01620	Hypothetical protein	Eps1		
	Reverse	PRJNA498492:EHW64_01615	Hypothetical protein	Eps2		
	Reverse	PRJNA498492:EHW64_05540	Hypothetical protein	Eps3		
	Reverse	PRJNA498492:EHW64_05535	Hypothetical protein	Eps4		
	Reverse	PRJNA498492:EHW64_18375	Hypothetical protein	Eps5		
	Forward	PRJNA498492:EHW64_05680	Type III secretion system effector EOP1	EOP1	*Erwinia amylovora*	6E-173
	Reverse	PRJNA498492:EHW64_18360	Type III effector hopW1	HopW1	*Pseudomonas syringae* pv. *maculicola*	0
	Reverse	PRJNA498492:EHW64_18370	Hrp-associated systemic virulence protein HsvB	HsvB	*Erwinia pyrifoliae* DSM 12163	0
	Reverse	PRJNA498492:EHW64_18365	Hrp-associated systemic virulence protein HsvC	HsvC	*Erwinia amylovora* ATCC BAA-2158	0
	Reverse	PRJNA498492:EHW64_05595	YopN family type III secretion system gatekeeper subunit	YopN	*Erwinia psidii* IBSBF435	0

LPF640	Forward	PRJNA498492:EHW65_03230	Hypothetical protein	Eps1		
	Forward	PRJNA498492:EHW65_03235	Hypothetical protein	Eps2		
	Reverse	PRJNA498492:EHW65_10010	Hypothetical protein	Eps3		
	Reverse	PRJNA498492:EHW65_10005	Hypothetical protein	Eps4		
	Reverse	PRJNA498492:EHW65_16300	Hypothetical protein	Eps5		
	Forward	PRJNA498492:EHW65_10150	Type III secretion system effector EOP1	EOP1	*Erwinia amylovora*	6E-173
	Reverse	PRJNA498492:EHW65_01995	Type III effector HopE1	HopE1	*Erwinia tracheiphila* PSU-1	4E-130
	Reverse	PRJNA498492:EHW65_16285	Type III effector hopW1	HopW1	*Pseudomonas syringae* group	0
	Reverse	PRJNA498492:EHW65_16295	Hrp-associated systemic virulence protein HsvB	HsvB	*Erwinia pyrifoliae* DSM 12163	0
	Reverse	PRJNA498492:EHW65_16290	Hrp-associated systemic virulence protein HsvC	HsvC	*Erwinia amylovora* ATCC BAA-2158	0
	Reverse	PRJNA498492:EHW65_10065	YopN family type III secretion system gatekeeper subunit	YopN	*Erwinia psidii* IBSBF435	0
LPF681	Reverse	PRJNA498492:EHW66_00045	Hypothetical protein	Eps1		
	Reverse	PRJNA498492:EHW66_00040	Hypothetical protein	Eps2		
	Forward	PRJNA498492:EHW66_03495	Hypothetical protein	Eps3		
	Forward	PRJNA498492:EHW66_03500	Hypothetical protein	Eps4		
	Forward	PRJNA498492:EHW66_21010	Hypothetical protein	Eps5		
	Reverse	PRJNA498492:EHW66_03355	Type III secretion system effector EOP1	EOP1	*Erwinia amylovora*	2E-166
	Forward	PRJNA498492:EHW66_01245	Type III effector HopE1	HopE1	*Erwinia tracheiphila* PSU-1	4E-130
	Forward	PRJNA498492:EHW66_21025	Type III effector hopW1	HopW1	*Pseudomonas syringae* group	0
	Forward	PRJNA498492:EHW66_21015	Hrp-associated systemic virulence protein HsvB	HsvB	*Erwinia pyrifoliae* DSM 12163	0
	Forward	PRJNA498492:EHW66_21020	Hrp-associated systemic virulence protein HsvC	HsvC	*Erwinia amylovora* ATCC BAA-2158	0
	Forward	PRJNA498492:EHW66_03440	YopN family type III secretion system gatekeeper subunit	YopN	*Erwinia psidii* IBSBF435	0

Table 5. Subcellular localization of *Erwinia psidii* predicted type III effectors according to WolfPsort.

Effector	IBSBF435	LPF640	LPF534	LPF681
Eps1	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic
Eps2	Peroxisome	Peroxisome	Peroxisome	Peroxisome
Eps3	Multiple	Multiple	Multiple	Multiple
Eps4	Multiple	Multiple	Multiple	Multiple
Eps5	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic
Eps6	Nuclear	-	-	-
Eps7	Nuclear	-	-	-
EOP1	Mitochondrial	Mitochondrial	Mitochondrial	Multiple
HopE1	Cytoplasmic	Cytoplasmic	-	Cytoplasmic
HopW1	Nuclear	Nuclear	Nuclear	Nuclear
HsvB	Cytoplasmic	Cytoplasmic	Cytoplasmic	Cytoplasmic
HsvC	Chloroplast	Chloroplast	Chloroplast	Chloroplast
YopN	Multiple	Multiple	Multiple	Multiple

- Putative T3SE not present on the respective strain.

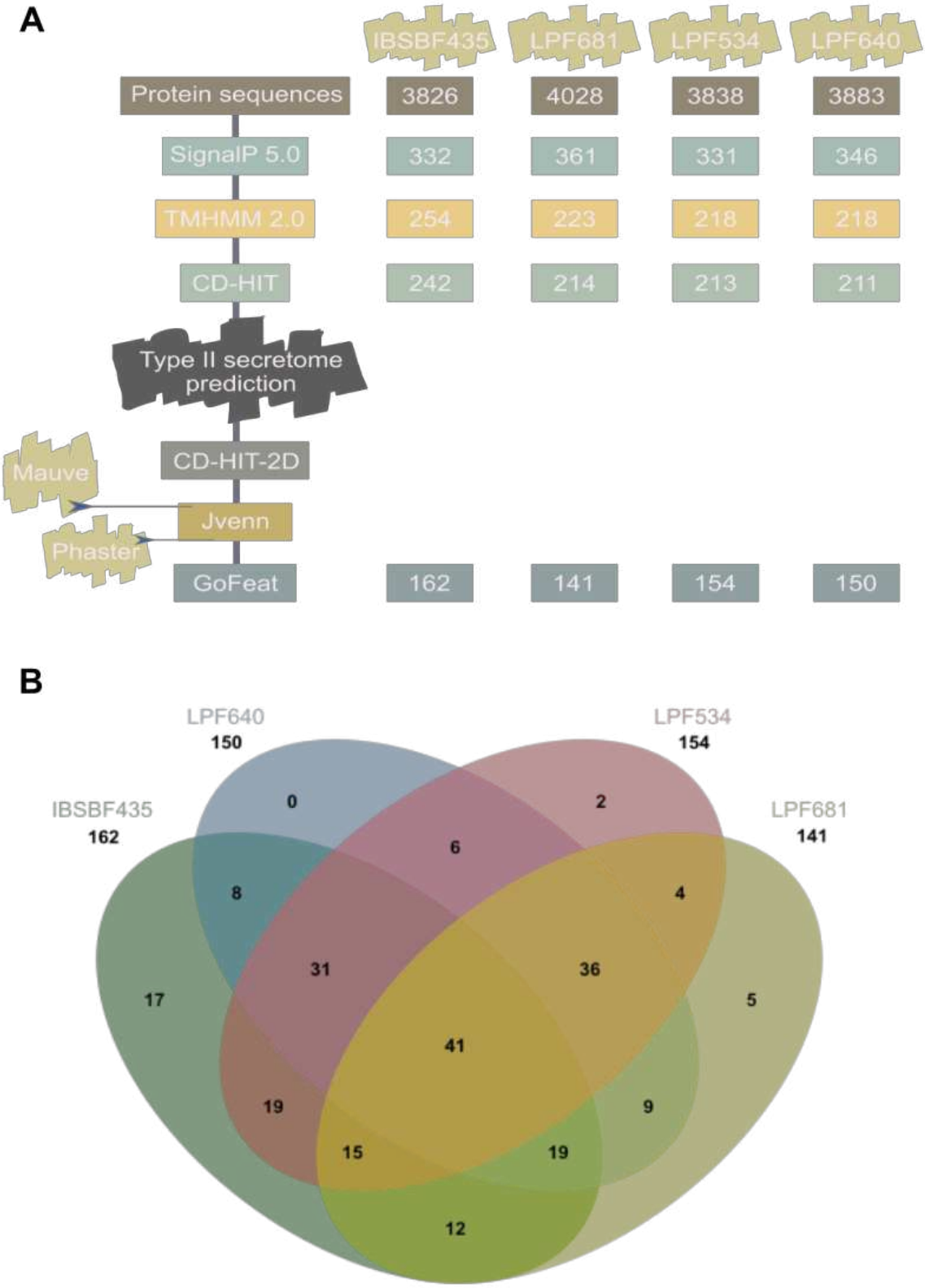


Figure 1. A. Bioinformatics workflow for prediction of the type II secretome of *Erwinia psidii* strains. Numbers on the right indicate the initial protein number to the remnant number after each step of the prediction process. **B.** Type II secretome of *E. psidii*. Strains IBSBF435 and LPF681 were isolated from *Eucalyptus* spp; strains LPF534 and LPF640, were isolated from *Psidium guajava.* The number below each strain identification shows the number of type II effector proteins predicted. Overlapped regions represent proteins shared between *E. psidii* strains.

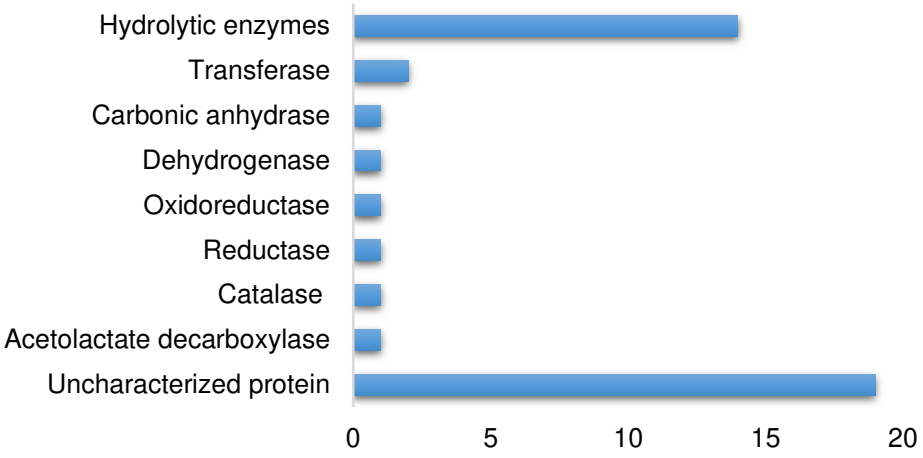


Figure 2. Functional enrichment obtained with GoFeat for 47 predicted type II proteins common to the four strains of *Erwinia psidii*.

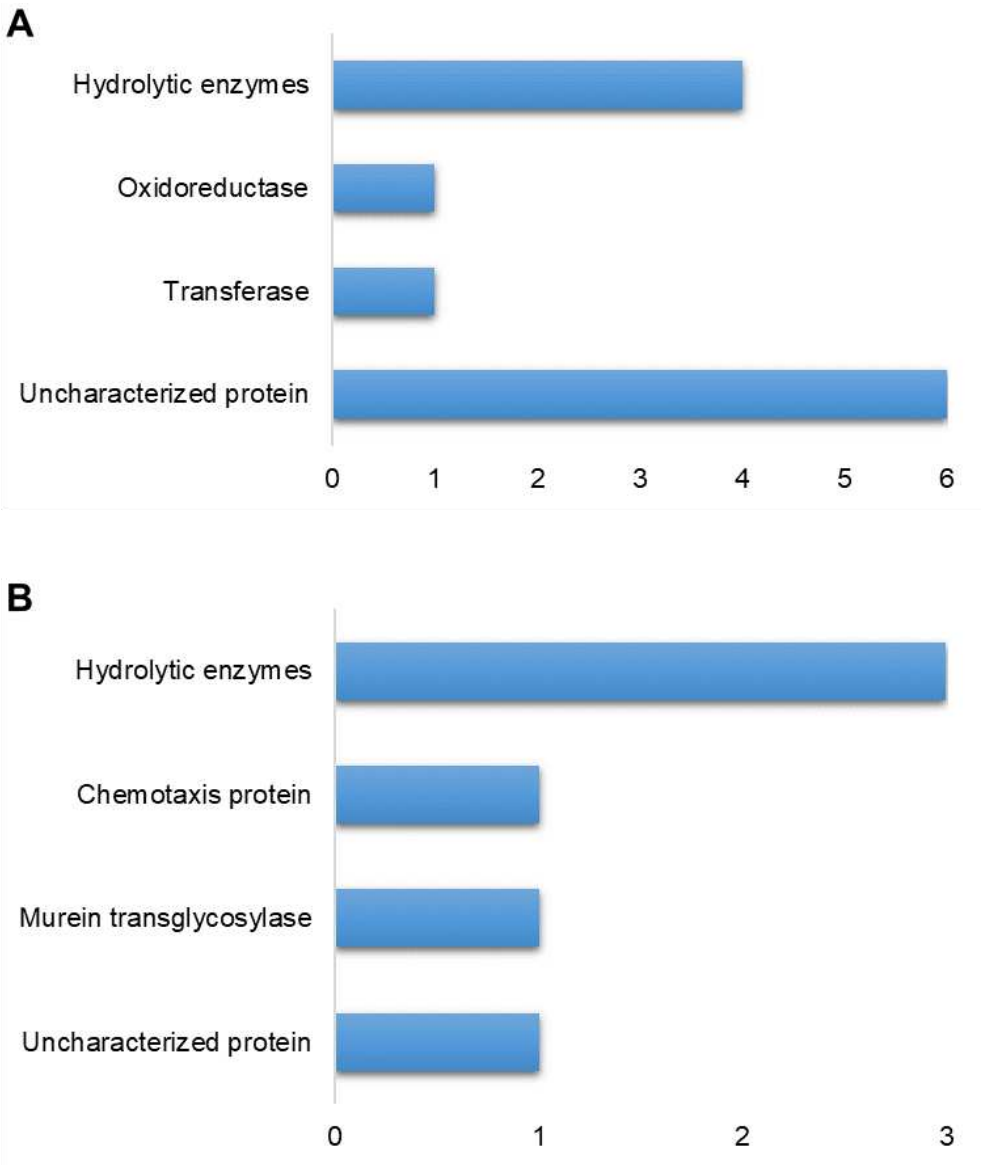


Figure 3. Functional enrichment obtained with GoFeat. **A.** Proteins common to *Erwinia psidii* IBSBF435 and LPF681, strains isolated from *Psidium guajava*. **B.** Proteins common to *E. psidii* LPF534 and LPF640, strains isolated from *Eucalypt* spp.

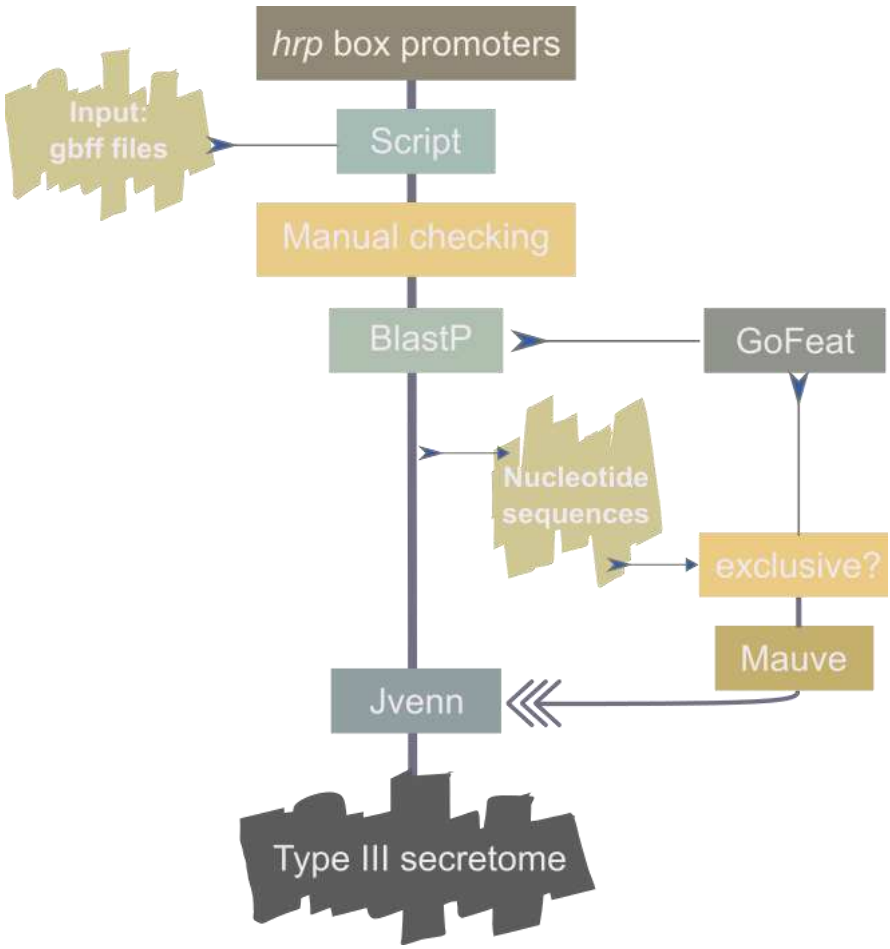


Figure 4. Bioinformatics workflow for prediction of the Type III secretome of *Erwinia psidii* strains.

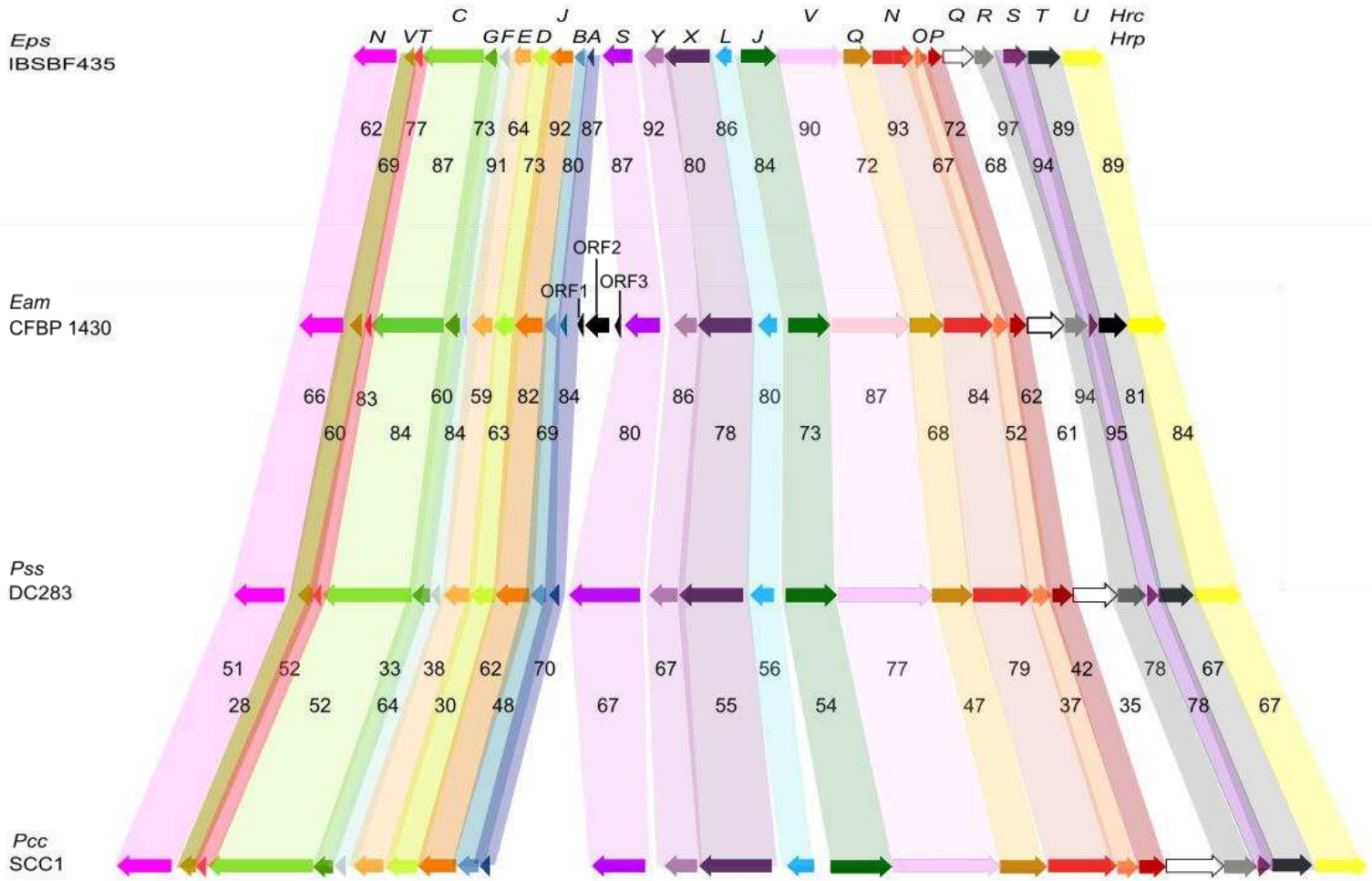


Figure 5. Characterization of the *hrp/hrc* gene cluster of *Erwinia psidii* (*Eps*; strain IBSBF435) and synteny comparison with the cluster of *Erwinia amylovora* (*Eam;* strain CFBP1430), *Pantoea stewartii* subsp. *stewartii* (*Pss*; strain DC283) and *Pectobacterium carotovorum* subsp. c*arotovorum* (*Pcc*; SCC1). Each gene is represented by one arrow and a specific color. Arrows direction indicate the orientation of transcription. The numbers above each species *hrp/hrc* cluster represents the percentage of identity shared by each protein with the predicted proteins from the putative cluster of *E. psidii*. Open reading frames (ORFs) not previously described on the cluster of *E. amylovora* were designated as ORF1, 2 and 3.

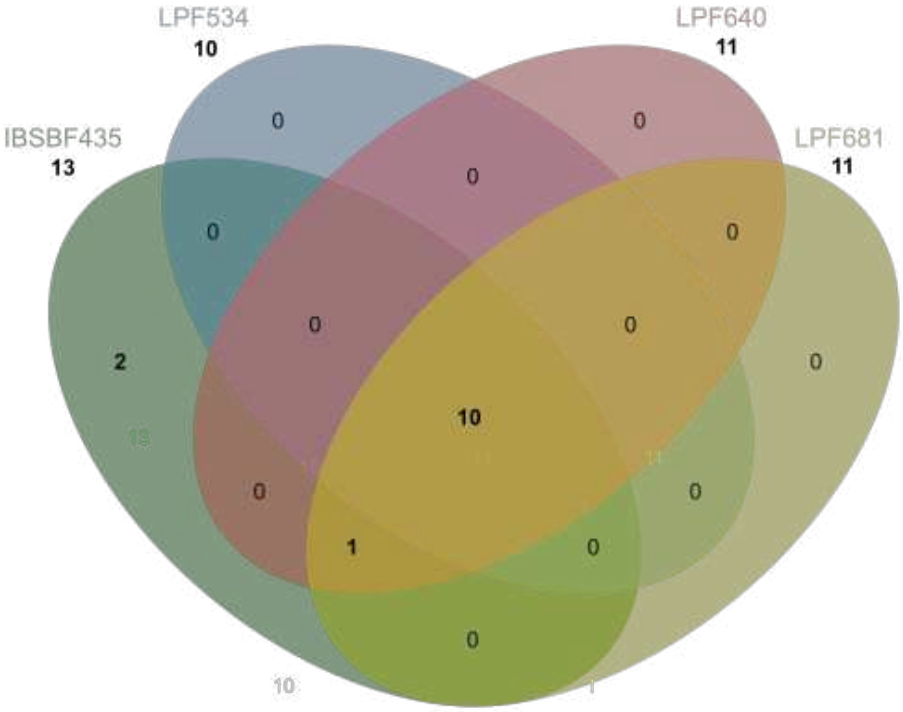


Figure 6. Predicted type III secretome of *Erwinia psidii*. Strains IBSBF435 and LPF681 were isolated from *Eucalyptus* spp.; LPF534 and LPF640, isolated from *Psidium guajava.* The number below each strain name represents the total number of putative type III secreted effectors identified. Overlapped regions represent proteins shared between *E. psidii* strains.

SUPPLEMENTARY TABLE

Supplementary Table S1. Functional characterization of exclusive type II secreted proteins from each *Erwinia psidii* strains.

Strain	Protein identification	Annotation
LPF534	EHW64_08205	Alginate lyase
	EHW64_06870	Hypothetical protein
LPF681	EHW66_00260	Hypothetical protein
	EHW66_13625	Sell repeat family protein
	EHW66_00695	DUF2511 domain-containing protein
	EHW66_14040	Hypothetical protein
	EHW66_2668	Hypothetical protein
IBSBF435	WP_124234817.1	DNA ligase B
	WP_124232785.1	Protease
	WP_124232064.1	Endopeptidase
	WP_124234647.1	Phytochelatin synthase
	WP_124234859.1	Lytic transglycosylase
	WP_124234837.1	Thiol:disulfide interchange protein
	WP_124234446.1	Uncharacterized protein
	WP_124234897.1	Uncharacterized protein
	WP_124234916.1	Uncharacterized protein
	WP_124232762.1	Uncharacterized protein
	WP_124232147.1	Uncharacterized protein
	WP_124234917.1	Uncharacterized protein
	WP_124234919.1	Uncharacterized protein
	WP_124232801.1	Uncharacterized protein

ARTICLE 2: Homology-based prediction and sequence evolution of putative type III effectors from *E. psidii*

PEREIRA, I. C.[1], VIDIGAL, P. M. P.[3], GUIMARÃES, L. M. S.[2], SANTOS, S. A.[2], ALFENAS, A. C.[2], BADEL, J. L.[1].

[1]Laboratory of Molecular Phytobacteriology; [2]Laboratory of Forest Pathology/Bioagro; Department of Plant Pathology, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil.
[3]NuBioMol; Center of Biological Sciences, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil.

ABSTRACT

Little is known about the pathogenicity of *Erwinia psidii* and how this pathogen evolved, to become able to infect *Eucalyptus s*pecies, besides its native host, *Psidium guajava*. In this study, we performed *in silico* predictions in order to identify putative type III effector proteins, that may play an important role on the bacterium pathogenicity. Through homology analysis, we found two proteins in the genome of *E. psidii* sharing identity with DspE and Eop1, both known important effector proteins secreted by the close realive *E. amylovora*. DspE is required by *E. amylovora* to cause disease, while Eop1 functions as an avirulence protein on non-host plants. Motif searches, subcellular localization prediction within host cells, as well as phylogenetic and selection analysis were conducted in order to better understand the evolution and function of these two putative *E. psidii* effectors. The results of these analysis represent the first effector genes homology-based prediction of *E. psidii*, a bacterium that causes important economical losses on both guava and eucalypt production.

Keywords: Eucalypt, guava, secretion, pathogenicity.

INTRODUCTION

Erwinia psidii was first reported in São Paulo state causing disease in guava trees (*Psidium guajava*) (Rodrigues Neto, Robbs and Yamashiro, 1987). A few years later, it was demonstrated that *E. psidii* was also the aetiologic agent responsible for the dieback of eucalypt, which is currently considered one of the most severe diseases of this woody plant in South America (Arriel et al. 2014).

Of particular interest for this study are the evolutionary processes underlying bacterial adaptation to infect different host species, considering the current proposal that *E. psidii* suffered a host shift to infect *Eucalyptus* spp. (Coutinho et al. 2011). Since little is known about the mechanisms utilized by *E. psidii* to cause disease on its host plants and how its pathogenicity genes evolved over time, an evolutionary analysis of its virulence genes is necessary to further our knowledge on the pathogenic potential of this bacterial species.

It is known that Type III effector proteins (T3SE) injected through the type III secretion system (T3SS) directly into the cytoplasm of its host plant cells (Grant et al. 2006; Lindgren 1997) are essential for the virulence of different bacterial species (Jin et al. 2016). Basically, T3SE have the capacity to suppress host defense responses to provide a better plant cell environment for bacterial growth (Grant et al. 2006; Jones and Dangl 2006).

Erwinia amylovora, type-species for the genus *Erwinia*, has the T3SS as one of its most important pathogenicity factors (Malnoy et al. 2012). Several T3SE are known to be secreted through the T3SS of *E. amylovora*, such as: Eop1, HopX1, Eop4 and DspA/E (Bogdanove, Bauer, and Beer 1998; Zhao, He, and Sundin 2006; Nissinen et al. 2007).

These T3SE exhibit different requirements for disease symptom development (Kamber 2013). For instance, Eop4 was reported as a virulence factor since it enhances bacterial infection and growth in host tissue (Zhao, He, and Sundin 2006). However, Eop4 as well as HopX1 and Eop1, were shown to behave as avirulence proteins on some non-host plants (Bocsanczy et al. 2012; Asselin et al. 2011; Lewis et al. 2011). Interestingly, Eop1 and HopX1 seem not to be required for the bacterial pathogenicity, since *E. amylovora's* virulence on

host plants was not compromised by the mutation of these genes (Asselin et al. 2011; Bocsanczy et al. 2012).

Eop1 is a member of the YopJ family of type III secreted proteins (Asselin et al. 2011; Castiblanco, Triplett, and Sundin 2018). Some members of this family are virulence factors, that inhibit the activation of host immune responses, such as the MAPK (mitogen-activated protein kinase) pathway (Asselin et al. 2011; Orth et al. 2000; Mukherjee et al. 2006). However, several others are characterized as avirulence proteins, based on their capability to elicit plant defense responses when expressed in resistant or non-host plants (Asselin et al. 2011). Eop1 belongs to the avirulence protein group (Lewis et al. 2011). Most YopJ-family members possess a catalytic triad composed of histidine, glutamate and cysteine (H/E/C) that is required for their function (Orth et al. 2000; Lewis et al. 2011).

The ability of *E. amylovora* to cause disease relies on DspA/E (disease-specific effector A/E) (Piqué et al. 2015; Siamer et al. 2014). Studies based on the interruption of *dsp* genes revealed that mutants became nonpathogenic or had their virulence severely attenuated (Bogdanove, Bauer, and Beer 1998; Boureau et al. 2006). DspA/E is a large protein (198 kDa) that belongs to the AvrE family of effectors (Jin et al. 2001; Siamer et al. 2014; Oh and Beer 2005), which are known for their ability to suppress the plant immune system and contribute to bacterial growth *in planta* (Piqué et al. 2015; Debroy et al. 2004).

All the effector genes aforementioned are adjacent to the *hrp* pathogenicity island (PAI) of *E. amylovora* (Aksoy, Kaya, and Hamid 2017; Bogdanove, Bauer, and Beer 1998; Asselin et al. 2011; Zhao, He, and Sundin 2006). These PAIs are genomic regions that not only contain virulence-associated genes, but that are also important for genetic variability, since they can be horizontally transferred between different species, thus favoring bacterial evolution (Blum et al. 1994; Hacker et al. 1997).

In addition to horizontal gene transfer, bacterial genetic variability can be caused by mutations derived from non-synomymous substitutions. (Stavrinides, McCann, and Guttman 2008). It is known that the interaction with the host immune system components imposes a strong selective pressure on T3SE, which promotes a greater genetic variability. Therefore, the study of the

processes that shape the evolution of T3SE could provide insights into the adaptation of bacterial pathogens to infect their hosts (Stavrinides, McCann, and Guttman 2008; Sokurenko, Hasty, and Dykhuizen 1999). In this study, we conducted evolutionary analyses of predicted *E. psidii* T3SE showing high similarity with *E. amylovora* effectors that play important roles during its interaction with host plants.

MATERIAL AND METHODS

Erwinia psidii genome sequences

Four *E. psidii* strains that showed differential aggressiveness on the eucalypt clones CLR440 (*E. urophylla)* and CLR375 (*E. urophylla × E. grandis)*, which are known to be susceptible to *E. psidii* (Montoya-Estrada et al. 2018), had their whole genome sequenced, assembled and annotated (Hermenegildo et al., 2019; Hermenegildo *et al.*, unpublished data). The sequencing was executed on an Illumina NovaSeq 6000 platform. Resulting reads showing high quality were *de novo* assembled into scaffolds using SPAdes version 3.10.1 (Bankevich et al. 2012) and their annotation was performed using NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/). Amino acid sequences were recovered from the annotation process.

All four *E. psidii* strains used in this study are described on Table 1. *E. psidii* IBSBF435, the type strain of the species, has its draft genome deposited on the NCBI database (https://www.ncbi.nlm.nih.gov/) under the accession number RHHM00000000 (Hermenegildo et al., 2019).

Prediction of effector genes

The type III effector proteins of *E. psidii* were identified through sequence similarity searches between genes and proteins annotated from *E. psidii* strains genomes against the NCBI database (https://www.ncbi.nlm.nih.gov/), using Blast

tool (Altschul et al. 1990). Amino acid and nucleotide sequences of the effector genes DspA/E, Eop1, HopX1 and Eop4, already reported for the model phytopathogen *E. amylovora* (strain ATCC 49946; GenBank accession number: NC_013971.1) were used as queries in the Blast searches (Bogdanove, Bauer, and Beer 1998; Zhao, Sara E Blumer, and Sundin 2005; Bocsanczy et al. 2012; Castiblanco, Triplett, and Sundin 2018).

Construction of effector and housekeeping gene databases

Amino acid and nucleotide sequences showing high similarity to the *E. psidii* predicted effectors were retrieved from the genome sequences of different phytopathogenic bacterial species by Blast searches against NCBI database (https://www.ncbi.nlm.nih.gov/). For this, the maximum number of hits was set to 250, the minimum coverage to 60% and the E-value to 10^{-5} in order to select sequences to compose the databases. Both nucleotide and protein databases were used for effector evolutionary analysis (Supplementary Table S1 and S2).

In order to compare the evolution of the effectors with the rest of the genome, the same procedure was followed for the construction of databases of the four housekeeping genes *atpD*, *gyrB*, *infB* and *rpoB*, that have been shown to be reliable markers for phylogenetic inference of the genus *Erwinia* (Coutinho et al. 2011; Liu et al. 2016; Moretti et al. 2010). In this case, preference was given to use strains that were also used in the databases of effector genes (Supplementary Table S3). The sequences of the housekeeping genes were used for multilocus sequence analysis (MLSA) (Glaeser and Kämpfer 2015).

Prediction of functional motifs

Potential functional motifs in AvrE and Eop1 protein sequences of *E. psidii* strains were predicted using the PsortII program (Nakai and Kanehisa 1992) as well as searches on Pfam (El-Gebali et al. 2019) and Prosite (de Castro et al. 2006) protein family databases. Also, the AvrE sequences of all four *E. psidii*

*s*trains were compared to other AvrE-family proteins, in order to verify the presence of the conserved leucine zipper (LZ), coiled-coil (CC) region, nuclear localization signal (NLS) and endoplasmic reticulum membrane retention signals (ERMRS) motifs present in most of the members of this family (Ham et al. 2007).

The amino acid sequences of Eop1 from *E. psidii* strains were aligned with previously reported YopJ family members to verify the presence of a H/E/C intact catalytic triad (Asselin et al. 2011). Sequence alignments were obtained with Kalign (https://www.ebi.ac.uk/Tools/msa/kalign/).

Prediction of effector proteins subcellular localization

Two tools from the Psort server (https://psort.hgc.jp/) were used for computational prediction of the subcellular localization of AvrE and Eop1 proteins within plant cells. Firstly, the WolfPsort algorithm was used to make predictions based on correlative amino acid sequence features and known sorting signal motifs (Horton et al. 2007). These predictions were compared with those obtained with the PsortII algorithm (Nakai and Kanehisa 1992), previously used for the identification of potential motifs.

Phylogenetic trees construction

The coding and protein sequences deposited in the built databases of effectors genes were aligned by codon with the ClustalW algorithm using Mega v. 7 (https://www.megasoftware.net/).Then, if needed, sequences were submitted to Gblocks version 0.91b server, with the aim to exclude poorly aligned regions and to make alignments more appropriate for phylogenetic reconstruction (Cruickshank 2000).

Phylogenetic analyses based on nucleotide sequences were performed using Bayesian inference with MrBayes v.3.04 (Huelsenbeck and Ronquist 2001). The substitution model for each tree was chosen based on Akaike information criterion (AIC) with MrModelTest 2.3 (Nylander 2004). Posterior

probabilities on trees distributions were created using the MCMC algorithm, starting from a random tree, and running for one million generations. In order to obtain a consensus tree, 25% of the trees were discarded. Tree visualization and rooting was done using FigTree v. 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

In order to compare the topologies of trees based on nucleotide sequences with those of the corresponding encoded proteins, a Bayesian phylogenetic analysis was executed for the peptide sequences of the effector proteins. These were aligned similarly as aforementioned. However, the trees were built through Cipres Science gateway v. 3.3 (http://www.phylo.org/index.php/), using Mr. Bayes on XSEDE. For this, the parameters were set to the same values as those used to construct the trees based on nucleotide sequences.

First, phylogenetic trees were built for each effector gene to reveal relationships of *E. psidii* with other species belonging to the *Enterobacteriaceae* family. Then, in order to have a better resolution on how each effector gene has evolved with respect to the rest of the genome, the topology of the effector-based tree was compared with that of the MLSA tree, based on the concatenated sequences of *atpD*, *gyrB*, *infB* and *rpoB* housekeeping genes (Supplementary Table S3). For that purpose, sequences of each housekeeping gene were aligned separately with ClustalW algorithm on Mega 7. Next, sequence ends were trimmed manually in order to maintain all sequences with the same length. Then, the software SequenceMatrix was used as a concatenation tool in order to assemble the multi-gene dataset (Vaidya, Lohman, and Meier 2011). The construction of the MLSA-based tree was then performed as for the single-gene trees.

To confirm the results obtained by phylogenetic trees, estimates of evolutionary divergence between species for each effector gene were performed using Mega 7 (Kumar, Stecher, and Tamura 2016), by computing pairwise distances on the previously generated alignment file. Gaps were treated as pairwise deletions and the algorithm chosen was Proportional (p)-distance. With this parameter set, the distance is given as the proportion (p) of nucleotide sites at which two sequences being compared are different. This value is obtained by

dividing the number of nucleotide differences by the total number of nucleotides compared (Swofford et al. 1996).

Occurrence of mobile genetic elements

The Dfam database release 3.1 (https://dfam.org/home) was used to predict the occurrence of DNA mobile genetic elements (MGEs) in the nucleotide sequences of the predicted effector genes of *E. psidii* strains. These elements are sequences of DNA that can be transferred between prokaryotic cells and contribute to genome rearrangement. Thus, MGEs can be considered as major agents mediating horizontal gene transfer (HGT). Comprobation of the occurrence of HGT on the evolutionary history of a bacterium species requires the knowledge of the elements that mediate this process (Frost et al. 2005).

Selection and recombination analyses

To determine the rate of synonymous (dS) and non-synonymous (dN) substitutions on the *E. psidii* effector protein coding sequences, the Data Monkey online server (https://www.datamonkey.org/) was used. To this end, the AvrE and Eop1 coding sequences from different strains of *Erwinia* species were aligned by codon, using the ClustalW algorithm implemented in Mega 7. In order to check if the alignment was in frame and did not contain any stop codons, the nucleotide sequences were translated into amino acid sequences, which were then mapped back onto their corresponding codons. Gaps were trimmed manually so that insertions and deletions did not affect the evaluation of substitution rates. Hereafter, with the aim to better understand the selective pressure prevailing on each codon, three tools from Datamonkey server were applied: single-likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL) and internal branches fixed-effects likelihood (IFEL). Values of $dN/dS = 1$, < 1, and > 1 indicate neutral evolution, negative and positive selection, respectively (Pond and Frost 2005).

RESULTS

Two highly conserved T3SE are present in the *E. psidii* genome

Similarity searches between the *E. psidii* type-strain IBSBF435 proteome against known T3SE of *E. amylovora* ATCC 49946 revealed that protein WP_124232032.1 from strain IBSBF435 shares 70.4% of amino acid identity with DspE (WP_004155375.1). Likewise, WP_124231866.1 protein from IBSBF435 shares 59% identity with Eop1 (WP_004155371.1). This search was repeated for the other *E. psidii s*trains included in this study and proteins with similar identity values with AvrE and Eop1 were found. No significant identity with HopX1 and Eop4 was found for any of the *E. psidii* strains. To confirm these results, similarity searches through Blastn were also performed using the genome sequences of the four strains of *E. psidii* against AvrE and Eop1 coding sequences of *E. amylovora* ATCC 49946. A locus in the IBSBF435 genome exhibiting 73.2% identity with the *avrE* gene was found on the complementary strand between positions 227821 and 233304. No significant similarity was found between the IBSBF435 genome and *eop1*.

Also, the predicted amino acid sequence of AvrE was subjected to a sequence similarity BlastP search against the NCBI database. It was found that AvrE from *E. psidii* IBSBF435 shares similarity with AvrE-family proteins from the following species: *E. mallotivora* (75% identity), *Pantoea agglomerans* (75%), *E. oleae* (75%), *P. vagans* (75%), *E. piriflorinigrans* (70.5%), *E. pyrifoliae* (70.6%), *E. tracheiphila* (69.9%), *E. tasmaniensis* (68.9%), and *P. stewartii* subsp. *stewartii* (65.3%). Lower similarity was found with AvrE proteins from *Pectobacterium carotovorum* (41.4%), *P. polaris* (41.3%), *P. carotovorum* subsp. *odoriferum* (41.3%), *P. peruviense* (41.2%), *P. carotovorum* subsp. *brasiliense* (41.2%), *Dickeya dadantii* (39.2%), and *D. chrysanthemi* (38.9%).

As for *E. psidii* Eop1 amino acid sequence, it shows identity to either predicted or validated avirulence proteins of the species: *P. agglomerans* (69.34%), *E. tracheiphila* (68.70%), *E. piriflorinigrans* (61.50%), *E. tasmaniensis* (61%), *E. pyrifoliae* (60.5) and *E. mallotivora* (53.9%).

E. psidii AvrE proteins contain signals for retention in the endoplasmic reticulum

PsortII predicted potential ERMRS motifs within the first five amino acids of the N-terminus and within the five last amino acids of the C-terminus of the *E. psidii* AvrE proteins (Figure 1). Nonetheless, whereas the ERMRS present at the C-terminus was conserved in the AvrE of all four strains (Figure 1a), an ERMRS motif was not found in the N-terminus of the AvrE from strain IBSBF435 (Figure 1b). Searches in the Pfam and Prosite databases did not reveal additional motifs.

Erwinia psidii AvrE proteins are predicted to localize to the plant nucleus

According to the prediction of PsortII, with 89% reliability, the putative AvrE effector proteins from *E. psidii* may be delivered through the T3SS to the plant cell nucleus. Although Wolfpsort provided different values for the subcellular localization of the AvrE proteins of the strains (Table 2), for IBSBF435, the prediction indicated the plant nucleus (rate 11) or the nuclear membrane (7.5) as possible subcellular localizations. The AvrE from strains LPF534 and LPF640 received different rates for more than four subcellular localizations, the highest of which were 5 for mitochondria and 4 for the nucleus. A high probability for localization to the plant nucleus (7) and the cytoplasm (5) was predicted for the AvrE protein from strain LPF681. Even though Wolfpsort indicated more than one possible subcellular localization, all four proteins received rates for nuclear localization, supporting the results obtained with PsortII.

Evolutionary models for phylogenetic analysis

Complete sequences of approximately 5,500 and 1,246 base pairs (bp) were obtained for *avrE* and *eop1* genes, respectively. As *eop1* sequences had the same length, there was no need for trimming, so phylogenetic tree was built using their full lengths. In contrast, the lengths of *avrE* sequences were uneven, so trimming using Gblocks was performed. After trimming, a total of 3,372 bp of *avrE* sequence was used.

For the housekeeping genes, *atpD, gyrB, infB* and *rpoB,* both partial and complete sequences ranging from 700 – 1,400, 760 – 2,400, 630 – 2,700, 650 – 4,200 bp, respectively, were obtained from the NCBI database. However, for a better phylogenetic reconstruction, the ends of the gene sequences were trimmed manually to obtain the same lengths, as described on the Supplementary Table S4. The final alignment length for building the concatenated tree based on *atpD, gyrB, infB* and *rpoB* genes was 2,281 bp.

The best evolutionary model selected by MrModeltest 2.3 for Bayesian inference based on Akaike information criterion (AIC) for each gene are described on Supplementary Table S5. The same models used for housekeeping genes individually were used to build the concatenated tree.

Erwinia psidii AvrE clusters in a clade separate from most *Erwinia* species

High posterior probability values on *avrE* nucleotide-based tree indicate strong confidence in the Bayesian inference results. Placement of species in both a*vrE* nucleotide- and amino acid-based trees (Supplementary Figures S1 and S2) was highly consistent. The *Pectobacterium* genus formed a single clade, its species and subspecies showed a consistent clustering pattern among branches. This genus is closely related to the genus *Dickeya*, as expected both formed neighbouring clades. On the other hand, *Pantoea a*nd *Erwinia* were more distantly related to *Pectobacterium* and *Dickeya*, but more closely related to each other. Individual *Erwinia s*pecies formed consistent clades, each species remained separate from the others in specific branches. *Erwinia amylovora, E. tasmaniensis, E. piriflorinigrans,* and *E. pyrifoliae* were found phylogenetically close to each other.

All four *E. psidii* strains clustered in a monophyletic clade, demonstrating that all of them share a common ancestral. In both *avrE* nucleotide- and translated amino acid-based trees, *E. psidii* was most closely related to a group containing *Pantoea agglomerans* pv. *gypsophilae* Ehg824-1, *P. stewartii* subsp. *stewartii* DC283, *P. agglomerans* FDAARGOS_407, *E. mallotivora* MT-MARDI,

P. vagans FDAARGOS_160, *P. agglomerans* DAPP-PG734, *E. oleae* DAPP-PG53 and *E. tracheiphila* strains.

Distance matrix values demonstrated conservation of the *avrE* coding sequences among the four strains of *E. psidii* (Table 3), which supports their positions in the phylogenetic tree. However, the distance matrix shows that *avrE* genes of *E. psidii* possess less divergence with those of *E. mallotivora*, instead of the clade formed by *Pantoea agglomerans* pv. *gypsophilae* Ehg824-1 and six other strains as shown in the phylogenetic trees. The overall mean distance between *avrE* sequences within this clade was 21.6%, which means that these genes share 78.4% of conserved regions.

Erwinia psidii **clusters together with other** ***Erwinia*** **species in an MLSA tree**

Multi-locus sequence analysis using the four housekeeping genes *gyrB*, *atpD, rpoB a*nd *infB* showed that *E. psidii* clustered together with other *Erwinia* species (Figure 3). This analysis showed that *E. psidii* was most closely related to *E. tracheiphila*, both of which share a common ancestor with *E. mallotivora.* The closest proximity of *E. psidii* to *E. tracheiphila* was consistently observed in all phylogenetic trees using individual housekeeping genes (Supplementary Figures S5 to S8). As for the *Pectobacterium* species, in the MLSA tree it was observed the formation of neighbouring clades with *Dickeya* instead of sharing the same monophyletic group as noted in *avrE* trees (Figure 2).

Absence of mobile genetic elements (MGEs) in the ***avrE*** **gene**

Through searches against the Dfam database, no evidence was found for the presence of MGEs in the *avrE* nucleotide sequences of the four *E. psidii* strains. Two *avrE* sequences from *Pantoea* species that showed close relationship with *E. psidii* were also assessed for the occurrence of MGEs. No evidence was found for the *avrE* gene of *P. stewartii* subsp. *stewartii* strain DC283, but an interspersed repeat sequence with similarity to another from

Caenorhabditis elegans was found in the *avrE* from *P. agglomerans* pv. *gypsophilae* Ehg824-1 gene.

Selection pressures acting on the *avrE* gene

Based on a*vrE* nucleotide sequences from 16 different strains of *Erwinia* species, 1,530 codons were tested for the occurrence of selection pressures. The three different methods used, SLAC, FEL and IFEL, found no evidence of positive selection at any sites. Based on FEL, 100% of the codons have undergone neutral selection pressure. According to SLAC method, 9 out of 1,530 codons suffered negative selection (synonymous substitutions) whereas IFEL indicated that 61 sites (3.9%) have undergone negative selection pressure and 1,469 (96%) suffered neutral selection (Table 4).

***Erwinia psidii* Eop1 retains a conserved catalytic triad**

An alignment of *E. psidii* Eop1 with similar sequences from different *Erwinia s*pecies, all previously used by Asselin et al. (2011), revealed it to possess an intact H/E/C catalytic triad (Figure 5) commonly found in members of this protein family.

***Erwinia psidii* Eop1 is predicted to localize to different plant organelles**

PsortII predicted the localization of Eop1 putative effector to the plant nucleus. As for Wolfpsort, it provided the following rates for the strains IBSBF435, LPF534 and LPF640: 7 – mitochondria; 5 – chloroplast; 2 – nucleus. The putative effector of LPF681 was predicted to have multiple localization sites, since it received rate 5 for both mitochondria and chloroplast, and 4 for nucleus. By comparing the results of the two algorithms, the *E. psidii* Eop1 predicted subcellular localization within host cells remained uncertain for all strains (Table 5).

Erwinia psidii Eop1 clusters in a clade separate from those of other *Erwinia* species

High posterior probability values indicate strong confidence in the bayesian inference results for *eop1* nucleotide- and amino acid-based trees (Supplementary Figures S3 and S4). The clustering pattern among species was very consistent on both trees. In the nucleotide-based tree, despite all *E. psidii* strains clustering in the same clade, there is some divergence between *eop1* from strain LPF681 and the *eop1* of the other *E. psidii* strains. The latter remained together in a single branch. This separation of the strains from LPF681 was also observed in the amino acid-based tree. In both trees, *eop1* from *E. psidii* was most closely related to *P. agglomerans* DAPP-PG734 and *E. tracheiphila* BuffGH (Figure 4). Thus, the phylogeny of the amino acid-based tree is in agreement with that of its corresponding coding sequence tree.

Eop1 from *E. psidii* strains were found to have a close relationship with the following species, in a decrescent order: *E. tasmaniensis* ET1/99; *E. piriflorinigrans* CFBP5888; *E. amylovora* Ea510; *E. amylovora* Ea246; the other *E. amylovora* (Ea587, Ea566, Ea262, E-2) strains; *E. pyrifoliae* Ep4/97 and EpK1/15; and *E. pyrifoliae* WT3.

Distance matrix values confirmed that there is a small divergence in *eop1* coding sequence of strain LPF681 when compared to the other strains of *E. psidii* (Table 6). The distance matrix is also in agreement with the tree topology, since both demonstrated that *eop1* genes of *E. psidii* strains are more closely related to the genes of *P. agglomerans* DAPP-PG734 and *E. tracheiphila* BuffGH, and more distant to the ones of *E. pyrifoliae* strains. The overall mean distance between *eop1* nucleotide sequences of these 16 strains was 27.2%, which means that these genes share 72.8% of conserved regions.

Absence of mobile genetic elements (MGEs) in the *eop1* gene

Through searches against the Dfam database, no evidence of MGEs was found in the *eop1* nucleotide sequences of the four *Erwinia psidii* strains. The same result was obtained for the gene of *P. agglomerans* DAPP-PG734, which demonstrated a close relationship to *E. psidii* strains in the phylogenetic analysis.

Selection pressures acting on the *eop1* gene

Based on *eop1* nucleotide sequences from 11 *Erwinia* species, 391 codons were analyzed for the occurrence of selection pressures. SLAC and IFEL found no evidence of positive selection at any site, but found, respectively, 34 and 98 negative selected sites (synonymous substitutions). The remainder sites were under neutral selection pressure. Based on FEL, 109 codons (27.8%) have undergone negative selection and 281 (71.8%) were subjected to neutral selection pressure. FEL also pointed that codon number 30 might have been exposed to positive selection pressure (Table 7). This codon is composed of three guanines, which code for glycine, according to the standard genetic code table.

DISCUSSION

In the present study, through similarity searches we identified genes coding for two putative type III effector proteins in the genome of four *E. psidii* strains, an AvrE-family protein and an Eop1-family protein. In previous studies, both were described as secreted effector proteins by the type III secretion system of *E. amylovora* (Bogdanove et al., 1998; Zhao et al., 2005). No sequence similarity with HopX1 and Eop4 was found in the draft genomes of the four *E. psidii* strains studied. The result for Eop4 is consistent with previous reports indicating that this effector was only found in *E. amylovora* strains and not in other closely related *Erwinia* species (Zhao, He, and Sundin 2006). Effector proteins belonging to the AvrE-family are known to have a meaningful contribution to bacterial virulence on host plants (Degrave et al. 2015). It is known that the pathogenicity of *E. amylovora* is dependent on the function of a single secreted

effector: DspE, a member of the AvrE-family (Bogdanove et al., 1998; Boureau et al., 2006). Thus, it is possible that this putative AvrE effector protein may play an important role in the pathogenicity of *E. psidii.*

Motif searches on the AvrE putative effector protein identified ERMRS motifs on both C- and N-termini, in the same locations as reported in other AvrE-family effector proteins (Degrave et al. 2015). ERMRS motifs are known to be required for virulence and avirulence activities of this protein family (Ham et al. 2009). The importance of this motif was demonstrated by Ham et al. (2007) when its partial deletion abolished the pathogenicity function of WtsE effector of *P. stewartii s*ubsp. *stewartii*. WtsE is a member of the AvrE protein family (Ham et al. 2007). Also, when this motif was mutated in AvrE of *P. syringae* pv. *tomato* DC3000, the virulence of this bacterium became defective (Ham et al. 2009). It remains to be determined through *in vivo a*nalysis whether these motifs are required for the function of this putative effector in *E. psidii*.

The presence of ERMRS motifs in the *E. psidii* AvrE proteins also suggests that this protein may localize to the plant endoplasmic reticulum (Degrave et al. 2015). Interestingly, although the PsortII program identified this motif, it also indicated that this putative effector may be delivered to the plant cell nucleus. This nuclear localization prediction is consistent with previous predictions for other AvrE-family effectors using the same program (Ham et al. 2007). Even though the AvrE-family proteins of *E. psidii* do not harbor recognizable NLS, it is known that proteins can be imported into the plant nucleus without carrying a canonical NLS (Degrave et al. 2015).

An important question that needs to be answered is what were the evolutionary processes underlying *E. psidii* adaptation to infect different host species, considering that the hypothesis that this bacterium suffered a host shift from *P. guajava* to *Eucalyptus* spp. (Coutinho et al. 2011) has been raised. Thus, considering that AvrE is a key factor for the pathogenicity of *Erwinia* species and to better understand the evolutionary history of *E. psidii c*ompared to other phytopathogens of the *Enterobacteriaceae* family, phylogenetic analyses of this putative effector were performed (Kannan and Wheeler 2012). On AvrE trees (both based on nucleotide or amino acid sequences), *E. psidii* consistently clustered together in a clade closely related to another formed by other *Erwinia*

species and some *Pantoea* species. The topology of these trees was robust as indicated by the clustered position of most individual species. Nonetheless, clustering of *Pectobacterium* species and subspecies was inconsistent, which is supported by previously reported results, where *P. carotovorum* subsp. *carotovorum* and *P. carotovorum* subsp. *brasiliense* did not cluster separately, but scattered among other *Pectobacterium* subspecies clades (Ma et al. 2007; Zhang, Fan, and Loria 2016; Christen and Gardan 2019). *Pectobacterium* and *Dickeya* formed contiguous clades as expected, since both genus belong to the soft-rot *Enterobacteriaceae* group (Golanowska et al. 2018). The other two genus of the *Enterobacteriaceae* family included in the analysis, *Pantoea a*nd *Erwinia,* shared the same cluster (Young and Park 2007), supporting the known close relationship between them; *Pantoea* species emerged from a reclassification of the genus *Erwinia* (Mergaert, Verdonck, and Kersters 1993; Hauben et al. 1998).

The separation of species in the AvrE-based trees from the genus *Erwinia* was consistent, since each species remained separate from the others in specific branches. The close relationship among *E. amylovora, E. tasmaniensis, E. piriflorinigrans* and *E. pyrifoliae* was similar to the reported in other studies (Geider et al. 2006; López et al. 2011; Arriel et al. 2014). It was observed that *E. psidii* and *E. tracheiphila* share a common ancestor, and are placed in the same large cluster with *E. mallotivora*. The phylogenetic proximity between these three *Erwinia* species agrees with previously reported results (Geider et al. 2006, 2009; Young and Park 2007; Arriel et al. 2014).

An evolutionary history of the *avrE* effector in comparison with that of the housekeeping genes of the *Enterobacteriaceae* was obtained. The topology of the concatenated MLSA tree is consistent with results obtained by Arriel et al. (2014), indicating that the housekeeping genes of *E. psidii* shared the same cluster with *E. tracheiphila* and *E. mallotivora*. A careful comparison of the MLSA and *avrE* trees suggests that a recent divergence of the *E. psidii avrE* genes from those of *E. tracheiphila* occurred. This divergence may have resulted from a horizontal gene exchange event between *E. psidii* and *Pantoea agglomerans* pv. *gypsophilae*. Horizontal gene transfers are common to happen in virulence genes located in pathogenicity islands, since those are unstable regions of the genome that can be easily transferred to other bacteria (Arber 2014). The hypothesis of

the occurrence of a horizontal gene transfer between these two species is supported by the presence of mobile genetic elements on the *avrE* gene of *P. agglomerans* pv. *gypsophilae* (Hubley et al. 2015; Frost et al. 2005), and also by a previous report of *P. agglomerans* internally colonizing eucalypt as an endophyte (Ferreira et al. 2008). The proximity between cells of these two bacterial species inside host tissue could explain an event of horizontal gene transfer.

Evidence of no positive selection pressure was observed in the AvrE gene when analyzing the sequence of different *Erwinia* species; only footprints of neutral and negative selections were found acting on this gene. This result suggests that the evolution process of this gene was not due to accumulation of non-synonymous nucleotide substitutions. Considering that dieback of eucalypt caused by *E. psidii* is a recently emerged disease, and that, so far, resistant eucalypt species are still being selected and tested in the field, it is tempting to hypothesize that not enough time has passed in order to detect host selection pressure acting on this pathogen (Caires et al. 2019). It is currently widely accepted that the interaction of pathogens with their host cells impose strong selective pressure on the host, while the host defense system applies a reciprocal selective pressure on the pathogen (Ma et al. 2006).

Besides *avrE*, the putative *eop1* effector gene was found to be conserved in the *E. psidii* genome (Nissinen et al. 2007; Asselin et al. 2011; Castiblanco, Triplett, and Sundin 2018). According to Pfam, this putative effector was classified as a *Yersinia* outer protein J (YopJ) Serine/Threonine acetyltransferase. YopJ is one of the most widely distributed T3SE families in both plant and animal pathogens (Lewis et al. 2011; Ka-Wai Ma 2016). A common feature of YopJ family members is a conserved H/E/C catalytic triad, which is essential for the full function of YopJ homologous proteins (Orth et al. 2000). These catalytic triad was conserved in the Eop1 putative effector of *E. psidii* strains, suggesting that these proteins might be functional and play an important role in avirulence, resembling the function as host limiting factors exhibited by their *E. amylovora* counterparts (Asselin et al. 2011).

The predicted subcellular localization of Eop1 putative effector within host cells remains uncertain, based on the contrasting results obtained by the two

computational prediction methods applied. As stated by Lewis et al. (2011), the homologous members of the YopJ superfamily present very distinct subcellular localization sites, and therefore YopJ still does not have a well-defined subcellular localization within host cells. Thus, further *in vivo* analysis should be employed in order to determine the subcellular localization of this putative effector of *E. psidii.*

As for the phylogenetic analysis, comparison of the topologies of the *eop1* and housekeeping genes demonstrated a coinciding evolutionary history (Asselin et al. 2011), except for the phylogenetic proximity between *E. psidii* and *P. agglomerans* DAPP-PG734 observed in the e*op1* tree. This differential evolution of the e*op1* gene could have been due to a horizontal gene transfer between *Erwinia* and *Pantoea* species, although this hypothesis was not supported by the assessment for the presence of mobile genetic elements in the *eop1* genes of these species (Hubley et al. 2015). However, the possibility to reach clearer conclusions is compromised by the limited number of *eop1* sequences available in the databases.

Curiously, in the phylogenetic analysis, it was noticed that the *eop1* gene from *E. psidii* LPF681 diverged from those of the other strains belonging to this species. According to Montoya-Estrada (2018), this strain was first isolated from *P. guajava* and showed weak aggressiveness on an *E. urophylla* × *E. grandis* hybrid, and was classified as nonvirulent on an *E. urophylla* clone. This raises the possibility that indeed *eop1* is a secreted effector protein, this divergence in the LPF681 sequence might lead to its function as host limiting factor, and therefore be the cause of this strain non-virulence on some *Eucalyptus* clones.

Additionally, we found evidence of positive selection pressure acting on the codon 30 of the *eop1* gene. This result suggests that the *eop1* putative effectors are accumulating nucleotide substitutions, however no previous studies were found demonstrating the importance of this codon for Eop1 protein function.

To date, little was known about how *E. psidii* causes disease on its host plants. Thus, searches for putative effectors and evolutionary analyses of their sequences were necessary to further our knowledge on the pathogenic potential of this bacterial species. In this study, we identified two potential effector proteins and traced their evolutionary history, aiming to better understand the capacity of this bacterium to cause disease on eucalypt at different aggressiveness levels,

varying according to the strain. However, functional characterization of AvrE and Eop1 as well as additional predicted effectors is required in order to clearly understand the molecular and evolutionary mechanisms behind *E. psidii* pathogenicity. Here, we demonstrate that comparisons of the effector repertoires of strains with different aggressiveness levels can provide important insights into this bacterium pathogenicity. Such comparisons are likely to reveal the molecular bases for whether and/or how *E. psidii* underwent a host shift to gain the ability to infect eucalypt.

ACKNOWLEDGEMENTS

The authors would like to thank CAPES for the Master's degree scholarship provided to Isadora Cristófoli Pereira and for partially funding this research - Finance Code 001.

REFERENCES

Aksoy, H. M.; Kaya, Y.; Tengku, A. H. 2017. Expression of the *dspA/E* gene of *Erwinia amylovora* in non-host plant *Arabidopsis thaliana*. Biotechnol. Biotechnol. Equip. 31:85–90.

Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J.; Pennsylvania, T. 1990. Basic Local Alignment Search Tool. J. Mol. Biol. 215:403–410.

Arber, W. 2014. Horizontal gene transfer among bacteria and its role in biological evolution. Life. 4:217–224.

Arriel, D. A. A.; Fonseca, N. R.; Guimarães, L. M. S.; Hermenegildo, P. S.; Mafia, R. G.; Borges Júnior, N.; de Souza, H. P. Alfenas, A. C. 2014. Wilt and die-back of *Eucalyptus* spp. caused by *Erwinia psidii* in Brazil. For. Pathol. 44:255–265.

Asselin, J. E.; Bonasera, J. M.; Kim, J. F.; Oh, C.; Beer, S. V. 2011. Eop1 from a *Rubus* strain of *Erwinia amylovora* functions as a host-range limiting factor. Phytopatol. 101:935–944.

Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A. A.; Dvorkin, M.; Kulikov, A. S.; Lesin, V. M.; Nikolenko, S. I.; Pham, S.; Prjibelski, A. D.; Pyshkin, A. V.; Sirotkin, A. V.; Vyahhi, N.; Tesler, G.; Alekseyev, M. A.; Pevzner, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455–477.

Blum, G.; Ott, M.; Lischewski, A.; Ritter, A.; Imrich, H.; Tschape, H. 1994. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. Infect. Immun. 62:606–614.

Bocsanczy, A. M.; Schneider, D. J.; Declerck, G. A.; Cartinhour, S.; Beer, S. V. 2012. HopX1 in *Erwinia amylovora* functions as an avirulence protein in apple and is regulated by HrpL. J. Bacteriol. 194:553–560.

Bogdanove, A. J.; Bauer, D. W.; Beer, S. V. 1998. *Erwinia amylovora* secretes

DspE, a pathogenicity factor and functional AvrE homolog, through the Hrp (type III secretion) pathway. J. Bacteriol. 180:2244–2247.

Boureau, T.; ElMaarouf-Bouteau, H.; Garnier, A.; Brisset, M. N.; Perino, C.; Pucheu, I. 2006. DspA/E, a Type III effector essential for *Erwinia amylovora* pathogenicity and growth in planta, induces cell death in host apple and nonhost tobacco plants. Mol. Plant-Microbe Interact. 19:16–24.

Caires, N. P.; Hermenegildo, P. S.; Guimarães, L. M. S.; Mafia, R. G.; Zauza, E. A. V.; Júnior, N. B.; Badel, J. L.; Alfenas, A. C. 2019. Host range of *Erwinia psidii* and genetic resistance of *Eucalyptus* and *Corymbia* species to this pathogen. Forest. Pathol. 12527:1-10.

Castiblanco, L. F.; Triplett, L. R.; Sundin, G. W. 2018. Regulation of effector delivery by type III secretion chaperone proteins in *Erwinia amylovora*. Front. Microbiol. 9:1–10.

De Castro, E.; Sigrist, C. J. A.; Gattiker, A.; Bulliard, V.; Langendijk-Genevaux, P. S.; Gasteiger, E. 2006. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 34:362–365.

Christen, R.; Gardan, L. 2019. Elevation of three subspecies of *Pectobacterium carotovorum* to species level : *Pectobacterium atrosepticum* sp. nov., *Pectobacterium betavasculorum* sp. nov. and *Pectobacterium wasabiae* sp. nov. Int. J. Sys. Evol. Microbiol. 53:381–391.

Coutinho, T. A.; Brady, C. L.; Vaart, M.; Van Der, V. S. N.; Telechea, N.; Rolfo, M. 2011. A new shoot and stem disease of eucalyptus species caused by *Erwinia psidii*. Australas. Plant Pathol. 40:55–60.

Cruickshank, R. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17:540–52.

Debroy, S.; Thilmony, R.; Kwack, Y. B.; Nomura, K.; He, Y. S. 2004. A family of conserved bacterial effectors inhibits salicylic acid-mediated basal immunity and promotes disease necrosis in plants. PNAS. 101: 9927-9932.

Degrave, A.; Siamer, S.; Boureau, T.; Barny, M. A. 2015. The AvrE superfamily: ancestral type III effectors involved in suppression of pathogen-associated molecular pattern-triggered immunity. Mol. Plant Pathol. 16:899–905.

El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S. R.; Luciani, A.; Potter, S. C. 2019. The Pfam protein families database in 2019. Nucleic Acids Res. 47:D427–D432.

Ferreira, A.; Quecine, M. C.; Lacava, P. T.; Oda, S.; Azevedo, J. L.; Araújo, W. L. 2008. Diversity of endophytic bacteria from *Eucalyptus* species seeds and colonization of seedlings by *Pantoea agglomerans*. FEMS Microbiol. Lett. 287: 8–14.

Frost, L. S.; Leplae, R.; Summers, A. O.; Toussaint, A. 2005. Mobile genetic elements: the agents of open source evolution. Nat. Rev. Microbiol. 3:722-732.

Geider, K.; Auling, G.; Du, Z.; Jakovljevic, V.; Jock, S.; Vo, B. 2006. *Erwinia tasmaniensis* sp. nov., a non-phytopathogenic bacterium from apple and pear trees. Int. J. Syst. Evol. Microbiol. 56:2937–2943.

Geider, K.; Auling, G.; Jakovljevic, V.; Vo, B. 2009. A polyphasic approach assigns the pathogenic *Erwinia* strains from diseased pear trees in Japan to *Erwinia pyrifoliae*. Appl. Microbiol. 48:324–330.

Golanowska, M.; Potrykus, M.; Motyka-Pomagruk, A.; Kabza, M.; Bacci, G.; Galardini, M.; Bazzicalupo, M.; Makalowska, I.; Smalla, K.; Mengoni, A.; Hugouvieux-Cotte-Pattat; Lojkowska, E. 2018. Comparison of highly and weakly virulent *Dickeya solani* strains, with a view on the pangenome and panregulon of this species. Front. Microbiol. 9:1-19.

Glaeser, S. P.; Kampfer, P. 2015. Multilocus sequences analysis (MLSA) in prokaryotic taxonomy. Sys. Appl. Microbiol. 38:237-245.

Grant, S. R.; Fisher, E. J.; Chang, J. H.; Mole, B. M.; Dangl, J. L. 2006. Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. Annu. Rev. Microbiol. 60:425–449.

Hacker, J.; Blum-Oehler, G.; Mühldorfer, I.; Tschäpe, H. 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol. Microbiol. 23:1089–1097.

Ham, J. H.; Majerczak, D. R.; Arroyo-Rodriguez, A. S.; Mackey, D. M.; Coplin, D. L. 2007. WtsE, an AvrE-family effector protein from *Pantoea stewartii* subsp. *stewartii*, causes disease-associated cell death in corn and requires a chaperone protein for stability. Mol. Plant-Microbe Interact. 19:1092–1102.

Ham, J. H.; Majerczak, D. R.; Nomura, K.; Mecey, C.; Uribe, F.; He, S. Y. 2009. Multiple activities of the plant pathogen type III effector proteins WtsE and AvrE require wxxxe motifs. Mol. Plant-Microbe Interact. 22:703–712.

Hauben, L.; Moore, E. R. B.; Vauterin, L.; Steenackers, M.; Mergaert, J.; Verdonck, L. 1998. Phylogenetic position of phytopathogens within the Enterobacteriaceae. Syst. Appl. Microbiol. 21:384–397.

Hermenegildo, S.; Santos, S. A.; Guimarães, L. M. S.; Pereira, I. C.; Vidigal, P. M. P.; Badel, J. L.; Alfenas-Zerbini, P.; Mafia, G. R.; Ferreira, A. S. V. M.; Alfenas, A. C. 2019. Draft genome sequence of *Erwinia psidii*, causal agent of Bacterial Blight of Guava (*Psidium guava*) and Dieback of Eucalypt (*Eucalyptus* spp.) Microbiol. Resour. Announc. 8:30–32.

Horton, P.; Park, K. J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C. J. 2007. WoLF PSORT: Protein localization predictor. Nucleic Acids Res. 35:585–587.

Huang, Y.; Niu, B.; Gao, Y.; Fu, L.; Li, W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics. 26:680–682.

Hubley, R.; Finn, R. D.; Clements, J.; Eddy, S. R.; Jones, T. A.; Bao, W.; Smit, A. F. A.; Wheeler, T. J. 2015. The Dfam database of repetitive DNA families. Nucleic Acids Res. 44:81-89.

Huelsenbeck, J. P.; Ronquist, R. 2001. Mrbayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Jin, L.; Ham, J. H.; Hage, R.; Zhao, W.; Soto-Hernández, J.; Lee, S. Y. 2016.

Direct and indirect targeting of pp2a by conserved bacterial type-III effector proteins. PLoS Pathog. 12:1–33.

Jones, J. D. G.; Dangl, J. L. 2006. The plant immune system. Nature. 444:323–329.

Ka-Wai Ma, W. M. 2016. YopJ family effectors promote bacterial infection through a unique acetyltransferase activity. Microbiol. Mol. Biol. Rev. 80:1011–1027.

Kamber, T. 2013. Interactions characterization of *Erwinia amylovora* - host plant and - biocontrol agent interactions. University of Zurich, Faculty of Science.

Kannan, L.; Wheeler, W. C. 2012. Maximum parsimony on phylogenetic networks. Algorithms. Mol. Biol. 7:1-9

Kumar, S.; Stecher, G.; Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33:1870–4.

Lewis, J. D.; Lee, A.; Ma, W.; Zhou, H.; Guttman, D. S.; Desveaux, D. 2011. The YopJ superfamily in plant-associated bacteria. Mol. Plant Pathol. 12:928–937.

Lindgren, P. B. 1997. The role of *hrp* genes during plant-bacterial interactions. Annu. Rev. Phytopathol. 35:129–152.

Liu, B.; Luo, J.; Li, W.; Long, X.; Zhang, Y.; Zeng, Z. 2016. *Erwinia teleogrylli* sp. nov., a bacterial isolate associated with a chinese cricket. PLoS ONE. 11:1-12.

López, M. M.; Roselló, M.; Llop, P.; Ferrer, S.; Christen, R.; Gardan, L. 2011. *Erwinia piriflorinigrans* sp. nov., a novel pathogen that causes necrosis of pear blossoms. Int. J. Syst. Evol. Microbiol. 61:561–567.

Ma, B.; Hibbing, M. E.; Kim, H.; Reedy, R. M.; Yedidia, I.; Breuer, J. 2007. Host range and molecular phylogenies of the soft rot enterobacterial genera *Pectobacterium* and *Dickeya*. Phytopathology. 97:1150-1163.

Ma, W.; Dong, F. F. T.; Stavrinides, J.; Guttman, D. S. 2006. Type III effector diversification via both pathoadaptation and horizontal transfer in response

to a coevolutionary arms race. PLoS Genet. 2:2131–2142.

Malnoy, M.; Martens, S.; Norelli, J. L.; Barny, M. A.; Sundin, G. W.; Smits, T. H. M. 2012. Fire Blight: applied genomic insights of the pathogen and host. Annu. Rev. Phytopathol. 50:475–494.

Mergaert, J.; Verdonck, L.; Kersters, K. 1993. Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the genus *Pantoea* emend. as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb, nov., respectively, and description of *Pantoea stewartii* subsp. *indologenes* subsp. nov. Int. J. Syst. Bacteriol. 43:162–173.

Montoya-Estrada, C. N.; Costa, C. R.; Badel, J. L.; Guimarães, L. M. S.; Alfenas, A. C. 2018. Root infection and aerial colonization of eucalypt host plants by *Erwinia psidii*. Trop. Plant Pathol. 44:251–257.

Moretti, C.; Hosni, T.; Vandemeulebroecke, K.; Brady, C.; Vos, P; De, Buonaurio, R. 2010. *Erwinia oleae* sp. nov., isolated from olive knots caused by *Pseudomonas savastanoi* pv. *savastanoi*. Int. J. Syst. Evol. Microbiol. 61: 2745-2752.

Mukherjee, S.; Keitany, G.; Li, Y.; Wang, Y.; Ball, H.; Goldsmith, E. 2006. *Yersinia* YopJ acetylates and inhibits kinase activation by blocking phosporylation. Science. 80:1211–1214.

Nakai, K.; Kanehisa, M. 1992. A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics. 14:897–911.

Nissinen, R.; Ytterberg, A. J.; Institutet, K.; Wijk, K. J.; Beer, S. V. 2007. Analyses of the secretomes of *Erwinia amylovora* and selected *hrp* mutants reveal novel type III secreted proteins and an effect of HrpJ on extracellular harpin levels. Mol. Plant Pathol. 8:55–67.

Oh, C. S.; Beer, S. V. 2005. Molecular genetics of *Erwinia amylovora* involved in the development of fire blight. FEMS Microbiol. Lett. 253:185–192.

Orth, K.; Xu, Z.; Mudgett, M. B.; Bao, Z. Q.; Palmer, L. E.; Bliska, J. B. 2000. Disruption of signaling by *Yersinia* effector YopJ, a ubiquitin-like protein

protease. Science. 290:1594–1597.

Piqué, N.; Miñana-Galbis, D.; Merino, S.; Tomás, J. M. 2015. Virulence factors of *Erwinia amylovora*: a review. Int. J. Mol. Sci. 16:12836-12854.

Pond, K. L. S.; Frost, S. D. W. Not so different after all: a comparison of methods for detecting amino acid sites under selection. 2005. Mol. Biol. Evol. 22:1208-1222.

Rodrigues Neto, J.; Robbs, C. F.; Yamashiro, T. 1987. A bacterial disease of guava (*Psidium guajava*) caused by *Erwinia psidii* sp. nov. Fitopatol. Bras. 12:345-350.

Siamer, S.; Guillas, I.; Shimobayashi, M.; Kunz, C.; Hall, M. N.; Barny, M. A. 2014. Expression of the bacterial type III effector DspA/E in *Saccharomyces cerevisiae* down-regulates the sphingolipid biosynthetic pathway leading to growth arrest. J. Biol.Chem. 289:18466-18477.

Sokurenko, E. V.; Hasty, D. L.; Dykhuizen, D. E. 1999. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. Trends Microbiol. 7:191–195.

Stavrinides, J.; McCann, H. C.; Guttman, D. S. 2008. Host-pathogen interplay and the evolution of bacterial effectors. Cell. Microbiol. 10:285–292.

Vaidya, G.; Lohman, D. J.; Meier, R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics. 27:171–180.

Young, J. M.; Park, D. C. 2007. Relationships of plant pathogenic enterobacteria based on partial *atpD, carA*, and *recA* as individual and concatenated nucleotide and peptide sequences. Syst. Appl. Microbiol. 30:343–354.

Zhang, Y.; Fan, Q.; Loria, R. 2016. A re-evaluation of the taxonomy of phytopathogenic genera *Dickeya* and *Pectobacterium* using whole-genome sequencing data. Syst. Appl. Microbiol. 39:252–259.

Zhao, Y.; Blumer, S. E.; Sundin, G. W. 2005. Identification of *Erwinia amylovora* genes induced during infection of immature pear tissue. J. Bacteriol. 187:8088–8103.

Zhao, Y.; He, S. Y., Sundin, G. W. 2006. The *Erwinia amylovora avrRpt2* EA gene contributes to virulence on pear and *avrRpt2* EA is recognized by *Arabidopsis* RPS2 when expressed in *Pseudomonas syringae*. Mol. Plant-Microbe Interact. 19:644–654.

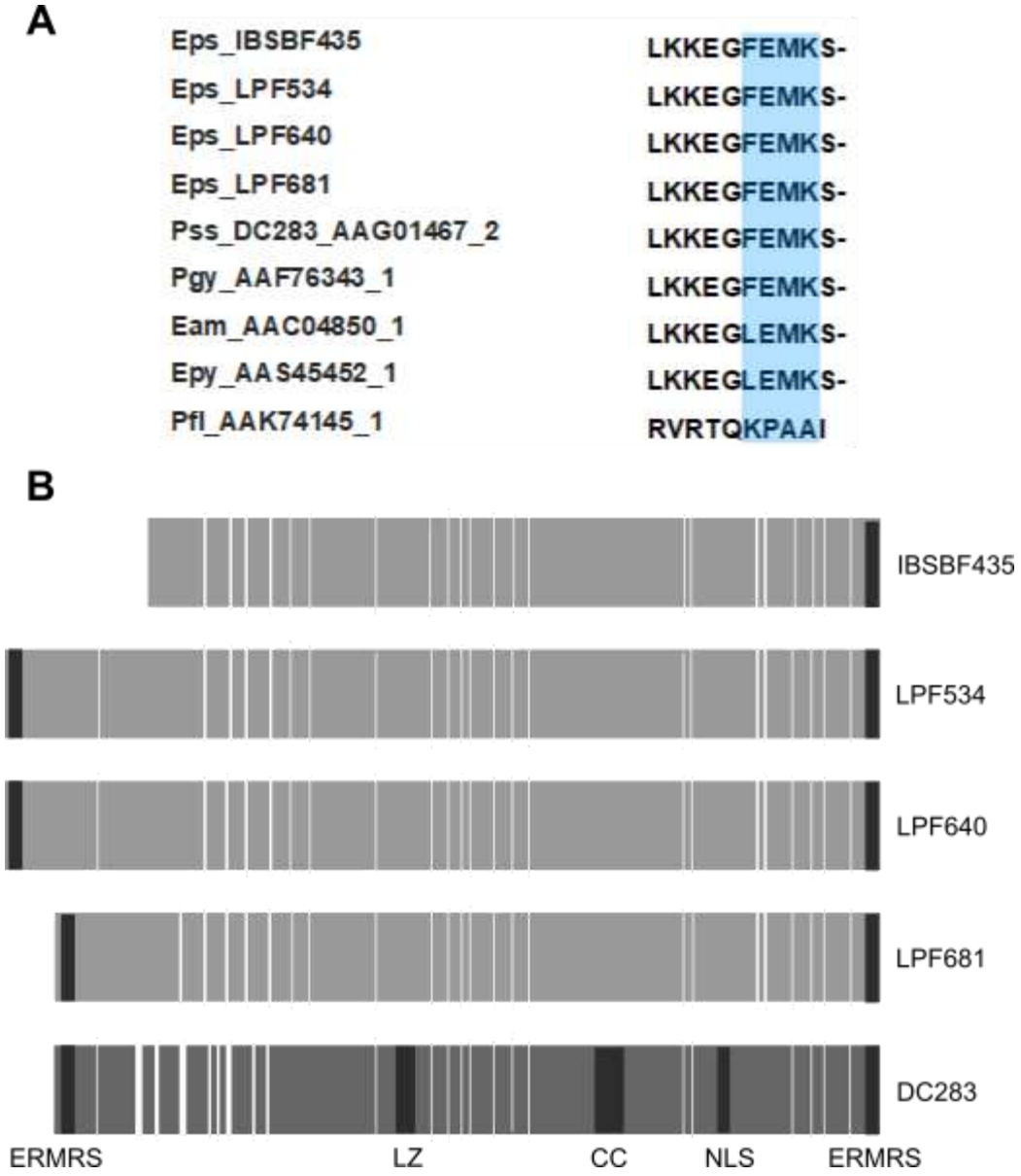


Figure 1. A. The last 10 amino acids at the C-terminus of AvrE-family proteins from four *Erwinia psidii* (Eps) strains, *Pantoea stewartii* subsp. *stewartii* DC283 (Pss; GenBank accession number AAG01467.2), *Pantoea agglomerans* pv. *gypsophilae* (Pgy; AAF76343.1), *Erwinia amylovora* (Eam; AAC04850.1), *Erwinia pyrifoliae* (Epy; AAS45452.1), and *Pseudomonas fluorescens* SBW25 (Pfl; AAK74145.1). The ERMRS motifs predicted by PsortII are highlighted in blue. **B.** Predicted functional motifs in AvrE. White bars indicate gaps in the alignment of amino acid sequences. Dark gray bars indicate a leucine zipper (LZ), coiled coil region (CC), nuclear localization signal (NLS) and ERMRS motifs.

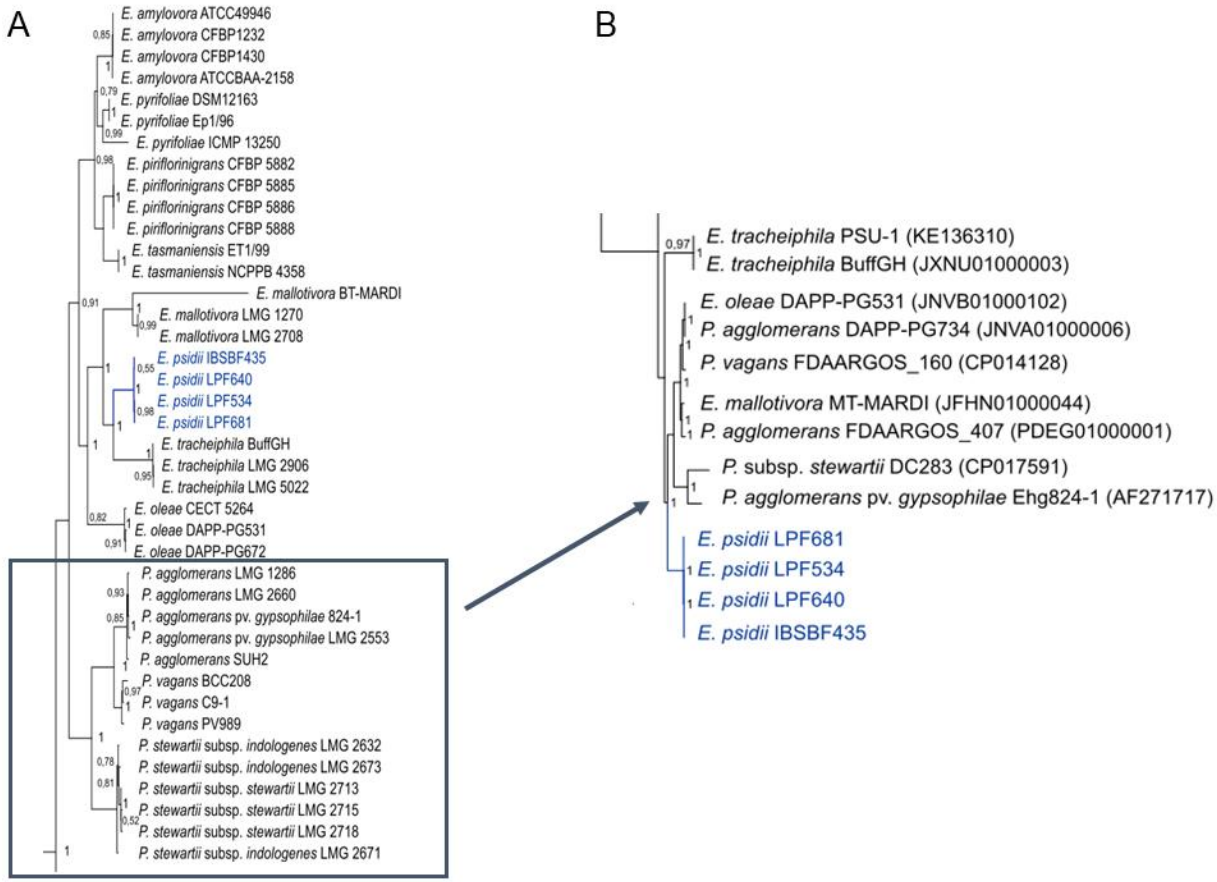


Figure 2. A. Bayesian phylogeny of bacterial species of the *Enterobacteriaceae* family using concatenated partial *gyrB*, *atpD, rpoB a*nd *infB* gene sequences **B.** Bayesian phylogenetic analysis of the *avrE* nucleotide sequences from bacteria of the *Enterobacteriaceae* family. Posterior probability values are indicated in each node. GenBank accession numbers are indicated in parentheses after strain identification. Bar denotes the rate of substitutions per site. *Erwinia psidii* strains are highlighted in blue. The positions of *Pantoea s*pecies are detached by the arrow and the square.

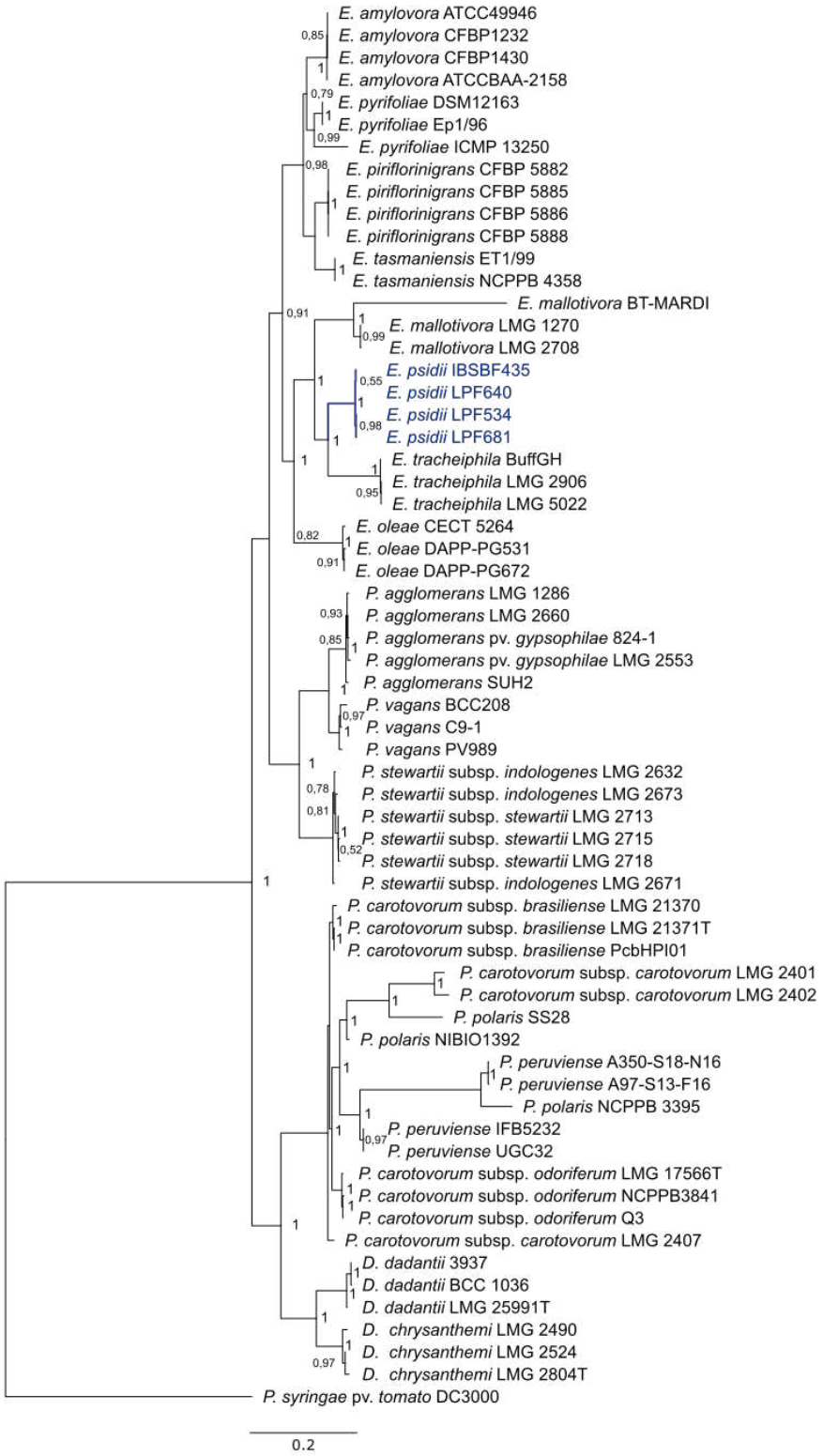


Figure 3. Bayesian phylogeny of 61 strains of bacterial species of the *Enterobacteriaceae* family using concatenated partial *gyrB*, *atpD, rpoB a*nd *infB* gene sequences. Posterior probability values are indicated in each node. *Pseudomonas syringae* pv. *tomato* was included as an outgroup. Bar denotes the rate of substitutions per site. *Erwinia psidii* strains are highlighted in blue.

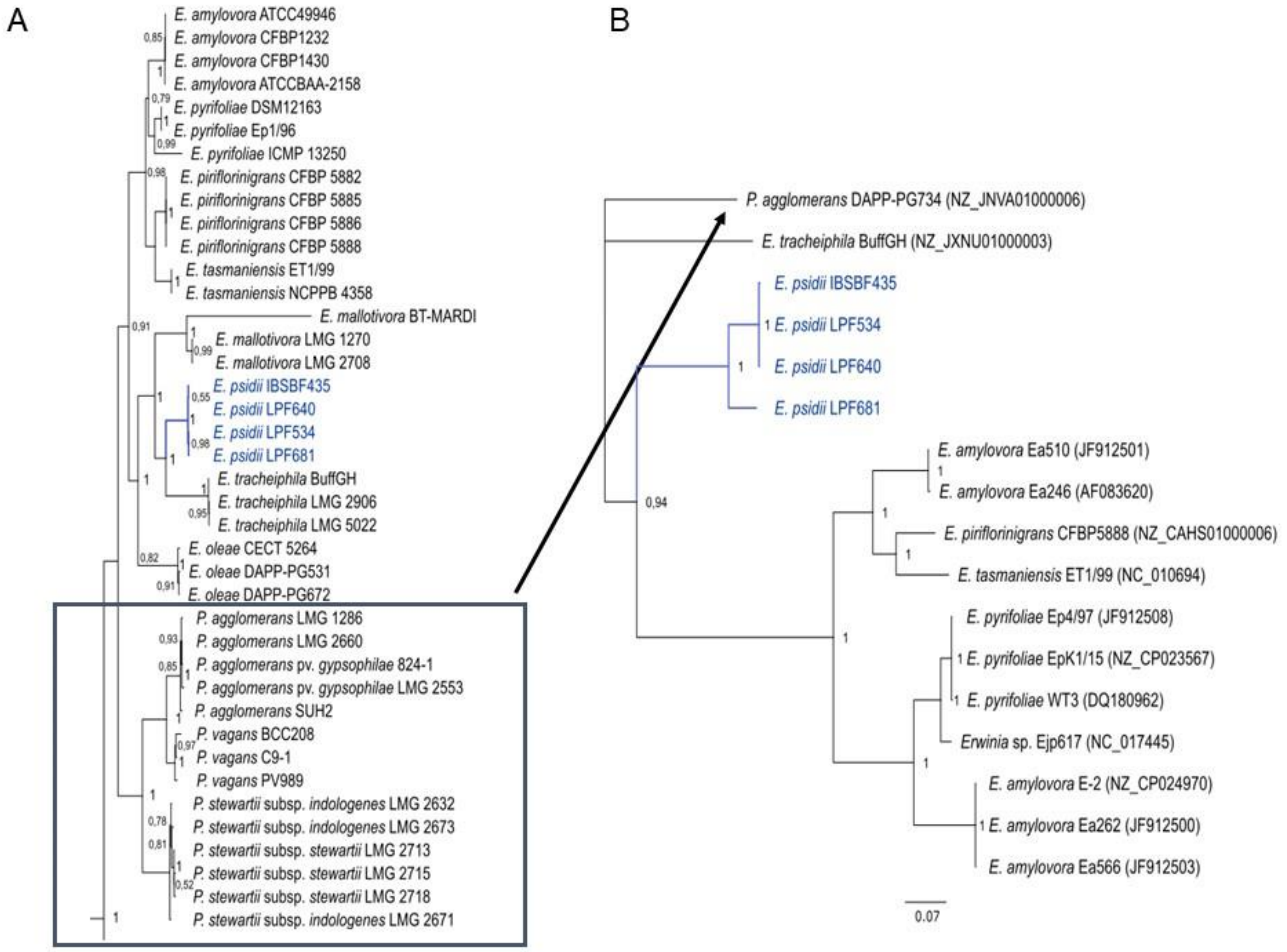


Figure 4. A. Bayesian phylogeny of bacterial species of the *Enterobacteriaceae* family using concatenated partial *gyrB*, *atpD, rpoB a*nd *infB* gene sequences **B.** Bayesian phylogenetic analysis of the *eop1* nucleotide sequences of *Erwinia* spp. and *Pantoea agglomerans*. Posterior probability values are indicated in each node. GenBank accession numbers are indicated in parentheses after strain identification. Bar denotes the rate of substitutions per site. *Erwinia psidii* strains are highlighted in blue. The position of *Pantoea s*pecies are detached by the arrow and the square.

Figure 5. Alignment of Eop1 amino acid sequences from *Erwinia psidii* strains with *E. amylovora* Ea246, *Erwinia* sp. Ejp617, *E. pyrifoliae* WT3 and *E. tasmaniensis* ET1/99. Alignments were performed using Kalign. GenBank accession numbers are indicated within parentheses. Amino acid residues of the catalytic triad are highlighted in blue. Numbers above the alignment indicate the amino acid position with respect to *E. psidii* IBSBF435 sequence. Asterisk (*) represents conserved amino acid residues among aligned sequences.

Table 1. *Erwinia psidii* strains used in this study. Adapted from Montoya-Estrada et al. (2018).

Strain	**Host**	**Geographic origin**	**Aggressiveness on *E. urophylla* × *E. grandis* (CLR375)**	**Aggressiveness on *E. urophylla* (CLR440)**
IBSBF435*	*Psidium guajava*	Brazil: Sâo Paulo, Valinhos	Highly Aggressive	Highly aggressive
LPF681	*P. guajava*	Brazil: São Paulo, Vista Alegre do Alto	Weakly aggressive	Non virulent
LPF534	*Eucalyptus dunnii*	Brazil: Rio Grande do Sul, Guaíba	Aggressive	Aggressive
LPF640	*E. urophylla*	Brazil: São Paulo, Conceição	Aggressive	Weakly aggressive

* Type-strain

Table 2. Predicted subcellular localization of *Erwinia psidii* AvrE effector protein.

	Wolfpsort							PsortII
Strains	Nu	Mi	Ch	Cy	Pm	Ck	Nm	
IBSBF435	11	0	0	1	2	0	7.5	Nu
LPF534	4	5	3	1	0	1	0	Nu
LPF640	4	5	3	1	0	1	0	Nu
LPF681	7	0	0	5	0	2	0	Nu

Nu - nucleus; Mi – mitochondria; Ch – chloroplast; Cy – cytoplasm; Pm – plasma membrane; Ck – cytoeskeleton; Nm – nuclear membrane.

Table 3. Calculation of the distance of *avrE* genes between strains present in the same monophyletic group as *Erwinia psidii* strains. The pairwise distance value is shown. A total of 3,372 nucleotide positions were included in the final dataset.

Strains	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
1. *E. psidii* IBSBF435													
2. *E. psidii* LPF534	0.002												
3. *E. psidii* LPF640	0.000	0.001											
4. *E. psidii* LPF681	0.001	0.002	0.001										
5. *P. agglomerans* pv. *gypsophilae* Ehg824-1	0.289	0.289	0.289	0.290									
6. *P.* subsp. *stewartii* DC283	0.310	0.311	0.310	0.310	0.240								
7. *P. agglomerans* FDAARGOS_407	0.225	0.226	0.225	0.225	0.251	0.288							
8. *E. mallotivora* MT-MARDI	0.222	0.223	0.222	0.223	0.248	0.280	0.053						
9. *P. vagans* FDAARGOS_160	0.228	0.229	0.228	0.228	0.255	0.288	0.085	0.074					
10. *P. agglomerans* DAPP-PG734	0.226	0.227	0.226	0.226	0.256	0.285	0.081	0.069	0.037				
11. *E. oleae* DAPP-PG531	0.226	0.226	0.226	0.226	0.256	0.284	0.081	0.068	0.036	0.000			
12. *E. tracheiphila* BuffGH	0.283	0.284	0.283	0.283	0.337	0.347	0.279	0.277	0.280	0.281	0.281		
13. *E. tracheiphila* PSU-1	0.283	0.284	0.283	0.283	0.337	0.347	0.279	0.277	0.280	0.281	0.281	0.000	

Table 4. Selection pressure on individual codons of *avrE* genes from different strains of *Erwinia* species.

Method	Number of codons	Global dN/dS	Neutral sites*	Positively selected sites*	Negatively selected sites*	Codons
SLAC	1,530	0.561	1521 (99.4%)	0	9 (0.6%)	53, 232, 376, 465, 923, 1148, 1195, 1246, 1265
FEL	1,530		1530 (100%)	0	0	0
IFEL	1,530		1,469 (96%)	0	61 (3,9%)	53, 62, 63, 76, 84, 85, 173, 183, 187, 240, 256, 287, 307, 309, 314, 320, 465, 503, 525, 544, 551, 568, 576, 618, 637, 727, 736, 758, 780, 799, 813, 817, 821, 923, 950, 952, 958, 959, 1023, 1035, 1050, 1080, 1087, 1101, 1118, 1148, 1157, 1184, 1201, 1220, 1265, 1285, 1301, 1303, 1311, 1327, 1368, 1369, 1372, 1411, 1417

*Significant values: $p < 0.05$.

Table 5. Predicted subcellular localization of the *Erwinia psidii* Eop1 effector protein

	Wolfpsort							PsortII
Strains	**Nu**	**Mi**	**Ch**	**Cy**	**Pm**	**Ck**	**Nm**	
IBSBF435	2	7	5	0	0	0	0	Nu
LPF534	2	7	5	0	0	0	0	Nu
LPF640	2	7	5	0	0	0	0	Nu
LPF681	4	5	5	0	0	0	0	Nu

Nu - nucleus; Mi – mitochondria; Ch – chloroplast; Cy – cytoplasm; Pm – plasma membrane; Ck – cytoeskeleton; Nm – nuclear membrane.

Table 6. Calculation of the distance of *eop1* genes between *Erwinia* and *Pantoea* species. The pairwise distance value is shown. A total of 1,246 nucleotide positions were included in the final dataset.

Strains	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1. *E. psidii* IBSBF435																
2. *E. psidii* LPF534	0.001															
3. *E. psidii* LPF640	0.001	0.000														
4. *E. psidii* LPF681	0.089	0.088	0.088													
5. *P. agglomerans* DAPP-PG734	0.290	0.290	0.290	0.285												
6. *E. tracheiphila* BuffGH	0.295	0.294	0.294	0.289	0.297											
7. *E. tasmaniensis* ET1/99	0.341	0.340	0.340	0.348	0.353	0.374										
8. *E. piriflorinigran*s CFBP 5888	0.347	0.346	0.346	0.355	0.343	0.361	0.131									
9. *E. amylovora* Ea246	0.347	0.346	0.346	0.360	0.347	0.365	0.172	0.157								
10. *E. amylovora* Ea510	0.346	0.345	0.345	0.359	0.347	0.363	0.169	0.153	0.003							
11. *E. amylovora* Ea566	0.373	0.372	0.372	0.372	0.370	0.383	0.262	0.257	0.252	0.250						
12. *E. amylovora* Ea262	0.373	0.372	0.372	0.372	0.370	0.383	0.262	0.257	0.252	0.250	0.000					
13. *E. amylovora* E-2	0.373	0.372	0.372	0.372	0.370	0.383	0.262	0.257	0.252	0.250	0.000	0.000				
14. *E. pyrifoliae* WT3	0.363	0.362	0.362	0.363	0.376	0.369	0.239	0.245	0.243	0.242	0.140	0.140	0.140			
15. *E. pyrifoliae* EpK1/15	0.362	0.361	0.361	0.362	0.375	0.368	0.238	0.244	0.243	0.241	0.139	0.139	0.139	0.001		
16. *E. pyrifoliae* Ep4/97	0.362	0.361	0.361	0.362	0.375	0.368	0.238	0.244	0.243	0.241	0.139	0.139	0.139	0.001	0.000	

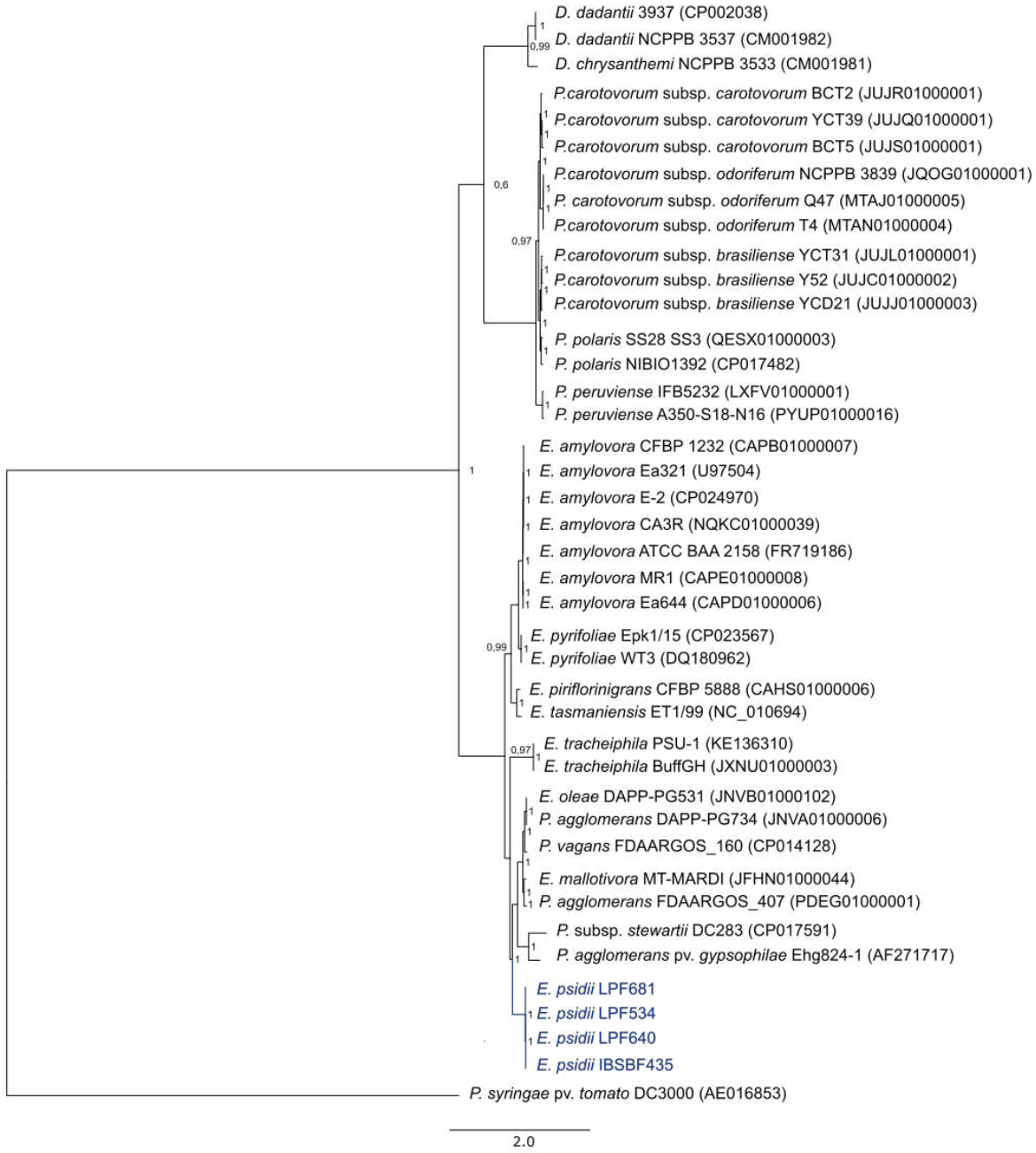
Table 7. Selection pressure acting on individual codons of the e*op1* gene from different species of the genus *Erwinia*.

Method	Number of codons	Global dN/dS	Neutral sites*	Positively selected sites*	Codon	Negatively selected sites*	Codons
SLAC	391	1.04	357 (91.3%)	0		34 (8.7%)	16, 29, 57, 71, 75, 92, 93, 100, 108, 117, 125, 142, 155, 183, 194, 197, 231, 235, 237, 262, 266, 267, 274, 293, 300, 301, 306, 308, 314, 315, 353, 376, 377, 380
FEL	391		281 (71.8%)	1 (0.25%)	30	109 (27.8%)	16, 22, 24, 29, 33, 36, 37, 41, 45, 49, 53, 55, 57, 64, 71, 74, 75, 80, 83, 85, 86, 92, 93, 100, 101, 108, 115, 117, 119, 120, 125, 132, 139, 140, 142, 154, 155, 160, 170, 176, 178, 181, 182, 183, 193, 194, 197, 198, 200, 209, 214, 218, 219, 222, 226, 228, 229, 231, 235, 237, 239, 240, 246, 247, 250, 261, 265, 266, 267, 269, 272, 274, 286, 287, 289, 291, 292, 293, 295, 296, 297, 300, 301, 302, 306, 308, 310, 313, 314, 315, 317, 321, 336, 337, 340, 346, 349, 353, 357, 360, 366, 370, 371, 372,

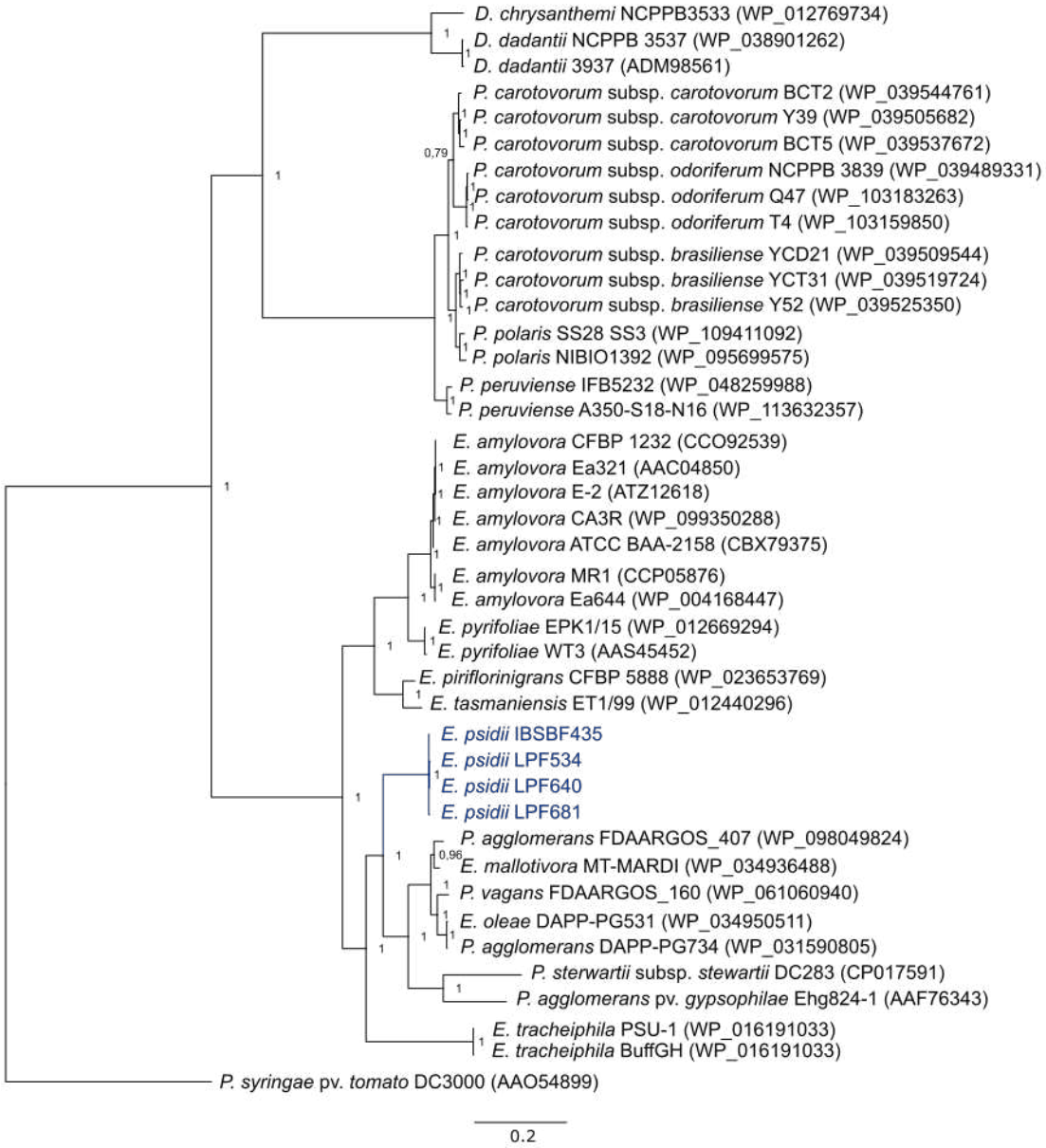
					376, 377, 378, 380, 384
IFEL	391	293 (74.9%)	0	98 (25%)	16, 22, 24, 26, 29, 36, 37, 45, 49, 55, 57, 64, 71, 75, 78, 83, 85, 86, 92, 93, 100, 101, 108, 117, 119, 120, 125, 132, 140, 142, 154, 155, 170, 178, 181, 182, 183, 193, 194, 197, 198, 209, 210, 214, 216, 218, 219, 222, 228, 229, 231, 235, 237, 239, 240, 246, 247, 248, 250, 263, 265, 266, 267, 269, 272, 274, 289, 291, 292, 293, 295, 296, 300, 301, 302, 304, 306, 308, 311, 313, 314, 315, 316, 317, 321, 336, 337, 346, 353, 357, 360, 366, 371, 376, 377, 378, 380, 384

*Significant values: $p < 0.05$.

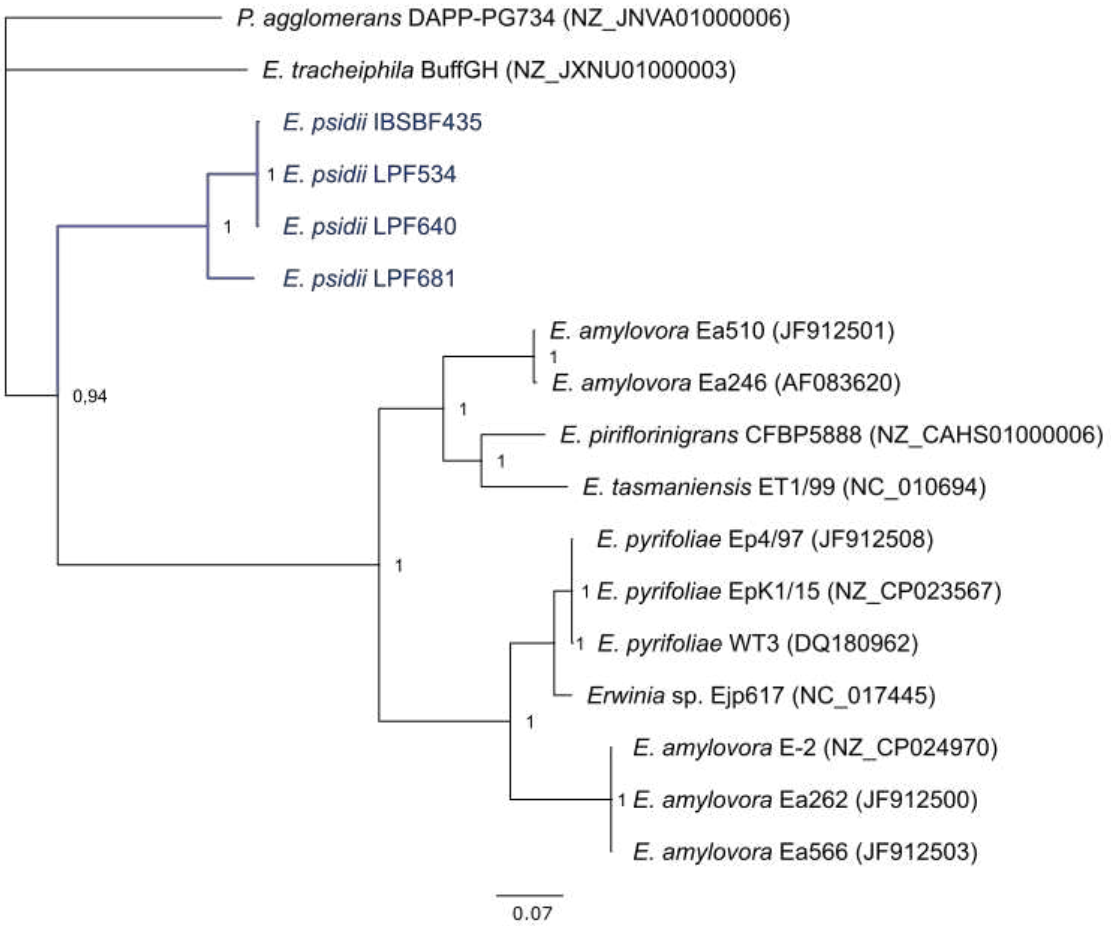
SUPPLEMENTARY INFORMATION



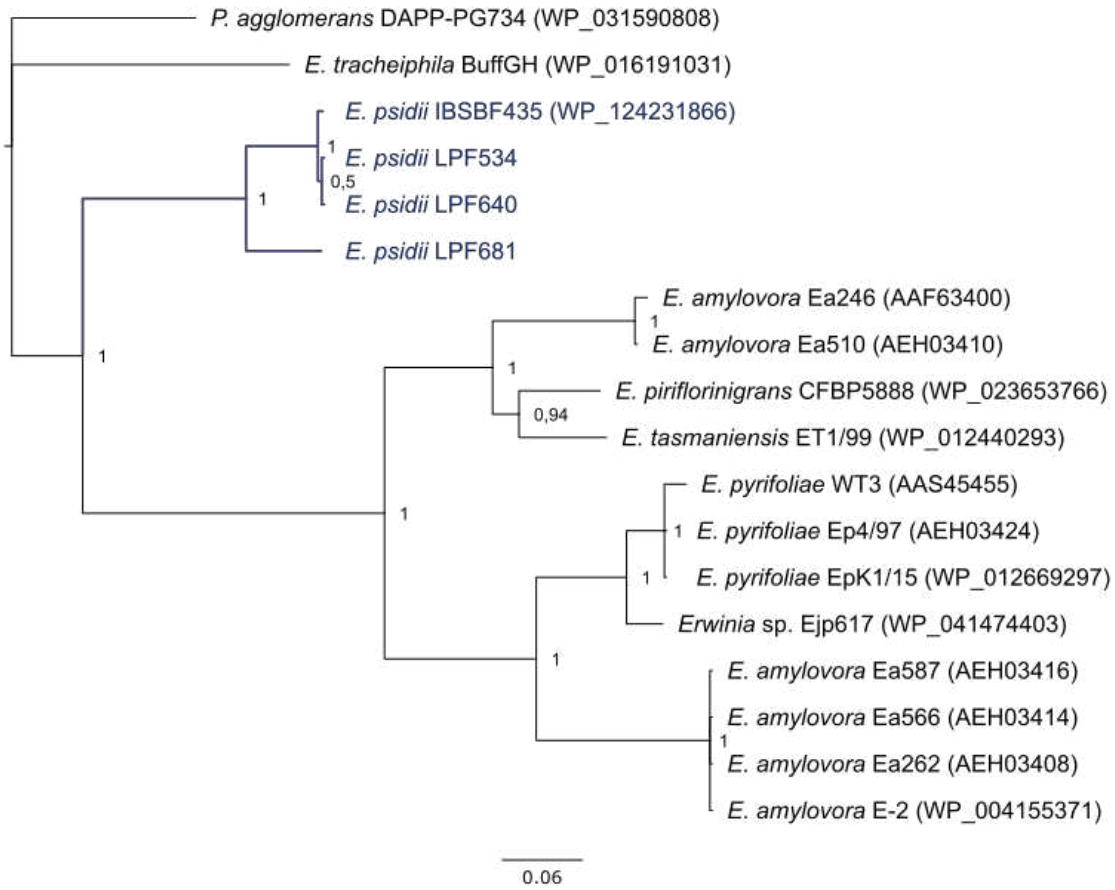
Supplementary Figure S1. Bayesian phylogenetic analysis of the *avrE* coding sequences from bacteria of the *Enterobacteriaceae* family. Posterior probability values are indicated in each node. *Pseudomonas syringae* pv. *tomato* was included as an outgroup. GenBank accession numbers are indicated in parenthesis after strain identification. Bar denotes the rate of substitutions per site. *Erwinia psidii s*trains are highlighted in blue.



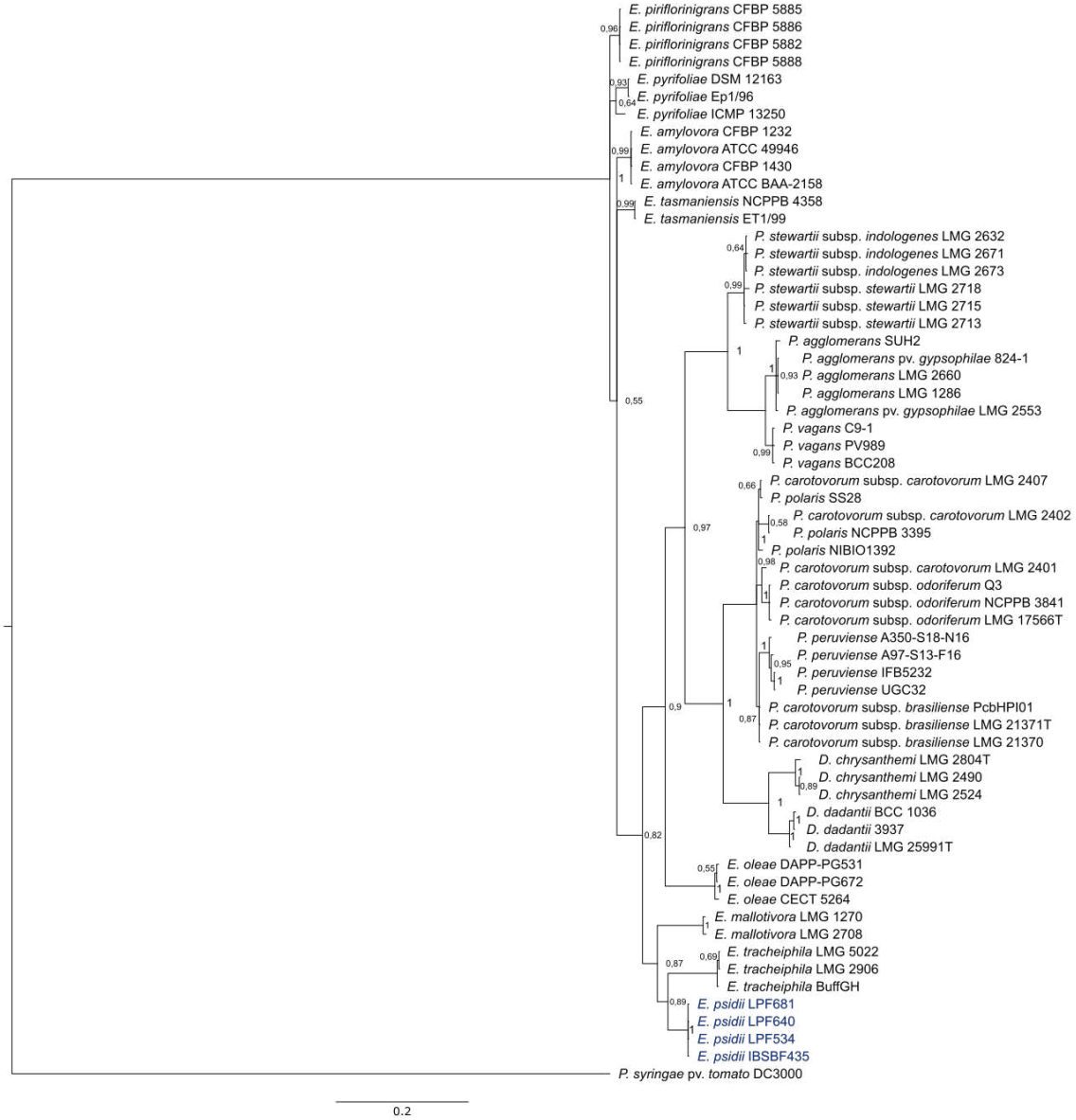
Supplementary Figure S2. Bayesian phylogenetic analysis of the AvrE amino acid sequences from bacteria of the *Enterobacteriaceae* family. Posterior probability values are indicated in each node. *Pseudomonas syringae* pv. *tomato* was included as an outgroup. GenBank accession numbers are indicated in parenthesis after strain identification. Bar denotes the rate of substitutions per site. *Erwinia psidii s*trains are highlighted in blue.



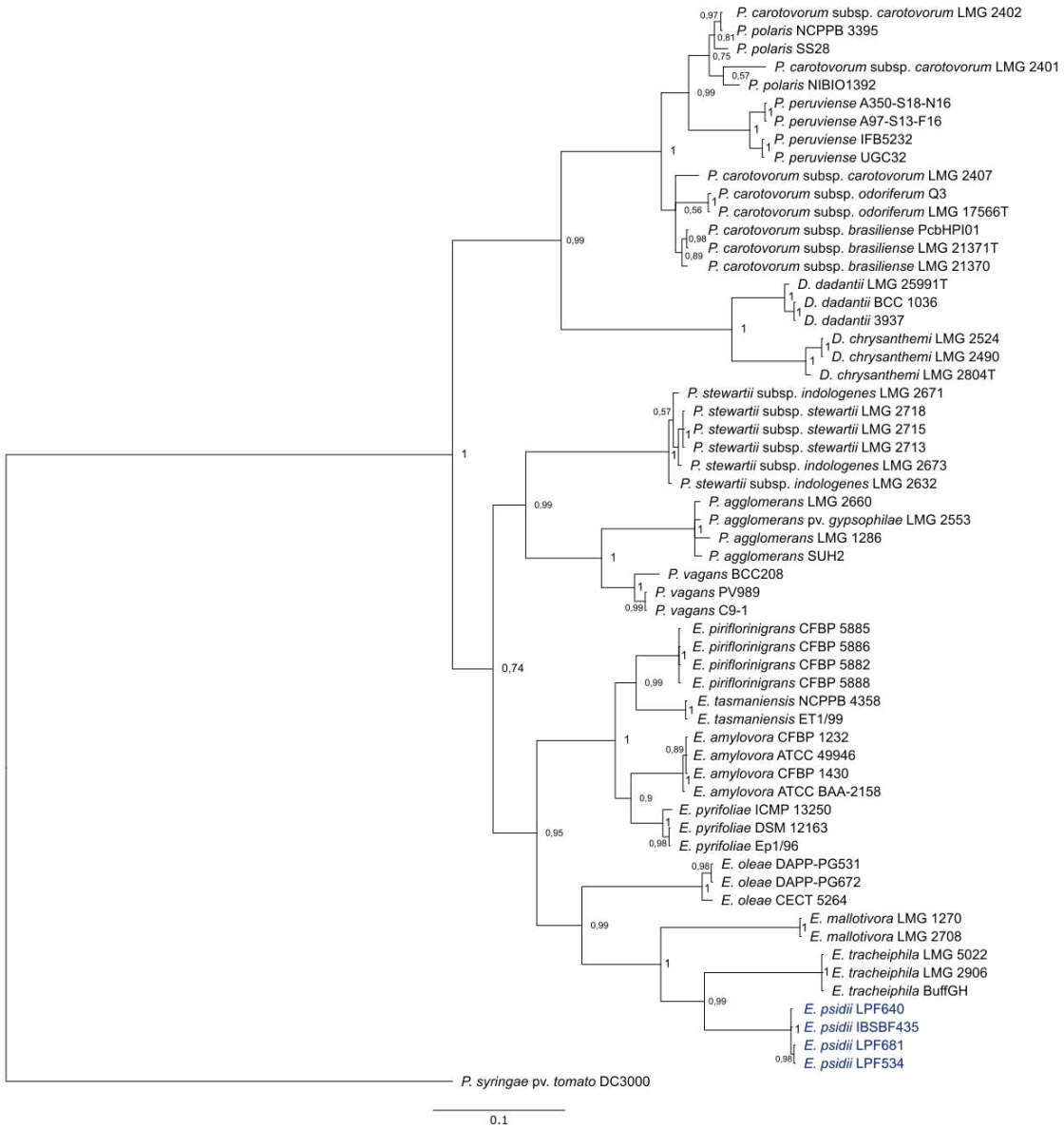
Supplementary Figure S3. Bayesian phylogenetic analysis of the *eop1* nucleotide sequences of *Erwinia* spp. and *Pantoea agglomerans*. Posterior probability values are indicated in each node. GenBank accession numbers are indicated in parentheses after strain identification. Bar denotes the rate of substitutions per site. *Erwinia psidii* strains are highlighted in blue.



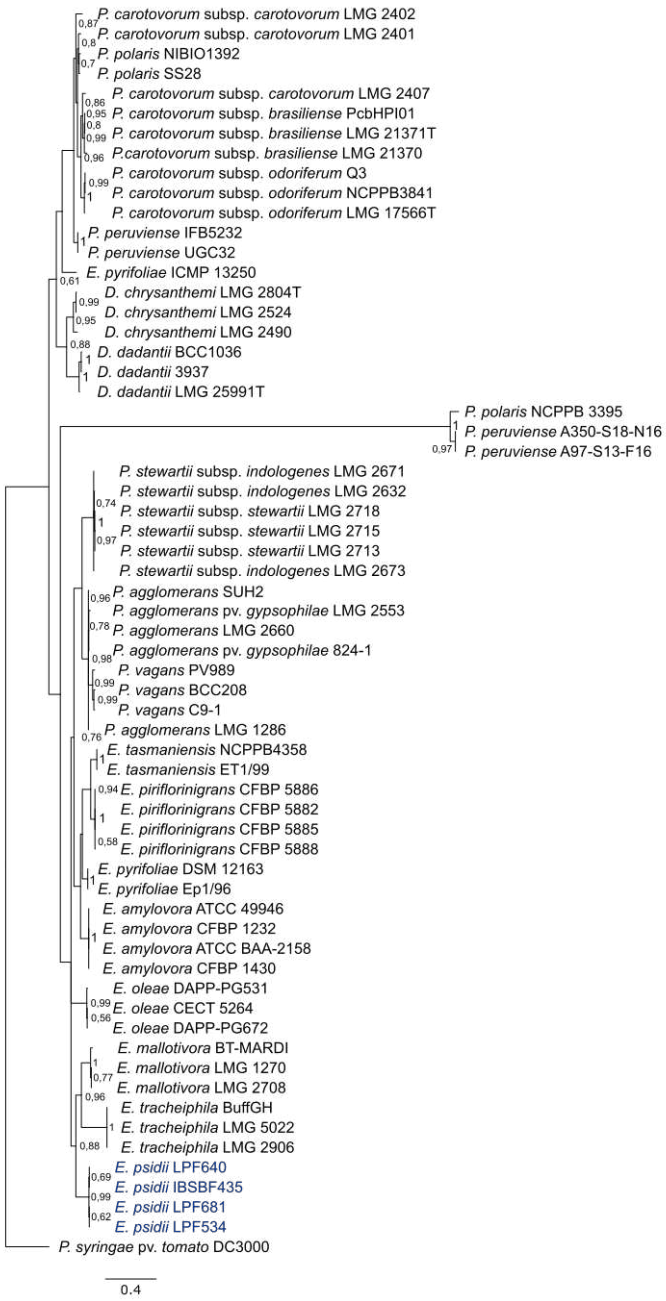
Supplementary Figure S4. Bayesian phylogenetic analysis of Eop1 amino acid sequences of *Erwinia* and *Pantoea* species. Posterior probability values are indicated in each node. GenBank accession numbers are indicated in parentheses after strain identification. Bar denotes the rate of substitutions per site. *Erwinia psidii* strains are highlighted in blue.



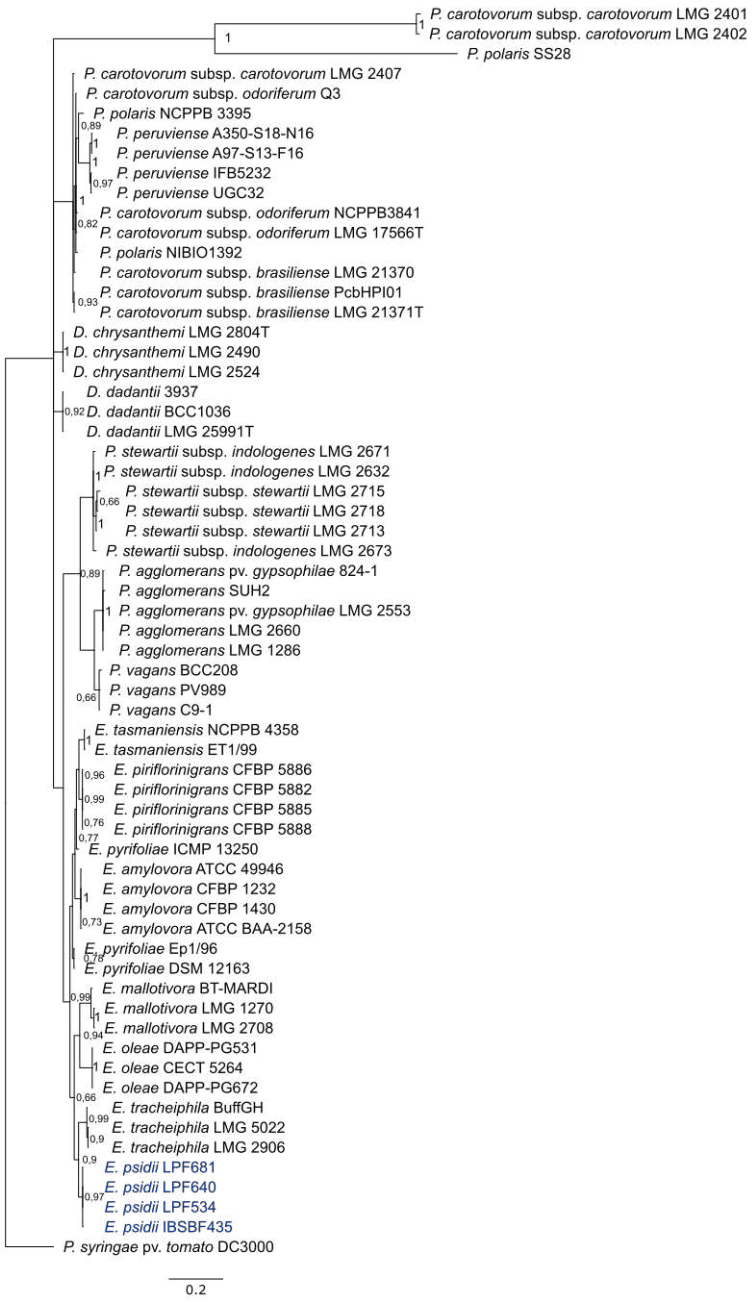
Supplementary Figure S5. Bayesian phylogeny of bacterial species of the *Enterobacteriaceae* family, using partial *atpD* gene sequences. Posterior probability values are indicated on each node. *Pseudomonas syringae* pv. *tomato* was included as an outgroup. *Erwinia psidii* strains are highlighted in blue. Bar denotes the rate of substitutions per site.



Supplementary Figure S6. Bayesian phylogeny of bacterial species of the *Enterobacteriaceae* family, using partial *gyrB* gene sequences. Posterior probability values are indicated in each node. *Pseudomonas syringae* pv. *tomato* was included as an outgroup. *Erwinia psidii* strains are highlighted in blue. Bar denotes the rate of substitutions per site.



Supplementary Figure S7. Bayesian phylogeny of bacterial species of the *Enterobacteriaceae* family, using partial *infB* gene sequences. Posterior probability values are indicated in each node. *Pseudomonas syringae* pv. *tomato* was included as an outgroup. *E. psidii* strains are highlighted in blue. Bar denotes the rate of substitutions per site.



Supplementary Figure S8. Bayesian phylogeny of bacterial species of the *Enterobacteriaceae* family, using partial *rpoB* gene sequences. Posterior probability values are indicated in each node. *Pseudomonas syringae* pv. *tomato* was included as an outgroup. *E. psidii* strains are highlighted in blue. Bar denotes the rate of substitutions per site.

Supplementary Table S1. Strains used for *avrE*-based phylogenetic reconstruction.

Species	Strain	GenBank accession (aa)	GenBank accession (nt)	Geographic location	Isolation source
Dickeya dadantii	NCPPB 3537	WP_038901262.1	NZ_CM001982.1		
	3937	ADM98561.1	CP002038.1		
Dickeya chrysanthemi	NCPPB 3533	WP_012769734.1	NZ_CM001981.1		
Erwinia amylovora	Ea644	WP_004168447.1	NZ_CAPD01000006.1		
	Ea321	AAC04850.1	U97504.1		
	CFBP 1232	CCO92539.1	CAPB01000007.1		
	CA3R	WP_099350288.1	NZ_NQKC01000039.1	USA: California	
	E-2	ATZ12618.1	CP024970.1	Belarus: Minsk area, Myadzel reg.	*Malus* sp.
	Ea266	CCO81193.1	CAOY01000005.1		
	ATCC BAA-2158	CBX79375.1	FR719186.1		
	MR1	CCP05876.1	CAPE01000008.1		
	MT-MARDI	WP_034936488.1	NZ_JFHN01000044.1	Malaysia	*Carica papaya*
Erwinia oleae	DAPP-PG531	WP_034950511.1	NZ_JNVB01000102.1	Italy: Perugia	Olive knot caused by *Pseudomonas savastanoi* pv. *savastanoi*
Erwinia piriflorinigrans	CFBP 5888	WP_023653769.1	NZ_CAHS01000006.1		
Erwinia psidii	IBSBF435	WP_124232032.1	NZ_RHHM01000002.1	Brazil: Sao Paulo, Valinhos	*Psidium guajava*
	LPF534			Brazil: Rio Grande do Sul, Guaiba	*Eucalyptus dunnii*
	LPF640			Brazil: Sao Paulo, Conceicao	*Eucalyptus* sp.
	LPF681			Brazil: Sao Paulo, Vista Alegre do Alto	*Psidium guajava*
Erwinia pyrifoliae	WT3	AAS45452.1	DQ180962.2		
	Epk1/15	WP_012669294.1	NZ_CP023567.1	South Korea	*Malus* sp.
Erwinia tasmaniensis	ET1/99	WP_012440296.1	NC_010694.1		
Erwinia tracheiphila	PSU-1	WP_016191033.1	NZ_KE136310.1		
	BuffGH	WP_016191033.1	NZ_JXNU01000003.1	USA: Pennsylvania	*Cucurbita foetidissima*
Pantoea agglomerans	DAPP-PG734	WP_031590805.1	NZ_JNVA01000006.1	Italy:Bari	Olive knot caused by *Pseudomonas savastanoi* pv. *savastanoi*
	FDAARGOS_407	WP_098049824.1	NZ_PDEG01000001.1	Missing	*Mangifera indica*
P. agglomerans pv. *gypsophilae*	Ehg824-1	AAF76343.1	AF271717.1		
Pantoea stewartii stewartii	DC283	WP_006121853.1	NZ_CP017591.1	USA: Blacksburg, VA	Pure culture
Pantoea vagans	FDAARGOS_160	WP_061060940.1	NZ_CP014128.2	USA:DC	Wound swab
	YC D21	WP_039509544.1	NZ_JUJJ01000003.1	China: Beijing	*Brassica rapa* ssp. *chinensis*

Pectobacterium carotovorum subsp. *brasiliense*	Y52	WP_039525350.1	NZ_JUJC01000002.1	China: Beijing	*Brassica rapa* ssp. *chinensis*
	YC T3	WP_039519724.1	NZ_JUJL01000001.1	China: Beijing	*Brassica rapa* ssp. *chinensis*
	CFIA1001	WP_039286320.1	NZ_JPSM01000022.1	Canada: Alberta	*Solanum tuberosum*
	CFIA1009	WP_039467402.1	NZ_JPSN01000037.1	Canada: Alberta	*Solanum tuberosum*
	ICMP 19477	KMK82535.1	ALIU01000006.1	New Zealand	*Solanum tuberosum*
	YC D49	WP_039518312.1	NZ_JUJB01000023.1	China: Beijing	*Brassica rapa* ssp. *chinensis*
	Y64	WP_040043269.1	NZ_JUJE01000001.1	China: Beijing	*Brassica rapa* ssp. *chinensis*
	YC D65	WP_040035778.1	NZ_JUJN01000018.1	China: Beijing	*Brassica rapa* ssp. *chinensis*
	YC T3	WP_014915532.1	NZ_JUJM01000003.1	China: Beijing	*Brassica rapa* ssp. *chinensis*
P. carotovorum subsp. *carotovorum*	BC S2	WP_039558455.1	NZ_JUJF01000014.1	China: Beijing	*Brassica rapa* ssp. *pekinensis*
	BC T2	WP_039544761.1	NZ_JUJR01000001.1	China: Beijing	*Brassica rapa* ssp. *chinensis*
	PC1	WP_015840410.1	NC_012917.1		
	ATCC 39048	WP_110163267.1	NZ_QHMC01000008.1	USA: New Jersey	
	RMIT1	GBO50684.1	BGPS01000036.1		
	BC D6	WP_039496420.1	NZ_JUJT01000001.1	China: Beijing	*Brassica rapa* ssp. *pekinensis*
	BC T5	WP_039537672.1	NZ_JUJS01000001.1	China: Beijing	*Brassica rapa* ssp. *pekinensis*
	YC D46	WP_039550546.1	NZ_JUJH01000007.1	China: Beijing	*Brassica rapa* ssp. *chinensis*
	YC D57	WP_040031508.1	NZ_JUJG01000004.1	China: Beijing	*Brassica rapa* ssp. *chinensis*
	Y39	WP_039505682.1	NZ_JUJQ01000001.1	China: Beijing	*Brassica rapa* ssp. *chinensis*
P. carotovorum subsp. *odoriferum*	NCPPB 3839	WP_039489331.1	NZ_JQOG01000001.1	France	
	Q106	WP_103199136.1	NZ_MTAK01000013.1	China: Beijing	
	Q47	WP_103183263.1	NZ_MTAJ01000005.1	China: Beijing	
	T4	WP_103159850.1	NZ_MTAN01000004.1	China: Beijing	
Pectobacterium peruviense	A350-S18-N16	WP_113632357.1	NZ_PYUP01000016.1	France: Bleone, Durance River	Fresh water from alpine river
	IFB5232	WP_048259988.1	NZ_LXFV01000001.1	USA: Arizona	S*olanum tuberosum*
Pectobacterium polaris	NIBIO1392	WP_095699575.1	NZ_CP017482.1	Norway	S*olanum tuberosum*
	SS28 SS3	WP_109411092.1	NZ_QESX01000003.1	Pakistan: Punjab	
Pseudomonas syringae pv *tomato*	DC300	AAO54899.1	AE016853.1		

Supplementary Table S2. Strains used for construction of the *eop1*-based phylogenetic tree.

Species	Strain	GenBank accession (aa)	GenBank accession (nt)	Geographic location	Isolation source
Erwinia psidii	IBSBF435	WP_124231866.1	NZ_RHHM01000002.1	Brazil: Sao Paulo, Valinhos	*Psidium guajava*
	LPF534	EHW64_05680		Brazil: Rio Grande do Sul, Guaiba	*Eucalyptus dunnii*
	LPF640	EHW65_10150		Brazil: Sao Paulo, Conceicao	*Eucalyptus* sp.
	LPF681	EHW66_03355		Brazil: Sao Paulo, Vista Alegre do Alto	*Psidium guajava*
Erwinia amylovora	Ea510	AEH03410.1	JF912501.1	Canada: Alberta	*Rubus idaeus*
	Ea587	AEH03416.1	JF912504.1	USA: Idaho	*Prunus salicina*
	Ea566	AEH03414.1	JF912503.1	Germany	*Fragaria × ananassa*
	Ea262	AEH03408.1	JF912500.1	Netherlands	Unknown
	E-2	WP_004155371.1	NZ_CP024970.1	Belarus: Minsk area, Myadzel reg.	*Malus* sp.
	Ea246	AAF63400.1	AF083620.1	USA: Illinois	*Rubus idaeus*
Erwinia piriflorinigrans	CFBP 5888	WP_023653766.1	NZ_CAHS01000006.1		
Erwinia tasmaniensis	ET1/99	WP_012440293.1	NC_010694.1		
Erwinia sp.	Ejp617	WP_041474403.1	NC_017445.1		
Erwinia pyrifoliae	DSM 12163	CAY75696.1	FN392235.1		
	WT3	AAS45455.1	DQ180962.2		
	Ep4/97	AEH03424.1	JF912508.1	South Korea	*Pyrus pyrifolia*
	EpK1/15	WP_012669297.1	NZ_CP023567.1	South Korea	*Malus* sp.
Pantoea agglomerans	DAPP-PG734	WP_031590808.1	WP_031590808.1	Italy:Bari	Olive knot caused by *Pseudomonas savastanoi* pv. *savastanoi*
Erwinia tracheiphila	BuffGH	WP_016191031.1	NZ_JXNU01000003.1	USA: Pennsylvania	*Cucurbita foetidissima*

Supplementary Table S3. Strains used to build the MLSA-based tree indicating the NCBI accession numbers for each housekeeping genes.

Species	Strain	*gyrB*	*atpD*	*rpoB*	*infB*
Erwinia amylovora	CFBP1430	FN434113.1	CAPB01000042.1	FN434113.1	FN434113.1
	ATCC 49946	FN666575.1	FN666575.1	FN666575.1	FN666575.1
	CFBP 1232	FJ617419.1	FR719198.1	CAPB01000004.1	CAPB01000005.1
	ATCC BAA-2158	FR719198.1	FN434113.1	FR719184.1	FR719186.1
Erwinia oleae	DAPP-PG 672	HM439617.1	HM439616.1	HM439619.1	HM439618.1
	CECT 5264	HM439613.1	HM439612.1	HM439615.1	HM439614.1
	DAPP-PG 531	GU991654.1	GU991653.1	GU991656.1	GU991655.1
Erwinia piriflorinigrans	CFBP 5888	CAHS01000023.1	JF311469.1	JF311808.1	JF311695.1
	CFBP 5885	JF311584.1	JF311471.1	JF311810.1	JF311697.1
	CFBP 5882	JF311583.1	JF311470.1	JF311809.1	JF311696.1
	CFBP 5886	JF311585.1	JF311472.1	JF311811.1	JF311698.1
Erwinia pyrifoliae	DSM 12163	FN392235.1	CP023567.1	FN392235.1	FN392235.1
	Ep1/96	FP236842.1	FP236842.1	FP236842.1	FP236842.1
	ICMP 13250	JF311575.1	JF311462.1	JF311801.1	JF311688.1
Erwinia tasmaniensis	ET1/99	CU468135.1	CU468135.1	CU468135.1	CU468135.1
	NCPPB 4358	HQ393607.1	HQ393595.1	HQ393631.1	HQ393619.1
Erwinia tracheiphila	BuffGH	JXNU01000003.1	JXNU01000003.1	JXNU01000003.1	JXNU01000003.1
	LMG 2906	HQ393603.1	HQ393591.1	HQ393627.1	HQ393615.1
	LMG 5022	HQ393604.1	HQ393592.1	HQ393628.1	HQ393616.1
Erwinia mallotivora	LMG 2708	HQ393601.1	HQ393589.1	HQ393625.1	HQ393613.1
	BT-MARDI	NZ_JFHN01000056.1	*	JFHN01000024.1	JFHN01000045.1
	LMG 1270	HQ393602.1	HQ393590.1	HQ393626.1	HQ393614.1
Dickeya dadantii	3937	CP002038.1	CP002038.1	CP002038.1	CP002038.1
	LMG 25991T	JF311644.1	JF311531.1	JF311869.1	JF311757.1
	BCC 1036	JF311647.1	JF311534.1	JF311872.1	JF311760.1
Dickeya chrysanthemi	LMG 2524	JF311639.1	JF311526.1	JF311864.1	JF311752.1
	LMG 2490	JF311637.1	JF311637.1	JF311862.1	JF311750.1
	LMG 2804T	JF311636.1	JF311523.1	JF311861.1	JF311749.1
Pectobacterium carotovorum subsp. *carotovorum*	ATCC 39048	NZ_QHMC01000018.1	NZ_QHMC01000025.1	NZ_QHMC01000025.1	NZ_QHMC01000010.1
	NCPPB 312	NZ_JQHJ01000010.1	NZ_JQHJ01000010.1	NZ_JQHJ01000018.1	NZ_JQHJ01000001.1
	RMIT1	NZ_BGPS01000048.1	NZ_BGPS01000040.1	NZ_BGPS01000047.1	NZ_BGPS01000007.1
	LMG 2407	JF311604.1	JF311491.1	JF311829.1	JF311717.1
	LMG 2401	JF311602.1	JF311489.1	JF311489.1	JF311715.1
	LMG 2402	JF311603.1	JF311490.1	JF311490.1	JF311716.1
P. carotovorum subsp. *brasiliense*	LMG 21370	JF311606.1	JF311493.1	JF311831.1	JF311719.1
	LMG 21371T	JF311605.1	JF311492.1	JF311830.1	JF311718.1
	PcbHPI01	LKKQ01000067.1	NZ_LKKQ01000067.1	LKKQ01000006.1	LKKQ01000056.1
	NCPPB3841	*	NZ_JQOF01000025.1	NZ_JQOF01000021.1	NZ_JQOF01000003.1

P. carotovorum subsp. *odoriferum*	Q3	NZ_MTAG01000018.1	NZ_MTAG01000006.1	NZ_MTAG01000034.1	NZ_MTAG01000002.1
	LMG 17566T	JF311607.1	JF311494.1	JF311832.1	JF311720.1
Pectobacterium peruviense	UGC32	AODU01000021.1	AODU01000022.1	AODU01000003.1	AODU01000005.1
	A97-S13-F16	NZ_PYUO01000013.1	NZ_PYUO01000015.1	NZ_LXFV01000037.1	NZ_PYUO01000007.1
	IFB5232	LXFV01000013.1	NZ_LXFV01000036.1	NZ_PYUO01000023.1	LXFV01000006.1
	A350-S18-N16	NZ_PYUP01000020.1	NZ_PYUP01000005.1	NZ_PYUP01000035.1	NZ_PYUP01000010.1
Pectobacterium polaris	NIBIO1392	NZ_CP017482.1	NZ_CP017482.1	NZ_CP017482.1	NZ_CP017482.1
	SS28	NZ_QESX01000009.1	NZ_QESX01000009.1	NZ_QESX01000016.1	NZ_QESX01000001.1
	NCPPB 3395	NZ_JQHN01000011.1	NZ_JQHN01000035.1	NZ_JQHN01000033.1	NZ_JQHN01000014.1
Pantoea agglomerans pv. *gypsophilae*	LMG 2553	EF988810.1	EF988723.1	EF988982.1	EF988896.1
	824-1	*	NZ_LXSX01000013.1	NZ_LXSX01000036.1	NZ_LXSX01000022.1
P. agglomerans	LMG 2660	EF988823.1	EF988736.1	EF988995.1	EF988909.1
	LMG 1286	FJ617401.1	EF988711.1	EF988970.1	EF988884.1
	SUH2	EF988835.1	EF988748.1	EF989007.1	EF988921.1
Pantoea stewartii subsp. *indologenes*	LMG 2632	EF988822.1	EF988735.1	KF482743.1	EF988908.1
	LMG 2673	EF988827.1	EF988740.1	EF988999.1	EF988913.1
	LMG 2671	EF988826.1	EF988739.1	EF988998.1	EF988912.1
P. stewartii subsp. *stewartii*	LMG 2718	EF988832.1	EF988745.1	EF989004.1	EF988918.1
	LMG 2713	EF988830.1	EF988743.1	EF989002.1	EF988916.1
	LMG 2715	EF988831.1	EF988744.1	EF989003.1	EF988917.1
Pantoea vagans	C9-1	CP002206.1	CP002206.1	CP002206.1	CP002206.1
	PV989	CP028349.1	NZ_CP028349.1	NZ_CP028349.1	NZ_CP028349.1
	BCC208	EF988837.1	EF988750.1	EF989009.1	EF988923.1
Pseudomonas syringae pv. *tomato*	DC3000	AE016853.1	AE016853.1	AE016853.1	AE016853.1

*The lack of accession number denotes lack of suitable sequences in the data bases.

Supplementary Table S4. Gene sequences used for MLSA-based tree construction. Character set explains the order of the genes in the concatenation process. Range values indicate in which positions the sequences were trimmed using as reference *E. psidii* IBSBF435 complete gene sequence.

Gene	Character set	Range (IBSBF435)
atpD	1 - 694	172 - 813
gyrB	695 - 1255	448 - 1145
infB	1256 - 1863	1375 - 1964
rpoB	1864 - 2281	1429 - 2065

Supplementary Table S5. Best fitting nucleotide substitution models for each gene, suggested by MrModeltest 2.3.

Gene	Model	I+G*
avrE	GTR+I+G	0.0473; 0.7158
eop1	GTR+I+G	0.1725; 1.4602
atpD	GTR+I+G	0.1293; 0.2388
gyrB	GTR+I+G	0.1177; 0.3827
infB	GTR+I+G	0.1036; 0.3519
rpoB	HKY+G	0; 0.2674

* I = Proportion of invariable sites; G = Gamma distribution shape parameter for variable sites.

GENERAL CONCLUSION

In conclusion, this study performed the first prediction of genes that might be involved in the pathogenicity of *Erwinia psidii*. Here we present relevant information on the interaction between this bacterium and its host plants, based on the type II and type III secretion systems, two remarkably important virulence factors for closely related bacterial species. Through different bioinformatic approaches, we were able to characterize the composition and arrangement of the *hrp/hrc* gene cluster of *E. psidii*, well-known for its important role of encoding proteins that compose the Type III secretion apparatus. We also predicted the type II and type III effector repertoire of this bacterium, and found evidence of effectors previously described for other bacterial species, as well as novel candidate effectors exclusively present on the genome of *E. psidii*. Additionally, we gained a better understanding on how this bacterium evolved over time in comparison to other Enterobacteriaceae. This result provided further knowledge on how or whether this bacterium might have suffered a host shift. Thus, this study represents an important advance in the knowledge about the pathogenicity mechanisms underlying *E. psidii* ability to cause disease.