

Characterization of the new species *Xanthomonas* *euroxanthea* and comparative pathogenomics to unveil a sympatric lifestyle with *Xanthomonas* *arboricola*

Maria Leonor Ribeiro Martins

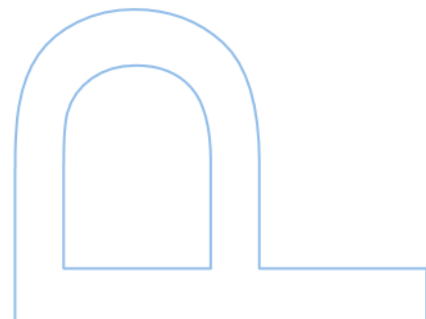
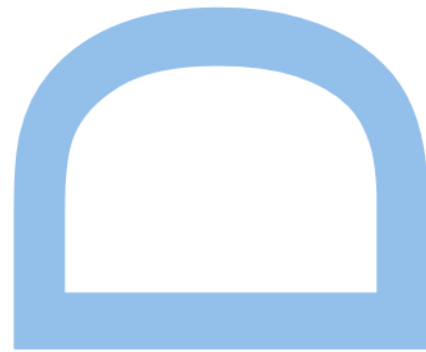
Programa Doutoral em Biologia

Faculdade de Ciências da Universidade do Porto

2023

Supervisor

Fernando Manuel dos Santos Tavares, Professor Associado,
Departamento de Biologia, Faculdade de Ciências da Universidade
do Porto

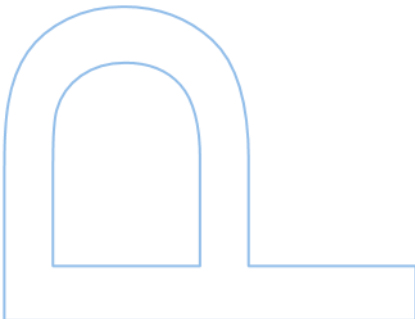
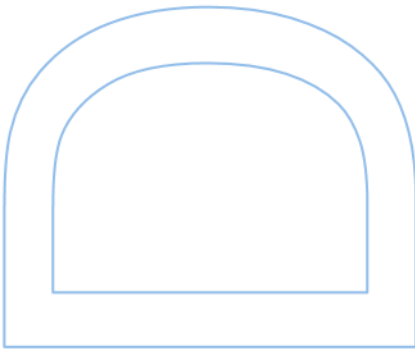


Note: This page should only be included in case the jury the thesis with recommendation of correction of errors or inaccuracies identified and expressly mentioned during the defence.

All the corrections determined by the jury, and only those, were made.

The Supervisor,

Porto, ____/____/____



Characterization of the new species *Xanthomonas euroxanthea* and comparative pathogenomics to unveil a sympatric lifestyle with *Xanthomonas* *arboricola*

The work presented in this thesis was carried out at the *Faculdade de Ciências da Universidade do Porto* (FCUP) and *Microbial Diversity and Evolution Group* (MDE) of Centro de Investigação em Biodiversidade e Recursos Genéticos, InBio Laboratório Associado (CIBIO-InBio); within the scope of the PhD studentship from Fundação para a Ciência e Tecnologia (FCT) SFRH/BD/137079/2018, entitled: "PathoXAJ – *Xanthomonas arboricola* pv. *juglandis* beyond detection: DNA markers of virulence and pathogenicity to make epidemiological risk assessments".



Characterization of the new species *Xanthomonas euroxanthea* and comparative pathogenomics to unveil a sympatric lifestyle with *Xanthomonas* *arboricola*

The work developed in this thesis was financially supported by:

- PhD fellowship from Fundação para a Ciência e a Tecnologia (FCT) SFRH/BD/137079/2018;
- EVOXANT (PTDC/BIA-EVF/3635/2014, POCI-01-0145-FEDER-016600), co-financed by the European Structural & Investment Funds (ESIFs) through the Operational Competitiveness and Internationalization Programme – COMPETE 2020 and by National Funds through Fundação para a Ciência e Tecnologia (FCT);
- European Cooperation in Science and Technology (COST) action CA16107 EuroXanth – “Integrating science on Xanthomonadaceae for integrated plant disease management in Europe”.
- EU’s Horizon 2020 Research and Innovation program through BIOPOLIS (grant agreement 857251).



UNIÃO EUROPEIA
Fundo Social Europeu



“Não é verdade que quanto mais se lê, mais se vê do mundo. Não importa o conhecimento que se enfie na cabeça, a menos que penses com a tua própria mente, caminhes com os teus próprios pés, o conhecimento que adquires será sempre um vazio e nunca te pertencerá.”

— Sōsuke Natsukawa, *O Gato que Salvava Livros*

“It’s not true that the more you read, the more you see of the world. No matter how much knowledge you cram into your head, unless you think with your own mind, walk with your own feet, the knowledge you acquire will never be anything more than empty and borrowed.”

— Sōsuke Natsukawa, *The Cat Who Saved Books*

Acknowledgements

As there is always a team behind every work, this PhD thesis project was no exception. Now is the time to thank everyone that in a more professional or personal way contributed to this thesis or played a significant role on this four-year journey.

Firstly, I would like to thank my supervisor Prof. Fernando Tavares for taking a chance on me when he first accepted me at his lab as a research fellow, introducing me to the phytopathogen universe. Thank you for believing in me and providing many opportunities, guidance, and mentorship from day one.

To Dr. Joel Pothier, Dr. Jochen Blom and Dr. Nay C. Dia for their contribution to the work presented in Chapter II of this thesis. Particularly to Jochen Bloom for the assistance in including the genomes and data analysis on EDGAR platform and Nay for the tireless work in Biolog assay. Thankfully the collaboration on the *X. euroxanthea* proposal as a new bacterial species prompted further collaboration with Dr. Joel Pothier whose brilliant ideas, encouragement, and always fruitful comments, were precious and greatly benefited the work that extends to all the Chapters of this thesis.

To Dr. Nuno Fonseca, Cátia Chaves e Joana Pinto for Nanopore genome sequencing and hybrid assembly. Particularly thank you to Miguel Teixeira for the bioinformatics work, always searching to learn more, finding new tools and ways to answer new research questions and for his important contribution to Chapters III and IV.

To 2.61 lab for the equipment shared and particularly to Prof. José Pissarra for helping in preparation of samples and obtaining ultrathin cuts for TEM. Thank you to SPReD lab for their help with electroporation.

Thank you to IB2 team for the collaboration particularly, Rafael, Sara e Nuno for your help and support. Special thanks to Nuno for support, idea sharing, problem solving and kind words since the masters.

I am also very grateful to Euroxanth COST Action for providing funding that allowed me to participate in scientific conferences, training schools and a scientific mission which was very rewarding for me career wise and at a personal level. Also, its contribution to the discovery, proposal and characterization of the new bacterial species that is the star of this thesis, will forever be immortalized in its name, *Xanthomonas euroxanthea*. Particularly to action's chair Dr. Ralf Koebnik for accepting me in your lab in Montpellier and for the constant patience, guidance and mentoring. And most of all to you and Brigitte

for opening your lovely home to me and always making me feel welcome, for sharing life views, experiences and always giving advice also changing my perspective on life as a scientist. Also, thank you for all the traveling and events suggestions and for all the planning and tour guiding on the hiking trips to the most incredible views and villages!

To XPLAIN team in Montpellier, especially to Dr. Boris Szurek for accepting me at the lab and all the team for the warm welcome, constant availability to help, tips and problem solving. To my colleagues that quickly became friends Marlenita, Angeliki, Carlos, Gabriel and Coline thank you for all the good times, amazing plans, advice and most of all for taking me in and being my family away from home. Merci Aurore, amazing *coloc*, and kindness in person, thanks for your friendship. Merci beaucoup, muchas gracias!

Just on a general note, I have to say that I am truly blessed for my friends, family and colleagues that even if far away were always there for me, I cannot thank you enough.

For the MDE team, thanking you to Sofia, Camila and Pedro for all the support when I was having my first steps at the lab. Further acknowledgments are due to Camila for her contribution in Chapters II and III. To Sofia whose guidance before, when I was starting throughout my time at the lab, and until now were precious. Although you always manage to convince me to do crazy experiments (eheh) we were great lab partners, and I thank you for all the help, motivation and good disposition!! Mafalda Pereira and Mafalda Pinto, and Kayla that not only are brilliant scientists but also amazing and loving people, always positive and ready for a laugh. Thank you, Mafalda Pereira, for your friendship and providing words of comfort and motivation through my insecurities and not so good times. To Kayla for your kindness and always being ready to help!. Also, thank you for the valuable contribution to this thesis concerning chapter V. Thank you Jessy and Maria for lots of memorable moments and lots of laughs.

A big thank you to Kayla, Rita, Helena and Filipe for the care packages that were a great help during the writing of this thesis, thank you for all the kind words, it meant a lot!!

Às minhas incríveis meisqueteiras, Sara e Cristiana, obrigada por todos os planos, risos descontrolados e conversas de tapete... and also for keeping me in the writing schedule!! Obrigada Cristiana pela tua amizade, energia contagiante e por estares sempre pronta a ajudar com palavras de motivação e confiança e por toda a paciência e apoio, principalmente nesta fase final da tese. Sarita, apesar de termos estado em continentes diferentes durante metade da duração desta tese, finalmente conseguimos parar na

mesma cidade, e ainda bem que tivemos esta oportunidade! Obrigada pelo tudo, todos os devaneios, conversas e momentos marcantes que levaremos sempre connosco.

Aos meus pais, obrigada por sempre me incentivarem a seguir o meu rumo e perseguir os meus objetivos e ambições. Obrigada por tudo!

À Helena, Filipe e Lúcia, obrigada pela vossa amizade e apoio constante.

Last but not least, I want to thank my Pedro for his unconditional support, and persistent motivation. As all the acknowledgments would never be enough, I will just say this, thank you for putting up with me no matter what and for always believing me in “*I think we make a real sharp couple of coconuts*”.

Abstract

Xanthomonas arboricola pv. *juglandis* (*Xaj*), is the etiological agent of important walnut diseases causing major yield losses worldwide. An epidemiological assessment of walnut-associated xanthomonads from Portuguese orchards carried out previously at our lab disclosed a high genetic diversity. The genomes of walnut-associated divergent lineages (CPBF 367, CPBF 424^T, CPBF 426) and *Xaj* (CPBF 427) strains isolated from the same isolated tree in Oeiras, Portugal were determined using Illumina sequencing. Average nucleotide identity (ANI) calculations failed to attribute CPBF 367, CPBF 424^T and CPBF 426 to known *Xanthomonas* species. Meaning that these strains belonged to a new species, that we characterized and proposed as *X. euroxanthea*.

As the implications of *Xaj* and *X. euroxanthea* co-colonization were still poorly understood, further whole-genome comparisons were carried out to unveil niche-specific adaptations. Hybrid-assembled genomes from *X. euroxanthea* isolates showed a slight reduction in size as well as CDS number comparatively to *Xaj* strains. Additionally, whole-genome comparisons revealed remarkable genomic differences between *Xaj* and *X. euroxanthea* strains, which translates into different pathogenicity and virulence features, namely regarding T3SS and T3E, other secretory systems, chemotaxis related proteins and extracellular enzymes.

Genomes of four *Xanthomonas* spp. strains isolated from a different host, pecan, were also obtained again confirming co-colonization amongst *X. euroxanthea* (CPBF 761 and CPBF 766 showing ANI > 97%) and *X. arboricola* (CPBF 1494 and CPBF 765). Also, the two pecan-isolated *X. euroxanthea* strains are pathogenic in walnut but not in the host of isolation. The identification of additional *X. euroxanthea* strains revealed that this species is also found in other hosts besides walnut and pecan, such as tomato, bean and ground cherry, usually sharing the host with *X. arboricola*. In an attempt to disclose host-specificity in these species, comparative genomics analysis was carried out between *X. arboricola* and *X. euroxanthea* isolated together from walnut, pecan and tomato for assessing possible variation in genomic traits, with special focus on T3SS and T3E. No major differences were detected, regarding general genomic features, apart from genome size, and indeed the only differences coincided with regions of predicted genomic islands. Although differing in pathogenicity potential and host, some *X. euroxanthea* strains show a similar T3E profile, suggesting a non-host related pattern.

As *X. euroxanthea* comprises both pathogenic and nonpathogenic strains frequently found colonizing the same host plants as the pathogen *X. arboricola*, it was fundamental to develop a PCR method to specifically detect *X. euroxanthea* strains and distinguish

these from the closely related *X. arboricola*. Eight DNA markers (XEA1-XEA8) were selected by *in silico* approaches suggesting that all markers are included in conserved and stable genomic regions. A multiplex PCR targeting markers XEA1, XEA5 and XEA8 was shown to be efficient and specific for seven *X. euroxanthea* strains analysed. The markers genotyping potential was also evaluated revealing that overall, this data provides a method conciliating detection and genotyping of *X. euroxanthea*-strains, contributing to quickly track the presence of *X. euroxanthea* in *X. arboricola*-colonizing habitats.

In summary, this thesis presents the discovery of a new bacterial species, *X. euroxanthea* whose particularities of including pathogenic and non-pathogenic strains (reflected in a distinct presence-absence of virulence-coding genes) and the recurrent isolation with *X. arboricola* in multiple hosts presumes a generalist and sympatric lifestyle with *X. arboricola*. Promoting this species to a particularly interesting model to address pathogenicity and virulence evolution in *Xanthomonas*.

Keywords: *Xanthomonas euroxanthea*, phytopathogen, comparative genomics, pathogenicity, type III secretion system, effectors, detection markers, multiplex PCR, walnut, pecan, pathogen lifestyle, evolution.

Resumo

Xanthomonas arboricola pv. *juglandis* (*Xaj*), é o agente etiológico de importantes doenças da árvore da noqueira que causam grandes perdas de rendimento em todo o mundo. Um anterior estudo epidemiológico de *Xanthomonas* em pomares de noqueiras portuguesas, realizada no nosso laboratório, revelou uma elevada diversidade genética. Os genomas des estirpes de linhagens divergentes (CPBF 367, CPBF 424^T, CPBF 426) e de *Xaj* (CPBF 427) presentes numa única árvore isolada em Oeiras, Portugal foram sequenciados por Illumina. Os cálculos da identidade média dos nucleótidos (ANI) não atribuíram CPBF 367, CPBF 424^T e CPBF 426 a espécies já conhecidas, assim revelando que estas estirpes pertenciam a uma nova espécie, que denominámos *X. euroxanthea*.

Considerando as implicações da co-colonização de *Xaj* e *X. euroxanthea* ainda serem mal compreendidas, foram efectuadas mais comparações genómicas para desvendar adaptações específicas de nichos. Estirpes de *X. euroxanthea* mostraram uma ligeira redução no tamanho, e do número de CDS comparativamente às estirpes de *Xaj*, também revelando diferentes características de patogenicidade e virulência, nomeadamente em relação ao T3SS e T3E, outros sistemas de secreção proteínas relacionadas com quimiotaxia e enzimas extracelulares.

Genomas de quatro estirpes de *Xanthomonas* spp., isolados de um hospedeiro diferente, a noqueira-pecã, foram também obtidos revelando uma vez mais fenómenos de co-colonização entre *X. euroxanthea* (CPBF 761 e CPBF 766, com ANI > 97%) e *X. arboricola* (CPBF 1494 e CPBF 765). As duas estirpes de *X. euroxanthea* provenientes da noqueira-pecã são patogénicas em noqueira mas não no hospedeiro de isolamento. A identificação de estirpes adicionais de *X. euroxanthea* revelou que esta espécie também pode estar presente noutros hospedeiros para além da noqueira e da noqueira-pecã, como o tomate, feijão e fisális, geralmente partilhando o hospedeiro com *X. arboricola*. Numa tentativa de revelar determinantes de especificidade de hospedeiro nestas espécies, foi realizada uma análise genómica comparativa entre as estirpes de *X. arboricola* e *X. euroxanthea* isoladas dos mesmos hospedeiros para avaliar possíveis variações nas características genómicas. Não foram detectadas grandes diferenças no que diz respeito às características genómicas gerais. Embora diferindo no potencial de patogenicidade e hospedeiro, algumas estirpes de *X. euroxanthea* apresentam um perfil T3E semelhante, sugerindo um padrão não relacionado com o hospedeiro.

Como *X. euroxanthea* inclui tanto estirpes patogénicas como não patogénicas, frequentemente encontradas a colonizar o mesmo hospedeiro que estirpes patogénicas

de *X. arboricola*, foi fundamental desenvolver um método PCR para detectar especificamente as estirpes *X. euroxanthea* e distingui-las de *X. arboricola*. Marcadores moleculares XEA1-XEA8 foram seleccionados através de abordagens *em silico*. Análises de sintenia, o conteúdo em GC e os valores de CAI/eCAI sugerem que todos os marcadores se encontram em regiões genómicas conservadas. O PCR multiplex incluindo os marcadores XEA1, XEA5 e XEA8 revelou ser eficiente e específico para estirpes de *X. euroxanthea* contempladas na análise. O potencial de genotipagem dos marcadores também foi avaliado, revelando que globalmente, estes marcadores fornecem um método que concilia a detecção e genotipagem de *X. euroxanthea*, contribuindo para rastrear rapidamente a presença desta bactéria em nichos onde *X. arboricola* também poderá estar presente.

Em resumo, esta tese apresenta a descoberta de uma nova espécie de bactéria, *X. euroxanthea* cujas particularidades de incluir estirpes patogénicas e não patogénicas (reflectidas nos padrões presença-ausência distintos de genes de virulência) e o isolamento recorrente com *X. arboricola* em hospedeiros múltiplos pressupõe um estilo de vida generalista e simprático com *X. arboricola*. O que torna esta espécie um modelo especialmente interessante para o estudo de padrões evolutivos de patogenicidade em *Xanthomonas*.

Palavras-chave: *Xanthomonas euroxanthea*, fitopatógeno, genómica comparativa, patogenicidade, sistemas de secreção tipo III, efetores, marcadores de detecção, PCR multiplex, noqueira, noqueira-pecã, ciclo de vida patógeno, evolução.

Table of Contents

Acknowledgements	xix
Abstract.....	xxiii
Resumo	xxv
List of Tables	xxxii
List of Figures.....	xxxii
List of Abbreviations	xxxvii
Chapter I	39
General Introduction	41
1. <i>Xanthomonas</i> , a diverse group of phytopathogenic bacteria	41
1.1. <i>The winding road of Xanthomonas taxonomical reclassification.....</i>	<i>41</i>
1.2. <i>Describing a new bacterial species</i>	<i>43</i>
1.3. <i>The particular case of Xanthomonas arboricola.....</i>	<i>43</i>
1.4. <i>Non-pathogenic bacteria conquer their place - Xanthomonas pathogenic and commensal dynamics</i>	<i>44</i>
2. Recent advances in genomics tools prompted a BOOM on <i>Xanthomonas</i> genomics analysis	45
2.1. <i>Improvement of sequencing technologies.....</i>	<i>45</i>
2.2. <i>Genomics-facilitated taxonomic re-classification.....</i>	<i>46</i>
2.3. <i>Using comparative genomics to unlock lifestyle traits and DNA marker identification.....</i>	<i>47</i>
3. Mechanisms of virulence and pathogenicity in <i>Xanthomonas</i> spp.....	48
3.1. <i>Secretion systems are essential for Xanthomonas survival.....</i>	<i>50</i>
3.2. <i>A closer look into Type III secretion system cluster and effectors</i>	<i>51</i>
3.3. <i>Evolution of T3SS and effectors in Xanthomonas sp.....</i>	<i>53</i>
Framework and goals of this thesis	54
Thesis publications.....	55
References	56
Chapter II	69
Proposal of <i>Xanthomonas euroxanthea</i> sp. nov., a new xanthomonad species including pathogenic and non-pathogenic strains of walnut	71
Abstract.....	71
Introduction	72
Isolation and growth of bacteria.....	73
Pathogenicity assays	76

Phenotypic characterization.....	77
Multilocus sequence analysis	80
Genomic characterization	81
Type III Secretion System (T3SS) and Type III Effectors (T3E).....	84
Description of <i>Xanthomonas euroxanthea</i> sp. nov.	85
References.....	87
Supplementary material.....	91
Chapter III	93
Comparative Genomics of <i>Xanthomonas euroxanthea</i> and <i>Xanthomonas arboricola</i> pv. <i>juglandis</i> Strains Isolated from a Single Walnut Host Tree	
Abstract	95
1. Introduction	95
2. Materials and Methods	98
2.1. <i>Xanthomonas</i> Strains Used in This Study and Genome Sequencing	98
2.2. Average Nucleotide Identity.....	99
2.3. Comparative Genome Analysis.....	99
2.4. Homologous of Pathogenicity and Virulence-Associated Proteins Inferred by tBLASTn Analysis.....	99
3. Results	100
3.1. General Features of <i>X. euroxanthea</i> and <i>X. arboricola</i> pv. <i>juglandis</i> Genome Assemblies	100
3.2. Genomic Distance Assessed by ANI and Phylogenetic Analysis ...	102
3.3. Genetic Patrimony Retrieved from the <i>X. euroxanthea</i> and <i>X.</i> <i>arboricola</i> pv. <i>juglandis</i> strains Isolated from a Single Walnut Host Tree ..	107
3.4. Pathogenic and Virulence-Related Factor Prediction	108
3.5. Type 3 Secretion System and Its Effectors	110
4. Discussion	110
5. Conclusions	114
References.....	116
Supplementary material.....	123
Chapter IV.....	125
Genomics of <i>Xanthomonas arboricola</i> and <i>X. euroxanthea</i>: two phytopathogens, same plant host species, similar pathoadaptations	
Abstract	127
1. Introduction	127
2. Methods	129
2.1. Genome sequencing	129

2.2.	<i>Xanthomonas</i> genomes used in this study and taxonomic affiliation determined by the Average Nucleotide Identity.....	129
2.3.	Phylogenetic tree of core genome.....	131
2.4.	General genomic characteristics	131
2.5.	Pathogenicity assays and hypersensitive response (HR)	131
2.6.	Pathogenicity and virulence factors prediction	133
3.	Results	133
3.1.	<i>Xanthomonas</i> sp. isolates are evidenced as belonging to <i>X. euroxanthea</i> species	133
3.2.	Phylogenomics tree denotes a cluster with <i>X. euroxanthea</i> strains isolated from three different hosts	135
3.3.	<i>X. euroxanthea</i> CPBF 761 and CPBF 766 are pathogenic on walnut and nonpathogenic on pecan	135
3.4.	General genomic characteristics of <i>X. euroxanthea</i> and <i>X. arboricola</i> from different hosts	137
3.5.	<i>X. euroxanthea</i> type III effector (T3E) profiles follow a non-host related pattern.....	138
4.	Discussion	142
	References.....	146
	Supplementary material	152
Chapter V.....		159
	Abstract.....	161
1.	Introduction	161
2.	Materials and Methods	163
2.1.	<i>In silico</i> selection and validation of <i>X. euroxanthea</i> -specific DNA markers.....	163
2.2.	Bacterial strains, culture conditions and DNA extraction	165
2.3.	Experimental validation of putative <i>X. euroxanthea</i> -specific DNA markers by multiplex PCR	167
2.4.	PCR detection limit	169
2.5.	Typing potential of <i>X. euroxanthea</i> -specific DNA markers.....	169
3.	Results	169
3.1.	<i>In silico</i> selection of DNA markers for <i>X. euroxanthea</i>	169
3.2.	Genomic analysis unearths the stability of XEA DNA markers	171
3.3.	Multiplex PCR allows for the confident identification of <i>X. euroxanthea</i> strains.....	173
3.4.	Detection limit of multiplex PCR with XEA DNA markers.....	174
3.5.	Typing potential of informative XEA DNA markers	174
4.	Discussion.....	175
5.	Conclusions	181

References.....	182
Supplementary Material.....	186
Chapter VI.....	187
General Discussion.....	189
Daring to be different - <i>X. euroxanthea</i> discovery story	189
<i>X. euroxanthea</i> and <i>X. arboricola</i> partner up on multiple host plants - a sympatric powerhouse duo?	190
T3SS and T3E evolution in <i>Xanthomonas</i> - more is not necessarily more	191
Tackling <i>X. euroxanthea</i> and <i>X. arboricola</i> sympatric lifestyle with comparative genomics	192
Providing the tools for <i>X. euroxanthea</i> identification and typing.....	193
Evolution of <i>X. euroxanthea</i>	193
Current situation of <i>X. euroxanthea</i> – What we know so far	194
Final remarks and Future perspectives	197
References.....	198
Supplementary material.....	202
List of Figures.....	Erro! Marcador não definido.
List of Tables	Erro! Marcador não definido.

List of Tables

Table II.1. Strains used for pathogenicity tests and Biolog assays.	75
Table II.2. Biolog profiles of <i>anthomonas euroxanthea</i> strains in comparison with members of <i>Xanthomonas arboricola</i> pv. <i>juglandis</i> and other related <i>Xanthomonas</i> species.	78
Table III.1. General genomic features of the four strains analyzed in this study.	101
Table III.2. Comparison of the occurrence of homologs associated with chemotaxis, non-fimbrial adhesins, T2SS, extracellular enzymes, and T4P, between <i>Xanthomonas euroxanthea</i> strains (CPBF 367, CPBF 424 ^T , and CPBF 426) and the <i>Xanthomonas arboricola</i> pv. <i>juglandis</i> (<i>Xaj</i>) strain (CPBF 427).....	109
Table IV.1. General genomic features.....	130
Supplementary Table IV.S1. List of genomes used in this study	152
Supplementary Table IV.S2. Query proteins used in T3E/T3SS prediction	154
Table V.1. Bacterial strains used for MaGe synteny analysis to retrieve <i>Xanthomonas euroxanthea</i> specific coding sequences (CDSs).	164
Table V.2. List of bacterial strains used for validation of the <i>Xanthomonas euroxanthea</i> -specific DNA markers.	166
Table V.3. Selected <i>Xanthomonas euroxanthea</i> -specific DNA markers (XEA1 to XEA8), corresponding primer pair sequences, expected amplicon sizes, and best BLASTn hits of amplicons with non-X. <i>euroxanthea</i> genomes.....	168
Table VI.I. Database of <i>X. euroxanthea</i> strains described so far	195

List of Figures

Figure I.1 Workflow for genome based classification at the species level. Extracted from (doi.org/10.1099/ijsem.0.002516).....	47
Figure II.1. Walnut plantlets leaves seven days post-inoculation with (a) isolate CPBF 424 ^T showing walnut bacteria blight (WBB) symptoms, (b) isolate CPBF 367 no WBB symptoms were observed, (c) <i>Xanthomonas arboricola</i> pv. <i>juglandis</i> LMG 747 ^T and (d) <i>X. arboricola</i> pv. <i>juglandis</i> strain CPBF 1480, used as positive controls, show characteristic WBB symptoms.	76
Figure II.2 . Maximum likelihood tree based on the nucleotide alignments of 118 concatenated sequences of <i>atpD</i> , <i>dnaK</i> , <i>efp</i> , <i>fyuA</i> , <i>glnA</i> , <i>gyrB</i> and <i>rpoD</i> partial sequences (total length of 4528 bp), emphasizing the phylogenetic relatedness of <i>Xanthomonas euroxanthea</i> strains CPBF 367, CPBF 424 ^T and CPBF 426 (highlighted in bold) within the <i>Xanthomonas</i> genus. Bootstrap values >50 from 1000 iterations are shown over and below the branches.....	80
Figure II.3 Average nucleotide identity (ANI) between novel isolates CPBF 367 and CPBF 424 ^T (in bold) and 20 different <i>Xanthomonas</i> genomes using EDGAR. Pairwise ANI values are shown in percentages and indicated within the cells of the heat-map. Accession numbers of sequences are available in Supplementary Table II.S1.	82
Figure II.4 Distribution of protein homologs of the type III secretion system (T3SS) and type III effectors (T3Es) among the <i>Xanthomonas euroxanthea</i> strains (CPBF 367, CPBF 424 ^T , CPBF 426) and <i>Xanthomonas arboricola</i> pv. <i>juglandis</i> (CPBF 427 and CPBF 1521) strains based on tBLASTn analysis. Dark blue corresponds to the presence of T3SS protein homologous, and dark green corresponds to the presence of T3E homologous. The light blue and green indicate absence of T3SS and T3E homologous, respectively. The threshold criteria applied were minimal e-value of 1e-10, query length similarity cut-off ≥40%, and sequence identity cut-off ≥ 70%.	83
Figure III.1. Heatmap representing average nucleotide identity (ANI) (disclosed in Supplementary Table III.S2) and the respective distance tree cladogram, determined for 44 <i>Xanthomonas</i> genomes including the four genomes disclosed in this study, indicated with (*) and highlighted in orange (CPBF 367, CPBF 424 ^T , CPBF 426) or blue (CPBF 427). Pathogenic and nonpathogenic on walnut are marked with (+) or (−), respectively. The color scale ranging from white to dark red depicts lower to higher similarity values,	

respectively. The strain names refer to the code field from Supplementary Table III.S1
..... 104

Figure III.2. Maximum likelihood tree based on concatenated sequences of the 1149 core genes of 44 *Xanthomonas* genomes, including the *Xanthomonas euroxanthea* strains (branches in orange) and *Xanthomonas arboricola* strains (branches in blue). Phylogenetic relations were inferred using RaxML and the phylogram represented with the R package “ggtree”. Supporting values from 500 bootstrap replicates are indicated near nodes. Dot plot scheme represents the presence/absence scheme for type 3 secretion system (T3SS) and effectors (T3E) putative homologs; ●, present; ○, not present; considering a tBLASTn hit with a query coverage threshold $\geq 75\%$, and sequence identity with $\geq 40\%$ cut-off. Results for genomes disclosed in this study are marked with (*); also, *X. euroxanthea* strains CPBF 367, CPBF 424T, and CPBF 426 are highlighted in orange and for *X. arboricola* pv. *juglandis* CPBF 427 in blue. Pathogenic and nonpathogenic on walnut are marked with (+) or (–), respectively. The strain names refer to the code field from Supplementary Table III.S1. Best BLAST results and accession numbers of sequences used as query are disclosed in Supplementary Table III.S3 106

Figure III.3. Venn diagram highlighting the genomic patrimony from CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427 genomes. The core genome is given by the interception of the four strains and corresponds to the number of orthologous CDSs shared (3398 CDSs). Strain specific CDSs are also represented in the periphery of the diagram (208 for CPBF 367, 213 for CPBF 426, 186 for CPBF 424^T, and 621 for CPBF 427). The remaining combinations represent the number of orthologous shared between two to three genomes. The strain names refer to the code field from Supplementary Table III.S1. 107

Figure IV.1. Heatmap representing average nucleotide identity (ANI) (disclosed in Supplementary Table IV.S1) and the respective distance tree cladogram determined for 51 *Xanthomonas* genomes. *Carya illinoensis* and *Solanum lycopersicum* isolates with *X. euroxanthea* strains are highlighted in blue, while isolates with ANI values $>95\%$ for *X. arboricola* are highlighted in orange. The colour scale ranging from white to dark red depicts lower to higher similarity values, respectively..... 132

Figure IV.2. Phylogenomic maximum likelihood tree using RaxML from concatenated sequences of 1149 core genes from 51 genomes including the *Xanthomonas euroxanthea* strains (branches in blue) and *Xanthomonas arboricola* and *Xanthomonas*

sp. strains (branches in orange). The phylogram was constructed with the R package “ggtree” with supporting values from 100 bootstrap replicates indicated near nodes. 134

Figure IV.3. (A) Pathogenicity tests results four weeks after inoculation. No symptoms were observed on pecan leaflets for CPBF 761 and CPBF 766 (on the left), while small necrotic spots with chlorotic halos were produced by CPBF 761 and CPBF 766 on walnut (on the right); (B) Hypersensitivity response symptoms on leaves of *Nicotiana benthamiana* after 72 h after inoculation with the *X. euroxanthea* strains CPBF 761, CPBF 766 and *X. arboricola* CPBF 765 and CPBF 1494 (from left to right). 136

Figure IV.4 (A) Circular plot for genome comparison of *X. euroxanthea* (Xea) and *X. arboricola* (Xa) strains from two different hosts, walnut (Xea CPBF 367, Xaj CPBF 427) and pecan (Xea CPBF 761, CPBF 766, Xa CPBF 765, CPBF 1494), using *X. euroxanthea* CPBF 424^T as the reference genome. The analysis and plot were carried out with EDGAR 3.0. (B) Genomic island prediction of *X. euroxanthea* CPBF 424^T using IslandViewer4, which integrates data from three tools: IslandPath-DIMOB, SIGI-HMM and IslandPick. The T3SS of CPBF 424^T is highlighted in light grey (560320-592275bp) 138

Figure IV.5. Dot plot scheme represents the presence/absence scheme for type 3 secretion system (T3SS) and effectors (T3E) putative homologs; ■ present; □ not present; considering a tBLASTn hit with a query coverage threshold ≥ 75%, and sequence identity with ≥ 40% % cut-off. Pathogenicity tests results in walnut (W), pecan (P) and tomato (T) are displayed on the right (this study, Martins et al. 2020, Roach et al. 2020). ¹Coverage and identity values are borderline and within the defined threshold, respectively; ²genes automatically annotated in CPBF 424^T genome that were confirmed by blast using the threshold of ≥ 40% identity and ≥ 75% query sequence length..... 139

Figure IV.6 Synteny analysis of genes included in type 3 secretion system (T3SS) cluster of *X. euroxanthea* strains CPBF 424^T and CPBF 766 and *X. arboricola* CPBF 765 and CPBF 427 using Easyfig. T3SS and T3E genes and depicted as green and orange arrows, respectively. ^{1,2} not characterized as effectors, multiple sequence alignments and blast results of protein sequences corresponding to each strain are disclosed in Supplementary Figure IV.S1; ¹accession numbers XTG000468 (CPBF 424T), XCY000470 (CPBF 766); ²accession numbers XCY000604 (CPBF 765), XSP003898 (CPBF 427); amino acid sequence alignment between ¹ and ² is shown in Supplementary Figure IV.S1; ³accession numbers XTG000469 (CPBF 424T), XCY000471 (CPBF 766), XCY000605 (CPBF 765), XSP003896 (CPBF 427); 4XSP003897 is only present in CPBF 427. 141

Supplementary Figure IV.S1. Multiple sequence alignments and blast results	157
Figure V.1. Flowchart for selection and validation of <i>X. euroxanthea</i> -specific DNA markers.....	165
Figure V.2. Distribution of six <i>Xanthomonas euroxanthea</i> (<i>Xea</i>)-specific DNA markers (XEA1, XEA2, XEA3, XEA5, XEA6 and XEA8) in 11 <i>X. euroxanthea</i> genomes. The presence/absence of six XEA DNA markers was assessed by BLASTn analysis in Geneious, allowing to disclose three patterns, A to C, that do not translate strain-host affinities.	170
Figure V.3. Circular map of <i>Xanthomonas euroxanthea</i> strain CPBF 424 ^T chromosome. Outside to inner circles are showing genome coordinates (bp); <i>X. euroxanthea</i> -specific DNA markers XEA1 to XEA8 (red); housekeeping genes <i>gyrB</i> , <i>rpoD</i> , <i>fyuA</i> and <i>acnB</i> (yellow); transposases (green); recombinases (light blue); integrases (dark blue) and phage-related ORFs (purple). For each XEA DNA marker and housekeeping gene the number of SNPs (calculated based on 11 genomes of <i>X. euroxanthea</i>), GC content and CAI/eCAI values are shown.	171
Figure V.4. Comparative syntenic maps of four <i>Xanthomonas euroxanthea</i> -specific DNA marker-harboring-regions (a) XEA1 (designed from a conserved protein of unknown function sequence), (b) XEA5 (design from a MarR family transcriptional regulator), (c) XEA6 (designed from a conserved protein of unknown function sequence) and (d) XEA8 (designed from a protein of unknown function, and a TetR/AcrR family transcriptional regulator sequences) DNA markers across 11 <i>X. euroxanthea</i> (<i>Xea</i>) and two <i>Xanthomonas arboricola</i> (<i>Xa</i>) strains.	173
Figure V.5. Multiplex PCR using <i>Xanthomonas euroxanthea</i> -specific DNA markers XEA1 (819 bp), XEA8 (648 bp) and XEA5 (295 bp) on 7 <i>X. euroxanthea</i> strains, 10 <i>Xanthomonas arboricola</i> strains, 6 pathovars of <i>Xanthomonas arboricola</i> , and 9 non- <i>arboricola</i> <i>Xanthomonas</i> species. Markers XEA5 and XEA8 were successful in detecting <i>X. euroxanthea</i> strains, while XEA1 identified all <i>X. euroxanthea</i> strains, except for CFBP 7622. No amplification was observed for any of the other xanthomonads tested, namely <i>X. arboricola</i> and other <i>Xanthomonas</i> species. C-: negative control.	174
Figure V.6. Maximum-likelihood phylogenetic tree based on concatenated sequences of (a) DNA markers XEA5, XEA6 and XEA8 (1180 bp); and (b) partial housekeeping gene sequences for <i>acnB</i> , <i>fyuA</i> , <i>gyrB</i> and <i>rpoD</i> (2774 bp) extracted from 11 <i>X. euroxanthea</i> genomes. The tree was constructed using the Tamura-Nei model using MEGA X (Kumar	

et al., 2018). Supporting values from 1000 bootstrap replicates are indicated near nodes.
..... 175

Figure VI.I. Working model depicting *X. euroxanthea* lifestyle..... 196

Supplementary Figure VI.S1. Distribution of six *Xanthomonas euroxanthea* (*Xea*)-
specific DNA markers (XEA1 - XEA8) in hitherto known and available *X. euroxanthea*
genomes assessed by BLASTn analysis, disclosing patterns A-D 202

List of Abbreviations

ANI	Average Nucleotide Identity
BAN	Brown Apical Necrosis
BLAST	Blast Local Alignment Search Tool
Bp	Base Pairs
BRIP	Plant Pathology Herbarium (Australia)
CAI	Codon Adaptation Index
CDS	Coding DNA Sequence
CFBP	Collection Française De Bactéries Phytopathogènes (French Collection of Phytopathogenic Bacteria)
CCOS	Culture Collection of Switzerland
CPBF	Coleção Portuguesa De Bactérias Fitopatogénicas (Portuguese Collection of Phytopathogenic Bacteria)
dDDH	Digital DNA-DNA Hybridization
EDGAR	Efficient Database Framework for Comparative Genome analysis using BLAST Score Ratios
ENA	European Nucleotide Archive
EPPO	European And Mediterranean Plant Protection Organization
EPS	Extracellular Polysaccharides
ETI	Effector-Triggered Immunity
HGT	Horizontal Gene Transfer
HR	Hypersensitive Response
Hpa	Hypersensitive Reaction and Pathogenicity associated gene
Hrc	Hrp conserved gene
Hrp	Hypersensitive Reaction and Pathogenicity gene
LB	Luria-Bertani Culture Medium
LMG	LMG Culture Collection, Gent University, Belgium
LOD	Limit of Detection
LPS	Lipopolysaccharides
MLSA	Multilocus Sequence Analysis
NA	Nutrient Agar Culture Medium
NCBI	National Center for Biotechnology Information
NCPPB	National Collection of Plant Pathogenic Bacteria, UK
ONT	Oxford Nanopore Technologies
ORF	Open reading frame

PSA	Peptone Sucrose Agar Culture Medium
SDW	Sterile Distilled Water
SNP	Single Nucleotide Polymorphism
T3E	Type III Effector
T4P	Type IV Pilus
T2SS	Type II Secretion System
T3SS	Type III Secretion System
T4SS	Type IV Secretion System
T6SS	Type VI Secretion System
VOC	Vertical Oozing Canker
WBB	Walnut Bacterial Blight
XA	<i>Xanthomonas arboricola</i>
XAJ	<i>Xanthomonas arboricola</i> pv. <i>juglandis</i>
XEA	<i>Xanthomonas euroxanthea</i>
YDC	Yeast Extract Dextrose Carbonate Culture Media



Chapter I

General introduction

General Introduction

1. *Xanthomonas*, a diverse group of phytopathogenic bacteria

The *Xanthomonas* genus is composed of gram-negative plant-associated bacteria (Hayward, 1993; Vauterin *et al.*, 1995), owing its name to its characteristic yellow pigmentation (from the Greek *Xanthos* and *monas*, translates “yellow entity” (An *et al.*, 2019), with most strains forming yellow, shiny, slimy colonies on nutrient agar and potato media (Dowson, 1939; Van den Mooter & Swings, 1990). At this time the genus includes 34 species with validly published and correct names (<https://lpsn.dsmz.de/genus/xanthomonas> accessed in 3 November 2022) (Parte *et al.*, 2020). This genus is composed mostly by pathogenic species which can cause disease in an eclectic range of about 400 monocot or dicot hosts (Hayward, 1993), with most of species or pathovars (i.e. intraspecies subgrouping according to host or symptomatology) being specialized in a single or a reduced range of hosts. Furthermore, this host range may be underestimated due to a reduced surveillance of plant hosts with less or without economic importance and for that are unaccounted for (Vauterin *et al.*, 1991; Hayward, 1993). Altogether, the ubiquity and co-evolutionary potential with several hosts stirs up a considerable genetic diversity, making this genus especially taxonomically rich.

1.1. The winding road of *Xanthomonas* taxonomical reclassification

Bacteria classification reports back to the late 19th century, and at the time taxonomical groups would be distinguished according to growth requirements, *in vitro*, morphologic and pathogenicity potential. These have shifted to more physiology, biochemical-based criteria that accompanied the refinement of comparative tools to finally transition from phenotypic to genotypic approaches, which deliver phylogenetic relationships between individuals (Schleifer, 2009).

The *Xanthomonas* genus was firstly coined by (Dowson, 1939) and since then has undergone considerable rearrangements. Initially, classification at the species and pathovar levels were based on biochemical characteristics of strains (Van den Mooter & Swings, 1990) being followed by a profound revision of the genus according to DNA-DNA hybridization (Vauterin *et al.*, 1995). Classification based on Multilocus-sequence analysis (MLSA) using the housekeeping genes (i.e. essential and ubiquitous bacterial genes) chaperone protein (*dnaK*), tonB-dependent receptor (*fyuA*), DNA gyrase subunit

B (*gyrB*) and polymerase sigma factor (*rpoD*) was then later proposed for a more rapid and less laborious strain differentiation/identification (Young *et al.*, 2008). As an effort to easily identify and reveal the phylogenetic structure within the genus, (Parkinson *et al.*, 2009) performed an extensive *gyrB*-based phylogeny considering all available species, pathovars and strains not attributed to any species (*Xanthomonas* sp.). This analysis allowed a glimpse on between-species diversity and even pointed out groups of potential unidentified species, referred to as species-level clades (Slc 1-Slc 7) (Parkinson *et al.*, 2009). With the increase of whole-genome sequences of strains from economically unattractive plant species, and the availability of bioinformatic identification tools, some of these species like-clades are now confirmed as novel species, namely, *Xanthomonas cannabis* (Slc 1), *Xanthomonas badrii* (Slc 4), *Xanthomonas indica* (Slc 7) and yet-to-be-named members of Slc5 and Slc 6 groups (Peduzzi *et al.*, 2022). The exponential surge of complete i.e. end-to-end genomic sequences together with computational tools that allow more holistic comparative analyses triggered a step-up in more refined prokaryote taxonomy and systematics (Chun & Rainey, 2014; Venter *et al.*, 2022), quickly becoming established as the go-to basis for a refined comparative/structural genomics study, being commonplace for many labs to sequence and assemble their microbial genomes in-house. This also opened the door for phylogenomics i.e. core genome-base phylogeny, which is a step-up from the standard 16S one marker-based phylogenetic inferences that may infer strong closeness between strains that disguise major genomic rearrangements or Multilocus Sequence Analysis (MLSA) phylogenies that are constricted to a reduced number of housekeeping genes. Phylogenomics rely instead on single-copy genes, running into hundreds of genes ubiquitously present in all genomes considered in the analysis (Ankenbrand & Keller, 2016; Chun *et al.*, 2018; Na *et al.*, 2018). Thus, allowing a more robust assessment of phylogenetic relatedness and taxonomy inferences at the genus level (Chun *et al.*, 2018).

As will be further detailed in the next section, the advances of whole genome sequencing-based taxonomy tools, particularly Average Nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) facilitated the long-overdue reclassification within *Xanthomonas arboricola* species (Fernandes *et al.*, 2021; Kałużna *et al.*, 2021; Zarei *et al.*, 2022). Furthermore, these tools also contributed for a few rearrangements of species within the genus, mostly associated with reclassifying pathovars within other already known species (da Gama *et al.*, 2018; Studholme *et al.*, 2020; Morinière *et al.*, 2020; Bansal *et al.*, 2022). Although sometimes both tools can become a source of dissent when ANI and dDDH values fall in the “transition zone” for species delineation (93-96% and 60-70%, respectively), hence reinforcing the importance of complementary assays

for assessing phenotypical traits and phylogenetic evaluation in order to avoid meaningless species separation (Rosselló-Móra & Amann, 2015; Timilsina *et al.*, 2019; Morinière *et al.*, 2020).

1.2. Describing a new bacterial species

When describing a new bacterial species, the proposed name must abide by the rules of the ICNP - International Code of Nomenclature of Prokaryotes (Parker *et al.*, 2019) in order to be recognized as validly published taxa. Essentially, the proposed taxon must be approved and published for publication on the *International Journal of Systematics and Evolutionary Microbiology* (IJSEM), also providing the etymology of the approved latin name, followed by a detailed description of the species morphological and phenotypic traits (Rule 27 of ICNP)(Parker *et al.*, 2019).

The current IJSEM requirements for proposing prokaryote species as a new taxon, include providing confirmation of genomic distance to described closely related species through Overall Genome Related Index (OGRI) calculations such as ANI and dDDH, phylogenetic analysis (preferentially using core genome data), general genomic information, and morphological physiological and chemotaxonomical characterization is required (Chun & Rainey, 2014; Chun *et al.*, 2018, <https://www.microbiologyresearch.org/content/journal/ijsem?page=about-journal>).

1.3. The particular case of *Xanthomonas arboricola*

Among the *Xanthomonas* genus, *Xanthomonas arboricola* is a particular heterogeneous group (Zarei *et al.*, 2022b), infecting numerous different plant species. According to the host of infection, *X. arboricola* has been known to infect stone fruit trees (peach, apricot, and plum), nut trees (almond, hazelnut, and walnut), fruit trees (banana), ornamental (niger, arum lily, and poplar) and herbaceous plants (arracacha and strawberry) (Lamichhane, 2014; Kałużna *et al.*, 2021) and as so, until recently, nine pathovars were acknowledged: *pruni*, *corylina*, *juglans*, *celebensis*, *guizotiae*, *zantedeschiae*, *populi*, *arracaciae* and *fragariae* (Fischer-Le Saux *et al.*, 2015). This number of *X. arboricola* pathovars may soon be reduced to seven, as a thorough genomic analysis of *X. arboricola* pathovars disclosed low DNA similarities between *X. arboricola* type strain and pathotype strains of *populi* and *guizotiae*, proposing these pathovars to be considered as species (Zarei *et al.*, 2022b). Particularly, *X. arboricola* species has also been associated with other diseases such as: bacterial leaf spot disease in bell pepper (*Capsicum annuum*) in Korea (Myung *et al.*, 2010) and tomato (*Solanum lycopersicum*)

in Tanzania (Mbega *et al.*, 2012) and more recently, bacterial blight in pecan (*Carya illinoensis*) in Portugal (Fernandes *et al.*, 2018) and bacterial leaf blight in blueberry (*Vaccinium corymbosum*) (Kałużna & Pothier, 2022).

The *X. arboricola* species also harbours an intriguing number of non-pathogenic strains (Merda *et al.*, 2016), which have been isolated from both symptomatic and non-symptomatic tissues of *Juglans regia* buds (Essakhi *et al.*, 2015) and leaves and buds of *Prunus persica* (peach) (Garita-Cambronero *et al.*, 2016a,b). Furthermore, a recent survey of bacterial canker and spot on stone fruits, and walnut bacterial blight in Iran revealed that more than half of the walnut isolates were non-pathogenic in walnut and all the stone fruit isolates were non-pathogenic except one that was weakly pathogenic in peach and nectarine (Zarei *et al.*, 2022a). Regarding phylogenetic structure of *X. arboricola*, while pathogenic strains are grouped in well-defined monophyletic clusters according to their host, which means organization according to pathovars, typically these non-pathogenic strains of *X. arboricola* do not cluster according to the host of isolation and instead usually group together (Merda *et al.*, 2016; Bansal *et al.*, 2020; Kałużna *et al.*, 2021; Zarei *et al.*, 2022a). Merda *et al.* 2016 proposed an epidemic population structure for *X. arboricola*, considering pathovars *pruni*, *juglandis* and *corylina* (the successful pathovars i.e. highly pathogenic and epidemically widespread around the globe) as epidemic clones, having mutation as the number one evolutionary strength. Meanwhile, less virulent pathovars and non-pathogenic strains from different hosts are under high recombination pressure, being included in the recombination network (Merda *et al.*, 2016).

1.4. Non-pathogenic bacteria conquer their place - *Xanthomonas* pathogenic and commensal dynamics

Interestingly, several new species of the *Xanthomonas* genus composed exclusively or partially by non-pathogenic strains (at least in the host of isolation) have been recently proposed, namely: *X. euroxanthea* from healthy buds of *Juglans regia* (Martins *et al.*, 2020), *X. indica* and *X. sontii* isolated from healthy seeds of *Oryza sativa* (Bansal *et al.*, 2021; Rana *et al.*, 2022), *X. bonasiae* and *X. yoonii* from crown gall tissues of *Ficus benjamina* and *Amaranthus* sp., respectively (Mafakheri *et al.*, 2022), and *X. maliensis* from rice plant leaves (Triplett *et al.*, 2015). The emergence of these novel species, so far constituted only by non-pathogenic strains, leads to wonder about their origin and the evolutionary pressures that these non-pathogenic strains exert in their hosts and how these contrast with pathogenic species in their respective hosts.

(Merda *et al.*, 2017) attempted to understand the evolution of pathogenicity in *Xanthomonas* by focusing on the evolution of the type III secretion system cluster, a key piece for pathogenicity in *Xanthomonas* (as it will be further detailed in the following sections).

It would be interesting to ascertain the origin of these nonpathogenic strains when they are genetically similar to pathogenic ones. The population structure of pathogenic and nonpathogenic xanthomonads in a host may alter drastically and swiftly due to frequent interactions leading to horizontal gene transfer and gain and/or loss of virulence genes. Speciation event of existing broad range bacteria that evolved in the way of becoming dependent of pathogenic strains.

This prompts the question as what the ecological role of these non-pathogenic strains may be so that evolutionary pressures keep on selecting them and one hypothesises that these strains may be included in a consortia with other pathogenic species that have the arsenal for plant infection and shape the host niche for successful colonization.

2. Recent advances in genomics tools prompted a BOOM on *Xanthomonas* genomics analysis

The boom of whole genome sequencing prompted sequencing of strains from the same genus and same species, including different pathovars, allowing comparative genomics analysis and considerable advances regarding pathogenicity and virulence evolution, host specificity, and clarification on phylogenetic relations (An *et al.*, 2019; Timilsina *et al.*, 2022).

2.1. Improvement of sequencing technologies

Long-read and short read Next generation sequencing (NGS) methods perform differently, particularly regarding read length, accuracy, nature of sequencing errors, cost-effectiveness and, ability to produce paired-end information (Chun & Rainey, 2014). By performing hybrid genome assemblies, one gets “the best of both worlds”, obtaining a reduced number of larger contigs from long reads (less fragmented sequences and more gap closure) with the high quality from short reads (Wick *et al.*, 2017). Within the *Xanthomonas* genus a lot of hybrid genomes have increasingly been published, including strains belonging to *Xanthomonas euroxanthea* (Teixeira *et al.*, 2020), *Xanthomonas arboricola* (Kałużna & Pothier, 2022), pv. *juglandis* (Teixeira *et al.*, 2021; Cuesta-Morrondo *et al.*, 2022), pv. *corylina* (Nunez *et al.*, 2021; Pothier *et al.*, 2022; Cuesta-

Morrondo *et al.*, 2022), *pv. pruni* (Cuesta-Morrondo *et al.*, 2022), *Xanthomonas hortorum* (Dia *et al.*, 2020), *Xanthomonas euvesicatoria pv. alfalfa*, *X. citri pv. glycines* (Ruh *et al.*, 2021), *pv. fuscans* and *X. phaseoli pv. phaseoli* (Briand *et al.*, 2021). High-quality, closed bacterial genomes, as those assured by hybrid assemblies of long and short read technologies (Cuesta-Morrondo *et al.*, 2022) are the best suitable for whole genome comparisons due to less gaps and high fidelity (Wick *et al.*, 2017; Amarasinghe *et al.*, 2020)

2.2. Genomics-facilitated taxonomic re-classification

For some time, the go-to method for prokaryote taxonomical classification was DNA-DNA hybridization (DDH)-based (Wayne *et al.*, 1988), which relied on the evaluation of the levels of hybrid association or thermal stability, and applied $\geq 70\%$ threshold for species differentiation (Rosselló-Móra *et al.*, 2011). As a more reliable, reproducible and less laborious tool, bioinformatic-based alternatives were demanded (Auch *et al.*, 2010; Chun *et al.*, 2018). Progressive attainable cost and accessibility of genome sequencing opened the door for whole genome comparisons-based tools, also known as, Overall genome related index (OGRI) tools (Chun & Rainey, 2014). OGRI include similarity-type index tools, best represented by Average nucleotide identity (ANI) (Konstantinidis & Tiedje, 2005); and distance-type genome relatedness indexes such as genome Blast distance phylogeny (GBDP) and MUM index (MUMi) (Deloger *et al.*, 2009; Chun & Rainey, 2014). The latter is a MUMmer-based tool that calculates maximal unique and exact matches with potential to be more reliable in within species differentiation. However, it does not seem to be widely used, probably because no threshold for species demarcation seems to have been attributed conversely to ANI and GBDP (Deloger *et al.*, 2009; Chun & Rainey, 2014). Furthermore, good correlation of a proposed threshold for ANI species delimitation with previous gold-standard DDH (Goris *et al.*, 2007) and 16 rRNA multiple sequence alignments (Konstantinidis & Tiedje, 2005) supported the use of these tools as the new gold-standard for species differentiation (Hu *et al.*, 2022), following the generally accepted threshold of 95–96 % same species individuals (Richter & Rosselló-Móra, 2009). Several reliable and user-friendly tools such as Genome-to-Genome Distance Calculator (GGDC 2.1; <http://ggdc.dsmz.de/ggdc.php>) (Meier-Kolthoff *et al.*, 2013, 2022), also referred to as dDDH, digital DNA-DNA hybridization (GBDP); and JspeciesWS (Richter *et al.*, 2016), EZGenome (<http://www.ezbiocloud.net/ezgenome/ani>), ANI calculator (Rodriguez-R & Konstantinidis, 2016), the EDGAR genomic comparisons platform (Blom *et al.*, 2009,

2016; Dieckmann *et al.*, 2021) (ANI) and LINbase (Tian *et al.*, 2020) are currently freely accessible to researchers.

The availability of these tools facilitated the long-overdue reclassification within *Xanthomonas* genus and particularly within *Xanthomonas arboricola* (Fernandes *et al.*, 2021b; Kalužna *et al.*, 2021; Zarei *et al.*, 2022b) (this aspect will be further addressed on the next section).

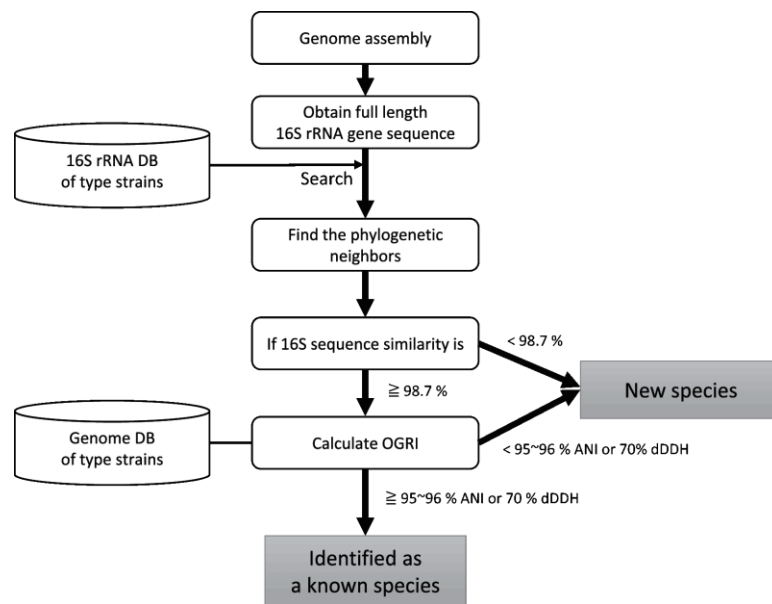


Figure I.1 Workflow for genome based classification at the species level. Extracted from (doi.org/10.1099/ijsem.0.002516).

2.3. Using comparative genomics to unlock lifestyle traits and DNA marker identification

Currently, there is a lot of genome sequences flooding the databases, it is important to explore this pool of data, instead of sequencing more isolates randomly (Timilsina *et al.*, 2022). The massive availability of *Xanthomonas* genomes from different hosts, from more or less aggressive strains, allowed the integration of these data in comprehensive comparative genomic studies of bacteria infecting or colonizing the same host plant (Fernandes *et al.*, 2021b) or host group (Garita-Cambronero *et al.*, 2017), causing similar disease symptoms (Garita-Cambronero *et al.*, 2019) or different pathogenicity or virulence levels (Cesbron *et al.*, 2015; Garita-Cambronero *et al.*, 2016b,c). Unravelling genetic determinants for host-specificity, for a particular disease and for general virulence. Particularly for *Xanthomonas arboricola*, good quality genomes have been made available in the last years, including non-pathogenic and pathogenic strains of *Prunus* (Garita-Cambronero *et al.*, 2016b,a).

For some time, identification of sampled isolates relied on phenotypic characteristics, biochemical tests as described by Schaad (Schaad *et al.*, 2001) or such MLSAs within species of the *Xanthomonas* genus. The above-mentioned advances in genome sequencing technologies and the reduction of genome sequencing costs fuelled the availability of bacterial genomes. These advances also allowed the use of comparative genomic tools for identification and selection of DNA markers/signatures (specific genomic markers), i.e. nucleotide sequence fragments that are specific to the target taxa (Albuquerque *et al.*, 2009). This strategy has been used to identify markers for phytopathogenic bacteria, achieving specificity at the species (Albuquerque *et al.*, 2012; Almeida *et al.*, 2018; Alves *et al.*, 2020; Silva *et al.*, 2022) and even pathovar level (Albuquerque *et al.*, 2011; Fernandes *et al.*, 2017). Upon identification, these DNA signatures can then be applied in designing highly reliable identification and detection methods by multiplex-PCR (Fernandes *et al.*, 2017; Silva *et al.*, 2022), DNA dot-blot (Alves *et al.*, 2020) and qPCR (Almeida *et al.*, 2018; Martins *et al.*, 2019).

3. Mechanisms of virulence and pathogenicity in *Xanthomonas* spp.

Plant infection is a challenging journey for bacterial pathogens because they must overcome a series of obstacles and difficult trials to be successful (Pfeilmeier *et al.*, 2016). The *Xanthomonas* lifecycle in plants starts as an epiphytic interaction and to remain attached to host, bacteria depend on fimbrial adhesins such as type IV pili for cell aggregation and non-fimbrial adhesins such as *Xanthomonas*-adhesion like proteins A, B (XadA, XadB) and the Yersinia auto- transporter-like protein H (YapH) for surface attachment (Das *et al.*, 2009; Büttner & Bonas, 2010). At this point, bacteria must endure the harsh environments of the leaf surface prone to UV radiation, dissection, nutrient shortage, competition, and predation (Alvarez-Martinez *et al.*, 2021), making use of xanthan (this genus' characteristic polysaccharide) and lipopolysaccharides (LPS) (Büttner & Bonas, 2010) to form stable biofilms (Sena-Vélez *et al.*, 2015). Bacteria can then make their way into the host plant through wounds or natural openings such as stomata or hydathodes (Stefani *et al.*, 2021). Once in the intercellular space, pathogens may mobilize DNA and effector proteins to the inside of the plant cell through a series of distinct secretion systems (Büttner & Bonas, 2010) (Figure 1.2). These effectors are key virulence factors that promote pathogen growth and establishment within the host by overthrowing the plants' immune responses (Mattoo *et al.*, 2007; Shames & Finlay, 2012).

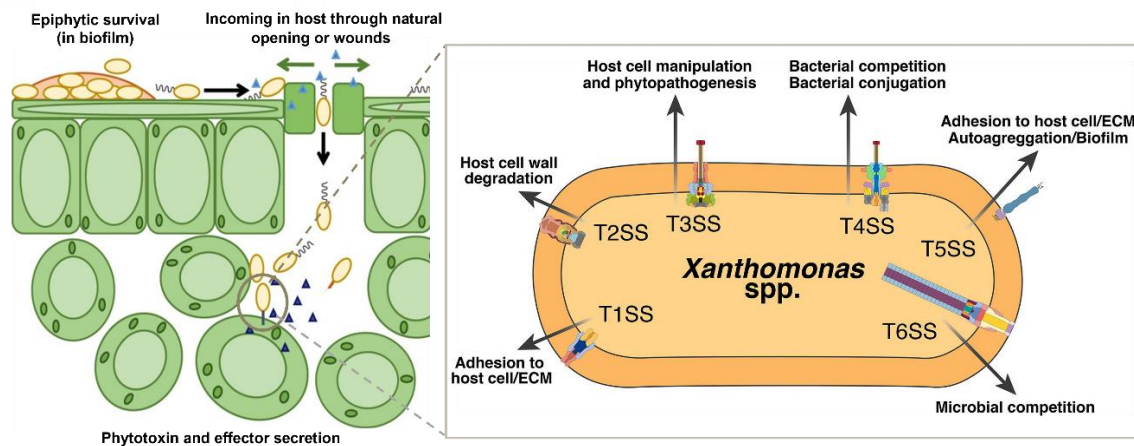


Figure I.2 Route for host infection by *Xanthomonas* pathogens and associated infection secretion systems. Adapted from (Pfeilmeier et al., 2016 (doi:10.1111/mpp.12427); Alvarez-Martinez et al., 2021 (doi:10.1016/j.csbj.2020.12.020)) <https://creativecommons.org/licenses/by-nc-nd/4.0/>

To protect against pathogens, the plant immune system acts as a fine-tuned and intricate complex network that perceives pathogenic activity and triggers signalling cascades to ultimately lead to defence responses (Delplace *et al.*, 2022). Plants have two layers of defence, which are co-regulated (Ngou *et al.*, 2021) and co-dependent: PAMP-triggered immunity (PTI) that perceives pathogenic stimuli by pattern recognition receptors (PRRs) at the cell surface; and Effector-triggered immunity (ETI) that recognizes pathogen effector proteins through nucleotide-binding, leucine-rich repeat-containing receptors (NLRs) at the intracellular level (Jones & Dangl, 2006; Yuan *et al.*, 2021). As these two immunity pathways share some common signalling components, their defence responses are similar, however ETI is stronger and usually culminates in a localized programmed cell death form called hypersensitive response (HR) to halt pathogenic growth (Chang *et al.*, 2014). A lot of what is known on which genes are responsible for ETI, the mode of action and mechanisms involved in this response owes to functional analysis of effectors from the type III secretion system (T3Es) from Gamma proteobacteria including *Xanthomonas* (Hurley *et al.*, 2014; Adlung *et al.*, 2016; Wei *et al.*, 2018; Martin *et al.*, 2020). The injectosome type III secretion system (T3SS) has an important role in suppressing PTI and ETI responses by directly injecting effector proteins (examples of these are the *Xanthomonas* outer proteins (Xop) avirulence proteins (Avrs)) that destabilize several cellular functions to increase bacterial fitness (White *et al.*, 2009; Cunnac *et al.*, 2011; Alvarez-Martinez *et al.*, 2021). However, the plant immune system can counteract this by directly or indirectly detecting these effectors intracellularly with R proteins (NLRs) and trigger ETI. As a result of intimate host-pathogen co-evolution, effector genes are lost or disrupted, new effector genes may be horizontally acquired or emerge through mutation to surpass recognition by the host,

favouring host expansion and virulence, diversifying the genetic repertoire involved in virulence among bacterial populations (Cunnac *et al.*, 2011).

3.1. Secretion systems are essential for *Xanthomonas* survival

The type III machinery plays an important and well-documented role in host immunity suppression and support infection, but lately, other secretion systems have also been gaining recognition not only for a direct role in disease virulence but also for increasing pathogen fitness and adaptation (Figure I.2) (Timilsina *et al.*, 2020; Alvarez-Martinez *et al.*, 2021). Besides playing a part in secretion of virulence-related factors, the type II secretion system (T2SS) also have a function in general metabolic processes and bacterial survival in the environment, and as such, it can be present in both pathogenic and nonpathogenic strains (Cianciotto, 2005; Rondelet & Condemine, 2013). In *Xanthomonas*, two sets of T2SS clusters are known: the *xps*, overall conserved in the genus and important for bacterial fitness and disease, and *xcs*, present in a restricted number of species, including *X. arboricola*. While *xps* contributes to virulence and host entry particularly through degradation of plant cell wall by secreting various hydrolases including extracellular proteases, xylanases, cellulases, esterases, pectinases, lipases, polygalacturonases and cellobiosidases and also toxins; the *xcs* system does not appear to have a significant role in virulence (Büttner & Bonas, 2010; Szczesny *et al.*, 2010; Rondelet & Condemine, 2013; Timilsina *et al.*, 2020; Teper *et al.*, 2021). The remaining secretion systems, Type I, IV, V and VI (T1SS-T6SS), all seem to have a role in bacterial competition while relying on different substrates, such as bacteriocins, RTX (repeat in toxins motif) proteins and adhesins (T1SS) and enzymatic cocktails and toxins (T4SS). Additionally, these systems also seem to take part in other functions such as adhesion and infiltration to plant cell and, extracellular matrix (T1SS) bacterial conjugation and environmental survival (T4SS), host attachment and epiphytic survival, biofilm formation and in cell-to-cell adhesion (T5SS). In *Xanthomonas*, the T6SS has been reported to have a role in bacterial competition and permanence in the environment (Alvarez-Martinez *et al.*, 2021; Teper *et al.*, 2021), however this system does not appear to have a significant role in all members of the genus, as three clusters (T6SS-I, II, III) are partially or totally absent in some species, including *X. arboricola* (Timilsina *et al.*, 2020; Teper *et al.*, 2021).

3.2. A closer look into Type III secretion system cluster and effectors

The type III secretion systems encompass two kinds of machinery, the flagellar and injectosome (non-flagellar) type III secretion systems. The flagellar is usually overlooked in the literature as it is more involved in the epiphytic phase from the life cycle of *Xanthomonas*, being necessary for motility, chemotaxis and biofilm establishment. While the injectosome (non-flagellar) type III secretion system usually steals the spotlight for its fundamental role in triggering disease as it secretes effector proteins (T3Es) and toxins that will disrupt host signalling pathways. (Abby & Rocha, 2012; Teper *et al.*, 2021; Wu *et al.*, 2021). As the injectosome type III secretion system is the one relevant for this thesis, from this point on it will be the only one mentioned and for simplification it will be referred to as T3SS. The T3SS includes three groups of proteins: structural proteins i.e. proteins that make up the structure of the injectosome; effector proteins (T3Es) i.e. proteins secreted by the T3SS structure; and chaperones i.e. proteins that lead and dock T3Es to the needle complex of T3SS structure while promoting their stability and avoiding degradation and aggregation within the bacteria cell (Büttner & He, 2009; Nazir *et al.*, 2017). In general, *Xanthomonas* species display the Hrp2 gene cluster organization family of T3SS, like other plant pathogens *Ralstonia solanacearum* and *Burkholderia* spp.), encoding around 20 structural genes that encompass hypersensitive reaction and pathogenicity (*hrp*), *hrp*-conserved (*hrc*) and *hrp*-associated (*hpa*) genes (Chang *et al.*, 2014; Nazir *et al.*, 2017; An *et al.*, 2019; Teper *et al.*, 2021). The T3SS structure is analogous to the shape of a chess game's pawn. At the base, the cargo sorting platform is formed by HrcL, HrpB7, HrcN (probably involved in cargo recognition, energy supply for translocation, and cargo unfolding), and HrcQ, with HrcN and HrcQ forming the ATPase complex (Chang *et al.*, 2014; Deng *et al.*, 2017; Alvarez-Martinez *et al.*, 2021). Pawn's body is formed by two ring structures that go through and connect the inner to the outer bacterial membranes, passing through the periplasmic space, encompassing: the export apparatus (HrcU, HrcV, HrcR, HrcS, HrcT), a homopolymeric ring of pore-forming secretins that include HrcC, the needle complex (HrcD, HrcJ, HrpB2, HrcC, HrpE), and the translocon protein (HrpF) (Chang *et al.*, 2014; Alvarez-Martinez *et al.*, 2021).

Two components of the Hrp2 cluster that are completely indispensable for pathogen ability for host colonization, promote HR but also cause disease are the transcriptional regulators HrpX and HrpG that work as master regulators of T3SS and T3Es (Teper *et al.*, 2021). Particularly, HrpG regulates the transcription of HrpX which regulates the expression of the Hrp2 T3SS cluster and of T3Es. This is achieved when HrpX binds to

the PIP (plant-inducible promoter)-box (TTCGB-N₁₅/N₈-TTCGB) of T3SS operons and T3Es (Teper *et al.*, 2021). Despite of the regulatory and structural importance of the T3SS elements mentioned above, there is another important part of the puzzle, the effectors (T3Es) that are injected into the cytosol of host cells by the T3SS machinery. So far, 74 T3E families have been described, accounting for 71 Xop (*Xanthomonas* outer proteins), XopB-XopBA and three Avr (avirulence proteins), AvrBs1-AvrBs3 families ([T3E Xanthomonas Resource Euroxanth](#)-accessed in January 2023). The effectome or effectorome (i.e. the total assortment of effector proteins) of a pathogen differs greatly in number and composition among pathogens and these variances determine the level of a pathogens' aggressiveness and host specificity (Arroyo-Velez *et al.*, 2020). Namely, pathogenic *Xanthomonas* strains usually possess around 25-35 T3Es that altogether are responsible for an array of actions targeting several different compartments of the host cell, interfering with immune signalling and metabolic pathways, all working towards host susceptibility, making nutrients available to the pathogen and causing disease (Büttner & He, 2009; Timilsina *et al.*, 2020; Teper *et al.*, 2021). Often, gene inactivation or deletion of individual T3Es does not lead to complete loss of virulence phenotypes owing to functional redundancy between effector proteins. Moreover, comparative genomics and developing bioinformatic tools has also prompted the identification of effector genes by prediction of gene homologs or conserved motifs based on sequence-genome similarity and bioinformatic approaches, ergo uncovering host-specificity patterns (Büttner & Bonas, 2010; Timilsina *et al.*, 2020).

Although type III secretion plays an indispensable role for pathogenicity and virulence of most *Xanthomonas* spp., it is interesting to notice that sugarcane pathogens *X. sacchari*, *X. albilineans* are still infective in spite of lacking the T3SS and associated effectors (Meline *et al.*, 2019), positing that under different host pressures, other host-pathogen mechanisms may become more relevant in adaptation and virulence than directly injecting effectors into host cell.

3.3. Evolution of T3SS and effectors in *Xanthomonas* sp.

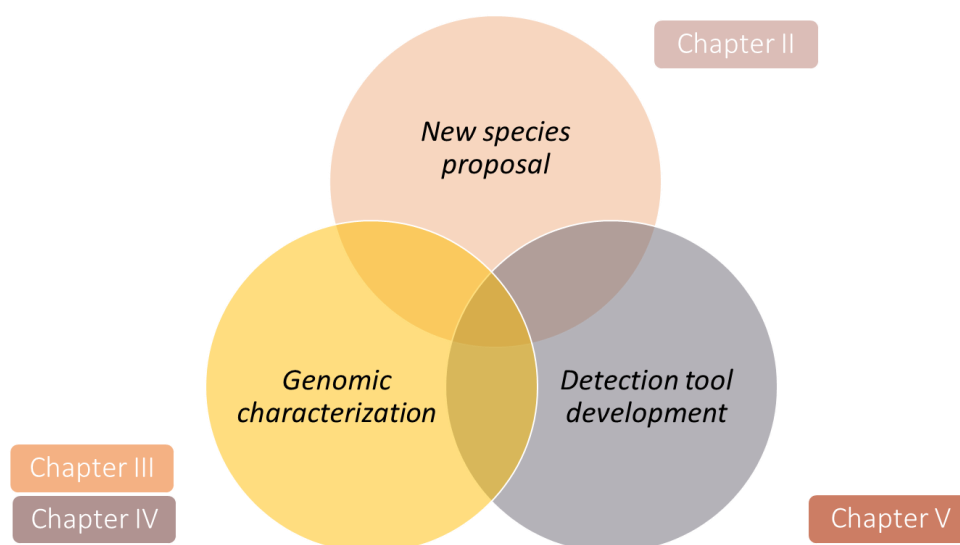
In the context of a plant-pathogen interaction, notably with bacterial pathogens, changes in genome structure at level of gene re-assortments or genome contractions lead to rapid adaptations or lifestyle shifts, and these effects are strongly associated with T3SS and T3E genes. Evolution of Hrp2 T3SS family secretion system and effectors in *Xanthomonas* spp., was the focus of Merda et al. (2017), proposing three ancestral acquisitions of this cluster, its loss in commensal non-pathogenic species and advocating for its ensuing re-acquisition in a few pathogenic species. The report also discloses high homologous recombination and horizontal gene transfer events among the different species. The first acquisition starts with a core effectome of four effectors, XopF1, XopM, XopR, and AvrBs2 present in both pathogenic and non-pathogenic strains (Merda et al., 2017). The set of T3Es composed of XopL, XopN, XopQ, XopX and XopZ in T3SS-harboring pathogenic *X. arboricola* may represent the turn-over from non-pathogenic to pathogenic as these effectors can disrupt host PTI defence response (Merda et al., 2017). It has also been proposed that the content of T3Es profiles could discern pathovar and host-specificity, the so called “repertoire-for-repertoire” premise (Hajri et al., 2009). Considering that non-pathogenic strains of *X. arboricola* are often isolated along with pathogenic *X. arboricola* strains it has also been hypothesised that these commensal strains benefit from the co-colonization with pathogenic strains expressing a complete T3SS (Merda et al., 2017), possibly enjoying a sympatric lifestyle.

Framework and goals of this thesis

This PhD project followed up a previous and thorough study from the same research group that collected an extensive library of *Xanthomonas* strains from walnut trees of different climatic regions. This work unveiled a great diversity of walnut-associated *Xanthomonas* spp. in Portuguese orchards (Fernandes *et al.*, 2021a). This was the starting point for the current PhD thesis.

The main goals for this PhD thesis were:

- Characterization and proposal of the recently discovered walnut-associated *X. euroxanthea* as a new bacterial species;
- Disclose the consortia between *X. euroxanthea* with the already well established walnut pathogen *X. arboricola* pv. *juglandis* (*Xaj*) and their lifestyle in walnut hosts
- Gain insights on general genomic differences, including virulence and pathogenicity determinants, with special emphasis on type III secretion system and type III effectors. The data resulting from this was crucial to develop molecular DNA markers;
- Identify DNA markers specific for *X. euroxanthea* species and transversal to all known strains;
- Optimize and validate a PCR multiplex protocol with *X. euroxanthea*-specific DNA markers for specific identification and typing of *X. euroxanthea* strains.



Thesis publications

The results obtained during this doctoral project were published or are currently in preparation to be submitted in international peer-reviewed scientific journals. All images and articles included in this PhD thesis are in accordance with copyright permission of the respective journals and Creative Commons Attribution-Non Commercial-No Derivatives License CC BY NC ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

The following publications are included in this thesis:

- Martins L, Fernandes C, Blom J, Dia NC, Pothier JF, Tavares F (2020) *Xanthomonas euroxanthea* sp. nov., a new xanthomonad species including pathogenic and non-pathogenic strains of walnut. International Journal of Systematic and Evolutionary Microbiology ijsem004386 **(Chapter II)**
- Fernandes C*, Martins L*, Teixeira M*, Blom J, Pothier JF, Fonseca NA, Tavares F (2021) Comparative Genomics of *Xanthomonas euroxanthea* and *Xanthomonas arboricola* pv. *juglandis* Strains Isolated from a Single Walnut Host Tree. Microorg 2021, Vol 9, Page 624 9: 624 **(Chaper III)**
- Martins L, Teixeira M, Fernandes C, Pothier JF, Koebnik R, Fonseca NA, Tavares F. Genomic features of *Xanthomonas arboricola* and *X. euroxanthea* sharing the same plant host species preferences (walnut, pecan and tomato) seems unbiased by the host species. *In preparation* **(Chapter IV)**
- Silva KG*, Martins L*, Teixeira M, Pothier JF, Tavares F (2022) DNA Markers for Detection and Genotyping of *Xanthomonas euroxanthea*. Microorg 2022, Vol 10, Page 1078 10: 1078 **(Chapter V)**

* These authors contributed equally to this work.

References

- Abby SS, Rocha EPC, 2012. The Non-Flagellar Type III Secretion System Evolved from the Bacterial Flagellum and Diversified into Host-Cell Adapted Systems. *PLOS Genetics* **8**, e1002983.
- Adlung N, Prochaska H, Thieme S *et al.*, 2016. Non-host resistance induced by the *Xanthomonas* effector XopQ is widespread within the genus *Nicotiana* and functionally depends on EDS1. *Frontiers in Plant Science* **7**, 1796.
- Albuquerque P, Caridade CMR, Marcal ARS *et al.*, 2011. Identification of *Xanthomonas fragariae*, *Xanthomonas axonopodis* pv. *phaseoli*, and *Xanthomonas fuscans* subsp. *fuscans* with Novel Markers and Using a Dot Blot Platform Coupled with Automatic Data Analysis. *Applied and Environmental Microbiology* **77**, 5619–5628.
- Albuquerque P, Caridade CMR, Rodrigues AS *et al.*, 2012. Evolutionary and experimental assessment of novel markers for detection of *Xanthomonas euvesicatoria* in plant samples. *PLoS ONE* **7**.
- Albuquerque P, Mendes M v., Santos CL, Moradas-Ferreira P, Tavares F, 2009. DNA signature-based approaches for bacterial detection and identification. *Science of the Total Environment* **407**, 3641–3651.
- Almeida E, Serra CR, Albuquerque P *et al.*, 2018. Multiplex PCR identification and culture-independent quantification of *Bacillus licheniformis* by qPCR using specific DNA markers. *Food Microbiology* **74**, 1–10.
- Alvarez-Martinez CE, Sgro GG, Araujo GG *et al.*, 2021. Secrete or perish: The role of secretion systems in *Xanthomonas* biology. *Computational and Structural Biotechnology Journal* **19**, 279–302.
- Alves BG, Martins L, Ribeiro JN, Tavares FMS, dos Santos M v., 2020. Application of a dot blot hybridization assay for genotyping *Streptococcus uberis* from Brazilian dairy herds. *Journal of Dairy Science*.
- Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q, 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biology* 2020 21:1 **21**, 1–16.
- An SQ, Potnis N, Dow M *et al.*, 2019. Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen *Xanthomonas*. *FEMS Microbiology Reviews* **44**, 1–32.

- Ankenbrand MJ, Keller A, 2016. BcgTree: Automatized phylogenetic tree building from bacterial core genomes. *Genome* **59**, 783–791.
- Arroyo-Velez N, González-Fuente M, Peeters N, Lauber E, Noël LD, 2020. From effectors to effectomes: Are functional studies of individual effectors enough to decipher plant pathogen infectious strategies? *PLOS Pathogens* **16**, e1009059.
- Auch AF, Jan M Von, Klenk H, Göker M, 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. , 117–134.
- Bansal K, Kaur A, Midha S, Kumar S, Korpole S, Patil PB, 2021. *Xanthomonas sontii* sp. nov., a non-pathogenic bacterium isolated from healthy basmati rice (*Oryza sativa*) seeds from India. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* **114**, 1935–1947.
- Bansal K, Kumar S, Patil PB, 2020. Phylogenomic Insights into Diversity and Evolution of Nonpathogenic *Xanthomonas* Strains Associated with Citrus . *mSphere* **5**.
- Bansal K, Kumar S, Patil PB, 2022. Phylo-Taxonogenomics Supports Revision of Taxonomic Status of 20 *Xanthomonas* Pathovars to *Xanthomonas citri*. <https://doi.org/10.1094/PHYTO-08-21-0342-SC> **112**, 1201–1207.
- Blom J, Albaum SP, Doppmeier D *et al.*, 2009. EDGAR: A software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics* **10**.
- Blom J, Kreis J, Spänig S *et al.*, 2016. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic acids research* **44**, W22–W28.
- Briand M, Ruh M, Darrasse A, Jacques M-A, Chen NWG, 2021. Complete and Circularized Genome Sequences of 17 *Xanthomonas* Strains Responsible for Common Bacterial Blight of Bean. *Microbiology Resource Announcements* **10**.
- Büttner D, Bonas U, 2010. Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiology Reviews* **34**, 107–133.
- Büttner D, He SY, 2009. Type III protein secretion in plant pathogenic bacteria. *Plant Physiology* **150**, 1656–1664.
- Cesbron S, Briand M, Essakhi S *et al.*, 2015. Comparative Genomics of Pathogenic and Nonpathogenic Strains of *Xanthomonas arboricola* Unveil Molecular and Evolutionary Events Linked to Pathoadaptation. *Frontiers in Plant Science* **6**, 1126.

- Chang JH, Desveaux D, Creason AL, 2014. The ABCs and 123s of Bacterial Secretion Systems in Plant Pathogenesis. *Annual Review of Phytopathology* **52**, 317–345.
- Chun J, Oren A, Ventosa A *et al.*, 2018. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology* **68**, 461–466.
- Chun J, Rainey FA, 2014. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *International Journal of Systematic and Evolutionary Microbiology* **64**, 316–324.
- Cianciotto NP, 2005. Type II secretion: A protein secretion system for all seasons. *Trends in Microbiology* **13**, 581–588.
- Cuesta-Morrondo S, Redondo C, Palacio-Bielsa A, Garita-Cambronero J, Cubero J, 2022. Complete Genome Sequence Resources of Six Strains of the Most Virulent Pathovars of *Xanthomonas arboricola* Using Long- and Short-Read Sequencing Approaches. *Phytopathology*®.
- Cunnac S, Chakravarthy S, Kvitko BH, Russell AB, Martin GB, Collmer A, 2011. Genetic disassembly and combinatorial reassembly identify a minimal functional repertoire of type III effectors in *Pseudomonas syringae*. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 2975–2980.
- Das A, Rangaraj N, Sonti R V, 2009. Multiple Adhesin-Like Functions of *Xanthomonas oryzae* pv. *oryzae* Are Involved in Promoting Leaf Attachment, Entry, and Virulence on Rice. / *73 MPMI* **22**, 73–85.
- Deloger M, el Karoui M, Petit MA, 2009. A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. *Journal of Bacteriology* **91**, 91–99.
- Delplace F, Huard-Chauveau C, Berthomé R, Roby D, 2022. Network organization of the plant immune system: from pathogen perception to robust defense induction. *Plant Journal* **109**, 447–470.
- Deng W, Marshall NC, Rowland JL *et al.*, 2017. Assembly, structure, function and regulation of type III secretion systems. *Nature Reviews Microbiology* **15**, 323–337.
- Dia NC, Rezzonico F, Smits THM, Pothier JF, 2020. Complete or High-Quality Draft Genome Sequences of Six *Xanthomonas hortorum* Strains Sequenced with Short- and Long-Read Technologies. *Microbiology Resource Announcements* **9**.

- Dieckmann MA, Beyvers S, Nkouamedjo-Fankep RC *et al.*, 2021. EDGAR3.0: comparative genomics and phylogenomics on a scalable infrastructure. *Nucleic Acids Research* **49**, W185–W192.
- Dowson DW, 1939. On the systematic position and generic names of the Gram negative bacterial plant pathogens. *Zentralblatt für Bakteriologie, Parasitenkunde und Infektionskrankheiten*, **2** **100**, 177–193.
- Essakhi S, Cesbron S, Fischer-Le Saux M, Bonneau S, Jacques M-A, Manceau C, 2015. Phylogenetic and Variable-Number Tandem-Repeat Analyses Identify Nonpathogenic *Xanthomonas arboricola* Lineages Lacking the Canonical Type III Secretion System. *Applied and Environmental Microbiology* **81**, 5395–5410.
- Fernandes C, Albuquerque P, Mariz-Ponte N, Cruz L, Tavares F, 2021a. Comprehensive diversity assessment of walnut-associated xanthomonads reveal the occurrence of distinct *Xanthomonas arboricola* lineages and of a new species (*Xanthomonas euroxanthea*) within the same tree. *Plant Pathology* **70**, 943–958.
- Fernandes C, Albuquerque P, Sousa R, Cruz L, Tavares F, 2017. Multiple DNA Markers for Identification of *Xanthomonas arboricola* pv. *juglandis* Isolates and its Direct Detection in Plant Samples. *Plant Disease* **101**, 858–865.
- Fernandes C, Martins L, Teixeira M *et al.*, 2021b. Comparative Genomics of *Xanthomonas euroxanthea* and *Xanthomonas arboricola* pv. *juglandis* Strains Isolated from a Single Walnut Host Tree. *Microorganisms* 2021, Vol. 9, Page 624 **9**, 624.
- Fernandes C, Sousa R, Tavares F, Cruz L, 2018. First report of *Xanthomonas arboricola* causing bacterial blight on pecan trees in Portugal. *Plant Disease* **102**, 2632.
- Fischer-Le Saux M, Bonneau S, Essakhi S, Manceau C, Jacques M-A, 2015. Aggressive Emerging Pathovars of *Xanthomonas arboricola* Represent Widespread Epidemic Clones Distinct from Poorly Pathogenic Strains, as Revealed by Multilocus Sequence Typing. *Applied and Environmental Microbiology* **81**, 4651–4668.
- da Gama MAS, Mariano R de LR, da Silva Júnior WJ *et al.*, 2018. Taxonomic Repositioning of *Xanthomonas campestris* pv. *viticola* (Nayudu 1972) Dye 1978 as *Xanthomonas citri* pv. *viticola* (Nayudu 1972) Dye 1978 comb. nov. and Emendation of the Description of *Xanthomonas citri* pv. *anacardii* to Include Pigmented Isolates Pa. *Phytopathology*® **108**, 1143–1153.

- Garita-Cambronero J, Palacio-Bielsa A, López MM, Cubero J, 2016a. Draft genome sequence of two strains of *Xanthomonas arboricola* isolated from *Prunus persica* which are dissimilar to strains that cause bacterial spot disease on *Prunus* spp. *Genome Announcements* **4**, 974–990.
- Garita-Cambronero J, Palacio-Bielsa A, López MM, Cubero J, 2016b. Draft genome sequence for virulent and avirulent strains of *Xanthomonas arboricola* isolated from *Prunus* spp. in Spain. *Standards in Genomic Sciences* **11**, 1–10.
- Garita-Cambronero J, Palacio-Bielsa A, López MM, Cubero J, 2016c. Comparative genomic and phenotypic characterization of pathogenic and non-pathogenic strains of *Xanthomonas arboricola* reveals insights into the infection process of bacterial spot disease of stone fruits. *PLoS ONE* **11**, e0161977.
- Garita-Cambronero J, Palacio-Bielsa A, López MM, Cubero J, 2017. Pan-genomic analysis permits differentiation of virulent and non-virulent strains of *Xanthomonas arboricola* that cohabit *Prunus* spp. and elucidate bacterial virulence factors. *Frontiers in Microbiology* **8**, 1–17.
- Garita-Cambronero J, Sena-Vélez M, Ferragud E, Sabuquillo P, Redondo C, Cubero J, 2019. *Xanthomonas citri* subsp. *Citri* and *Xanthomonas arboricola* pv. *Pruni*: Comparative analysis of two pathogens producing similar symptoms in different host plants. *PLoS ONE* **14**.
- Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM, 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* **57**, 81–91.
- Hajri A, Brin C, Hunault G *et al.*, 2009. A «repertoire for repertoire» hypothesis: Repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS ONE* **4**.
- Hayward AC, 1993. The hosts of *Xanthomonas*. In: *Xanthomonas*. Springer Netherlands, 1–119.
- Hu S, Li K, Zhang Y *et al.*, 2022. New Insights Into the Threshold Values of Multi-Locus Sequence Analysis, Average Nucleotide Identity and Digital DNA–DNA Hybridization in Delineating *Streptomyces* Species. *Frontiers in Microbiology* **13**, 1869.

- Hurley B, Subramaniam R, Guttman DS, Desveaux D, 2014. Proteomics of effector-triggered immunity (ETI) in plants. *Virulence* **5**, 752.
- Jones JDG, Dangl JL, 2006. The plant immune system. *Nature* 2006 **444**:7117 **444**, 323–329.
- Kałużna M, Fischer-Le Saux M, Pothier JF *et al.*, 2021. *Xanthomonas arboricola* pv. *juglandis* and pv. *corylina*: Brothers or distant relatives? Genetic clues, epidemiology, and insights for disease management. *Molecular Plant Pathology* **22**, 1481–1499.
- Kałużna M, Pothier JF, 2022. Complete Genome Sequence Data of Two *Xanthomonas arboricola* Strains Isolated from Blueberry Plants Displaying Bacterial Leaf Blight in Poland. *Phytopathology*®.
- Konstantinidis KT, Tiedje JM, 2005. Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 2567–2572.
- Lamichhane JR, 2014. *Xanthomonas arboricola* diseases of stone fruit, almond, and walnut trees: Progress toward understanding and management. *Plant Disease* **98**, 1600–1610.
- Mafakheri H, Taghavi SM, Zarei S *et al.*, 2022. *Xanthomonas bonasiae* sp. nov. and *Xanthomonas youngii* sp. nov., isolated from crown gall tissues. *International Journal of Systematic and Evolutionary Microbiology* **72**, 005418.
- Martin R, Qi T, Zhang H *et al.*, 2020. Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. *Science* **370**.
- Martins L, Fernandes C, Albuquerque P, Tavares F, 2019. Assessment of *Xanthomonas arboricola* pv. *juglandis* Bacterial Load in Infected Walnut Fruits by Quantitative PCR. *Plant disease* **103**, 2577–2586.
- Martins L, Fernandes C, Blom J, Dia NC, Pothier JF, Tavares F, 2020. *Xanthomonas euroxanthea* sp. nov., a new xanthomonad species including pathogenic and non-pathogenic strains of walnut. *International Journal of Systematic and Evolutionary Microbiology* **70**, 9.
- Mattoo S, Lee YM, Dixon JE, 2007. Interactions of bacterial effector proteins with host proteins. *Current Opinion in Immunology* **19**, 392–401.

- Mbega ER, Mabagala RB, Adriko J, Lund OS, Wulff EG, Mortensen CN, 2012. Five Species of Xanthomonads Associated with Bacterial Leaf Spot Symptoms in Tomato from Tanzania. *Plant Disease* **96**, 760–760.
- Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M, 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**, 60.
- Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M, 2022. TYGS and LPSN: A database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Research* **50**, D801–D807.
- Meline V, Delage W, Brin C *et al.*, 2019. Role of the acquisition of a type 3 secretion system in the emergence of novel pathogenic strains of Xanthomonas. *Molecular Plant Pathology* **20**, 33–50.
- Merda D, Bonneau S, Guimbaud JF *et al.*, 2016. Recombination-prone bacterial strains form a reservoir from which epidemic clones emerge in agroecosystems. *Environmental Microbiology Reports* **8**, 572–581.
- Merda D, Briand M, Bosis E *et al.*, 2017. Ancestral acquisitions, gene flow and multiple evolutionary trajectories of the type three secretion system and effectors in Xanthomonas plant pathogens. *Molecular Ecology* **26**, 5939–5952.
- Van den Mooter M, Swings J, 1990. Numerical analysis of 295 phenotypic features of 266 Xanthomonas strains and related strains and an improved taxonomy of the genus. *International Journal of Systematic Bacteriology* **40**, 348–369.
- Morinière L, Burlet A, Rosenthal ER *et al.*, 2020. Clarifying the taxonomy of the causal agent of bacterial leaf spot of lettuce through a polyphasic approach reveals that Xanthomonas cynarae Trébaol *et al.* 2000 emend. Timilsina *et al.* 2019 is a later heterotypic synonym of Xanthomonas hortorum Vauterin *et al.* *Systematic and Applied Microbiology* **43**, 126087.
- Myung IS, Jeong IH, Moon SY, Lee SW, Shim HS, 2010. A new disease, arboricola leaf spot of bell pepper, caused by Xanthomonas arboricola. *Plant Disease* **94**, 271.
- Na SI, Kim YO, Yoon SH, Ha S min, Baek I, Chun J, 2018. UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *Journal of Microbiology* **2018** 56:4 **56**, 280–285.

- Nazir R, Mazurier S, Yang P, Lemanceau P, van Elsas JD, 2017. The Ecological Role of Type Three Secretion Systems in the Interaction of Bacteria with Fungi in Soil and Related Habitats Is Diverse and Context-Dependent. *Frontiers in Microbiology* **8**.
- Ngou BPM, Ahn HK, Ding P, Jones JDG, 2021. Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *Nature* **592**, 110–115.
- Nunez P, Muster C, Lisperguer MJ, Vargas E, Bustos S, 2021. Complete Genome of *Xanthomonas arboricola* pv. *corylina* Strain A7 Isolated from Southern Chile. <https://doi.org/10.1094/MPMI-12-20-0363-A> **35**, 94–95.
- Parker CT, Tindall BJ, Garrity GM, 2019. International Code of Nomenclature of Prokaryotes. *International journal of systematic and evolutionary microbiology* **69**, S1–S111.
- Parkinson N, Cowie C, Heeney J, Stead D, 2009. Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *International Journal of Systematic and Evolutionary Microbiology* **59**, 264–274.
- Parte AC, Carbasse JS, Meier-Kolthoff JP, Reimer LC, Göker M, 2020. List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology* **70**, 5607–5612.
- Peduzzi C, Sagia A, Burokienė D *et al.*, 2022. Novel genetic traits in *Xanthomonas* unveiled by the complete genome sequencing of three clade-1 xanthomonads. *bioRxiv*, 2022.10.10.511549.
- Pfeilmeier S, Caly DL, Malone JG, 2016. Bacterial pathogenesis of plants: future challenges from a microbial perspective. *Molecular Plant Pathology* **17**, 1298–1313.
- Pothier JF, Kałużna M, Prokić A, Obradović A, Rezzonico F, 2022. Complete Genome and Plasmid Sequence Data of Three Strains of *Xanthomonas arboricola* pv. *corylina*, the Bacterium Responsible for Bacterial Blight of Hazelnut. *Phytopathology* **112**, 956–960.
- Rana R, Madhavan VN, Saroha T *et al.*, 2022. *Xanthomonas indica* sp. nov., a Novel Member of Non-Pathogenic *Xanthomonas* Community from Healthy Rice Seeds. *Current Microbiology* 2022 79:10 **79**, 1–10.
- Richter M, Rosselló-Móra R, 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences USA* **106**, 19126–19131.

- Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J, 2016. JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* **32**, 929–931.
- Rodriguez-R LM, Konstantinidis KT, 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes.
- Rondelet A, Condemine G, 2013. Type II secretion: the substrates that won't go away. *Research in Microbiology* **164**, 556–561.
- Rosselló-Móra R, Amann R, 2015. Past and future species definitions for Bacteria and Archaea. *Systematic and Applied Microbiology* **38**, 209–216.
- Rosselló-Móra R, Urdiain M, López-López A, 2011. DNA-DNA Hybridization. In: *Methods in Microbiology*. Academic Press, 325–347.
- Ruh M, Briand M, Bonneau S, Darrasse A, Jacques M-A, Chen NWG, 2021. Complete and Circularized Genome Sequences of Three Xanthomonas Strains Pathogenic on Soybean and Alfalfa. *Microbiology Resource Announcements* **10**.
- Schaad NW, Jones JB, Chun W, 2001. Laboratory guide for the identification of plant pathogenic bacteria. *Laboratory guide for the identification of plant pathogenic bacteria.*, 299–315.
- Schleifer KH, 2009. Classification of Bacteria and Archaea: Past, present and future. *Systematic and Applied Microbiology* **32**, 533–542.
- Sena-Vélez M, Redondo C, Gell I *et al.*, 2015. Biofilm formation and motility of Xanthomonas strains with different citrus host range. *Plant Pathology* **64**, 767–775.
- Shames SR, Finlay BB, 2012. Bacterial effector interplay: A new way to view effector function. *Trends in Microbiology* **20**, 214–219.
- Silva KG, Martins L, Teixeira M, Pothier JF, Tavares F, 2022. DNA Markers for Detection and Genotyping of Xanthomonas euroxanthea. *Microorganisms* 2022, Vol. 10, Page 1078 **10**, 1078.
- Stefani E, Obradović A, Gašić K, Altin I, Nagy IK, Kovács T, 2021. Bacteriophage-mediated control of phytopathogenic xanthomonads: A promising green solution for the future. *Microorganisms* **9**.
- Studholme DJ, Wicker E, Abrare SM *et al.*, 2020. Transfer of Xanthomonas campestris pv. arecae and X. campestris pv. musacearum to X. vasicola (Vauterin) as X.

- vasicola pv. arecae comb. nov. And *X. vasicola* pv. musacearum comb. nov. And description of *X. vasicola* pv. vasculorum pv. nov. *Phytopathology* **110**, 1153–1160.
- Szczesny R, Jordan M, Schramm C *et al.*, 2010. Functional characterization of the Xcs and Xps type II secretion systems from the plant pathogenic bacterium *Xanthomonas campestris* pv. vesicatoria. *New Phytologist* **187**, 983–1002.
- Teixeira M, Fernandes C, Chaves C, Pinto J, Tavares F, Fonseca NA, 2021. Complete Genome Sequence Obtained by Nanopore and Illumina Hybrid Assembly of *Xanthomonas arboricola* pv. juglandis CPBF 427, Isolated from Buds of a Walnut Tree (SR Gill, Ed.). *Microbiology Resource Announcements* **10**.
- Teixeira M, Martins L, Fernandes C *et al.*, 2020. Complete Genome Sequences of Walnut-Associated *Xanthomonas euroxanthea* Strains CPBF 367 and CPBF 426 Obtained by Illumina/Nanopore Hybrid Assembly (DA Baltrus, Ed.). *Microbiology Resource Announcements* **9**.
- Teper D, Pandey SS, Wang N, 2021. The HrpG/HrpX Regulon of Xanthomonads—An Insight to the Complexity of Regulation of Virulence Traits in Phytopathogenic Bacteria. *Microorganisms* **9**, 187.
- Tian L, Huang C, Mazloom R, Heath LS, Vinatzer BA, 2020. LINbase: a web server for genome-based identification of prokaryotes as members of crowdsourced taxa. *Nucleic Acids Research* **48**, W529–W537.
- Timilsina S, Goss EM, Koebnik R, Potnis N, Jones JB, 2022. Editorial: Population and comparative genomics of plant pathogenic bacteria. *Frontiers in Microbiology* **13**, 3465.
- Timilsina S, Kara S, Jacques MA *et al.*, 2019. Reclassification of *Xanthomonas gardneri* (ex Šutič 1957) Jones *et al.* 2006 as a later heterotypic synonym of *Xanthomonas cynarae* Trébaol *et al.* 2000 and description of *X. cynarae* pv. *cynarae* and *X. cynarae* pv. *gardneri* based on whole genome analyses. *International Journal of Systematic and Evolutionary Microbiology* **69**, 343–349.
- Timilsina S, Potnis N, Newberry EA *et al.*, 2020. *Xanthomonas* diversity, virulence and plant–pathogen interactions. *Nature Reviews Microbiology* **18**, 415–427.
- Triplett LR, Verdier V, Campillo T *et al.*, 2015. Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali and proposal of *Xanthomonas*

- maliensis* sp. nov. Antonie van Leeuwenhoek, *International Journal of General and Molecular Microbiology* **107**, 869–881.
- Vauterin L, Hoste B, Kersters K, Swings J, 1995. Reclassification of *Xanthomonas*. *International Journal of Systematic Bacteriology* **45**, 472–489.
- Vauterin L, Swings J, Kersters K, 1991. Grouping of *Xanthomonas campestris* pathovars by SDS-PAGE of proteins. *Journal of General Microbiology* **137**, 1677–1687.
- Venter SN, Palmer M, Steenkamp ET, 2022. Relevance of prokaryotic subspecies in the age of genomics. *New Microbes and New Infections* **48**, 101024.
- Wayne LG, Brenner DJ, Colwell RR *et al.*, 1988. International Committee on Systematic Bacteriology: Announcement of the Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Systematic and Applied Microbiology* **10**, 99–100.
- Wei HL, Zhang W, Collmer A, 2018. Modular Study of the Type III Effector Repertoire in *Pseudomonas syringae* pv. tomato DC3000 Reveals a Matrix of Effector Interplay in Pathogenesis. *Cell Reports* **23**, 1630–1638.
- White FF, Potnis N, Jones JB, Koebnik R, 2009. The type III effectors of *Xanthomonas*. *Molecular Plant Pathology* **10**, 749–766.
- Wick RR, Judd LM, Gorrie CL, Holt KE, 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads (AM Phillippy, Ed.). *PLOS Computational Biology* **13**, e1005595.
- Wu HP, Derilo RC, Chen HL *et al.*, 2021. Injectisome T3SS subunits as potential chaperones in the extracellular export of *Pectobacterium carotovorum* subsp. *carotovorum* bacteriocins Carocin S1 and Carocin S3 secreted via flagellar T3SS. *BMC Microbiology* **21**.
- Young JM, Park DC, Shearman HM, Fargier E, 2008. A multilocus sequence analysis of the genus *Xanthomonas*. *Systematic and Applied Microbiology* **31**, 366–377.
- Yuan M, Jiang Z, Bi G *et al.*, 2021. Pattern-recognition receptors are required for NLR-mediated plant immunity. *Nature* 2021 592:7852 **592**, 105–109.
- Zarei S, Taghavi SM, Osdaghi E, 2022a. Reassessment of the occurrence and distribution of *Xanthomonas arboricola* on fruit trees in Iran. *Plant Pathology*.

Zarei S, Taghavi SM, Rahimi T *et al.*, 2022b. Taxonomic Refinement of *Xanthomonas arboricola*. <https://doi.org/10.1094/PHYTO-12-21-0519-R>.



Chapter II

Proposal of *Xanthomonas*
euroxanthea sp. nov., a new
xanthomonad species including
pathogenic and non-pathogenic
strains of walnut

Chapter II include the following publication:

Martins L, Fernandes C, Blom J, Dia NC, Pothier JF, Tavares F (2020) *Xanthomonas euroxanthea* sp. nov., a new xanthomonad species including pathogenic and non-pathogenic strains of walnut. International Journal of Systematic and Evolutionary Microbiology ijsen004386

Article in national newspaper:

Teresa Sofia Serafim. "Descoberta em Portugal nova bactéria que pode adoecer as nozeiras", Público, 2020, <https://www.publico.pt/2020/10/21/ciencia/noticia/descoberta-portugal-nova-bacteria-adoecer-nogueiras-1936215>.

Part of the results from this chapter were presented at:

Martins L, Fernandes C, Blom J, Dia NC, Pothier JF, Tavares F. Genomic evidence for a new *Xanthomonas* species isolated from walnut (*Juglans regia* L.) 3rd Annual EuroXanth Conference, Lednice, Czech republic, 9-11 September 2019 (Oral presentation)

Martins, L. Fernandes, C. Teixeira, M. Blom, J. Dia, N.C. Pothier, J.F. Fonseca, N.A. Tavares, F. *Xanthomonas euroxanthea*: genomic insights from a new phytopathogenic bacteria species. IV Encontro Biologia Funcional e Biotecnologia de Plantas - Faculdade de Ciências da Universidade do Porto, Portugal, 2 December 2020 (Oral presentation)

Proposal of *Xanthomonas euroxanthea* sp. nov., a new xanthomonad species including pathogenic and non-pathogenic strains of walnut

Repositories: The GenBank/EMBL/DDBJ accession numbers for the 16S rRNA gene sequence of strain CPBF 424^T is MT036365. GenBank accession number for genome sequences of CPBF 367 and CPBF 424^T are UNRN000000000.1 and UIHB000000000.1, respectively.

Abstract

We describe a novel species isolated from walnut (*Juglans regia*) which comprises non-pathogenic and pathogenic strains on walnut.

The isolates, obtained from a single ornamental walnut tree showing disease symptoms, grew on yeast extract-dextrose-carbonate agar as mucoid yellow colonies characteristic of *Xanthomonas* species. Pathogenicity assays showed that while strain CPBF 424^T causes disease in walnut, strain CPBF 367 was non-pathogenic on walnut leaves. Biolog Gen III metabolic profiles disclosed some differences between strains CPBF 367 and CPBF 424^T and other xanthomonads. Multilocus sequence analysis with seven housekeeping genes (*fyuA*, *gyrB*, *rpoD*, *atpD*, *dnaK*, *efp*, *glnA*) grouped these strains in a distinct cluster from *Xanthomonas arboricola* pv. *juglandis* and closer to *Xanthomonas prunicola* and *Xanthomonas arboricola* pv. *populi*. Average nucleotide identity (ANI) analysis results displayed similarity values below 93% to *X. arboricola* strains. Meanwhile ANI and digital DNA-DNA hybridization similarity values were below 89 and 50% to non-*arboricola* *Xanthomonas* strains, respectively, revealing that they do not belong to any previously described *Xanthomonas* species. Furthermore, the two strains show over 98% similarity to each other. Genomic analysis shows that strain CPBF 424^T harbours a complete type III secretion system and several type III effector proteins, in contrast with strain CPBF 367, shown to be non-pathogenic in plant bioassays. Taking these data altogether we propose that strains CPBF 367 and CPBF 424^T belong to a new species herein named *Xanthomonas euroxanthea* sp. nov., with CPBF 424^T (= LMG 31037^T = CCOS 1891^T = NCPPB4675^T) as the type strain.

Introduction

The taxonomy of the genus *Xanthomonas* has been extensively studied and categorized over the last century, raising some disputes regarding classification and nomenclature of its members, largely due to the wide plant host diversity associated with the genus. Recorded *Xanthomonas* hosts include numerous economically relevant crop species, severely impacting agricultural productivity worldwide (Vauterin *et al.*, 2000).

Introduction of molecular methods such as DNA-DNA hybridization allowed the first big reclassification within *Xanthomonas* (Vauterin *et al.*, 1995). Recently, studies relying on multilocus sequence analysis (MLSA), whole-genome average nucleotide identity (ANI) and biochemical analysis have proved fruitful in the reclassification of *Xanthomonas* species and description amendments of several strains (Constantin *et al.*, 2016; da Gama *et al.*, 2018). These techniques allowed us to reach a deeper discriminatory insight on the genetic diversity of *Xanthomonas* at the intrasubspecific level, consequently leading to the proposal of novel species (Triplett *et al.*, 2015; Vicente *et al.*, 2017; López *et al.*, 2018).

The increasing reports of non-pathogenic *Xanthomonas* strains, such as *Xanthomonas arboricola* isolates recovered from asymptomatic and symptomatic host tissues, points to the uncertainty of several *Xanthomonas* strains regarding their host-associated lifestyle and likely host diversity (Essakhi *et al.*, 2015; Garita-Cambronero *et al.*, 2016; Jacques *et al.*, 2016). Interestingly, *X. arboricola* is the species with most non-pathogenic strains within the genus *Xanthomonas* (Merda *et al.*, 2016). Characterization of several of these non-pathogenic *X. arboricola* strains has revealed that they are phylogenetically distant from the disease-causing strains that were isolated from their specific hosts (Jacques *et al.*, 2016). Furthermore, plants are host to a consortium of both distant and genetically similar bacteria, displaying different host-associated phenotypes. It has been suggested that the sympatry of pathogenic and non-pathogenic strains may favour horizontal gene transfer events, likely leading to the emergence of new pathogenic lineages (Jacques *et al.*, 2016; Merda *et al.*, 2016).

The advent of increasingly powerful genomic tools and improved bacterial genome assemblers and accurate annotation has led to a more confident examination of bacterial evolution and phylogeny, allowing microbiologists and plant pathologists to assign specific differences to the existing taxonomy, even at the pathovar level. A pathovar of a pathogenic species refers to an infrasubspecific group that includes strains with the same host range (Yeates, 1978; Schulze-Lefert & Panstruga, 2011). To the best of our knowledge, all *Xanthomonas arboricola* pv. *juglandis* strains described previously have been isolated from diseased walnut trees.

Furthermore, MLSA results revealed that several *X. arboricola* isolated from walnut, including both pathogenic and non-pathogenic strains, do not cluster with the members of *juglandis* pathovar (Essakhi *et al.*, 2015; Fernandes *et al.*, 2018b). Indeed, a previous study has reported the identification of two *X. arboricola* strains isolated from symptomatic leaves of pecan tree (*Carya illinoensis*), which are pathogenic to both pecan and walnut trees (Fernandes *et al.*, 2018b). These data raise the need to understand how the genetic makeup of plant bacterial pathogens, such as *Xanthomonas*, links with a diversifying range of plant hosts. Walnut (*Juglans* species) is an economically important tree crop worldwide, with the Persian walnut (*Juglans regia* L.) being the main cultivated species for walnut fruits and timber production (Leslie *et al.*, 2006; Belisario *et al.*, 2010; Lamichhane, 2014). *X. arboricola* pv. *juglandis* is a major threat to walnut orchards and acknowledged as the etiological agent of walnut bacterial blight (WBB) and other walnut diseases such as brown apical necrosis (Moragrega *et al.*, 2011) and vertical oozing canker (Hajri *et al.*, 2010), altogether causing high yield losses (Leslie *et al.*, 2006).

In this study we characterize two novel strains, CPBF 367 and CPBF 424^T, isolated from asymptomatic buds of one walnut tree showing WBB symptoms. Genotyping and comparative genomics indicate that these strains, showing a distinct pathogenicity phenotype in walnut, belong to the genus *Xanthomonas*, although they could not be assigned to any of the known species of this genus. This work gathers phenotypic, genotypic and genomic evidence to support that strains CPBF 367 and CPBF 424^T are members of a new species of the genus *Xanthomonas*.

Isolation and growth of bacteria

Several isolates were obtained over three consecutive years (2014-2016) from an isolated ornamental walnut tree (*Juglans regia*) showing bacterial blight symptoms in leaves and fruits, located in a municipal garden in the city of Loures, Portugal (Supplementary Figure II.S1). Distinct organs were sampled including symptomatic leaves (CPBF 1521 in 2014) and asymptomatic buds (CPBF 367, CPBF 424^T, CPBF 426 and CPBF 427 in 2016). The procedure for preparation of plant material for bacteria isolation has been described previously (Fernandes *et al.*). Briefly, the excised plant material was disinfected by immersion in 70% ethanol followed by washing with sterile distilled water (SDW) and then macerated in extraction bags with 5ml SDW (Fernandes *et al.*, 2017). Suspension and correspondent dilutions were streaked on yeast extract dextrose carbonate (YDC) agar (Stolp' & Starr, 1964) and incubated at 26 ± 2°C for 4 days. Single characteristic mucoid yellow colonies of *Xanthomonas* were streaked on

fresh nutrient agar (NA) medium to ensure purity. Selected isolates (CPBF 424^T and CPBF 367) were stored at -80°C in cryovials containing LB medium (Difco) and glycerol to a final concentration of 30%. The strains were deposited in international bacterial collections with the accession numbers CPBF 367 (LMG 31036 = CCOS 1890) and CPBF 424^T (= LMG 31037^T = CCOS 1891^T = NCPPB 4675^T). Isolates CPBF 367 and CPBF 424^T were grown in YDC for analysis of colony morphology and pigment production for a 10 day period. The bacterial strains used throughout this study are disclosed in Table II.1.

Table II.1. Strains used for pathogenicity tests and Biolog assays.

Strain	Species	Host of isolation	Geographic origin	Year of isolation
CPBF 367 ^a	<i>X. euroxanthea</i>	<i>Juglans regia</i>	Loures, Portugal	2016
CPBF 424 ^{Ta}	<i>X. euroxanthea</i>	<i>Juglans regia</i>	Loures, Portugal	2016
CPBF 427 ^a	<i>X. arboricola</i> pv. <i>juglandis</i>	<i>Juglans regia</i>	Loures, Portugal	2016
CPBF 1521 ^a	<i>X. arboricola</i> pv. <i>juglandis</i>	<i>Juglans regia</i>	Loures, Portugal	2014
CPBF 1480	<i>X. arboricola</i> pv. <i>juglandis</i>	<i>Juglans regia</i>	Azeitão, Portugal	2014
LMG 747 ^T	<i>X. arboricola</i> pv. <i>juglandis</i>	<i>Juglans regia</i>	New Zealand	1956
CFBP 3123 ^{PT}	<i>X. arboricola</i> pv. <i>populi</i>	<i>Populus × euroamericana</i> cv. <i>robusta</i>	Netherlands	1979

^a Other known names for these strains: CPBF 367 = LMG 31036 = CCOS 1890; CPBF 424^T = LMG 31037^T = CCOS 1891^T = NCPPB 4675^T; CPBF 427 = LMG 31039 = CCOS 1893; CPBF 1521 = LMG 31040 = CCOS 1894 = NCPPB 4676. (CPBF, Portuguese Collection of Phytopathogenic Bacteria in Oeiras, Portugal; BCCM/LMG, the Belgian Coordinated Collections of Microorganisms/ LMG Bacteria Collection in Gent, Belgium, CCOS, Culture Collection of Switzerland in Wädenswil; NCPPB, National Collection of Plant Pathogenic Bacteria in York, United Kingdom.)

Pathogenicity assays

Plantlets used in the pathogenicity tests were grown from seeds collected from *Juglans regia* cv. Hartley on the same day. After 30 days of cold stratification treatment at 3-5°C to break dormancy, seeds were sown in sterilized sand substrate and germinated during 60 days at alternated temperatures, 16 h day at 30°C and 8 h night at 20°C (International Seed Testing Association, 1999). Walnut plantlets were then maintained in a climatic chamber under controlled environmental conditions of 16 hour photoperiod (16 h of light at 24°C and 8 h of darkness at 18°C). Pathogenicity determination of CPBF 424^T and CPBF 367 strains was carried out as previously described (Fernandes *et al.*). Briefly, inoculum suspensions of approximately 10⁸ c.f.u ml⁻¹ were prepared with SDW. Plantlets with at least four young leaves expanded were inoculated by spraying with a manual atomizer until runoff. SDW was used for negative control, while *X. arboricola* pv. *juglandis* strains LMG 747^T and CPBF 1480 were used as positive controls of infection. Three replicates were included for each strain tested. Symptoms were registered after 7 days and further followed for over 5 weeks. Necrotic bacterial spots were observed 7 days post inoculation on leaves of plants inoculated with walnut isolate CPBF 424^T, and with positive-control strains LMG 747^T and CPBF 1480 (Figure II.1). Walnut isolate CPBF 367 and negative controls did not produce any disease symptoms on walnut plantlets. In order to fulfil Koch's postulates typical *Xanthomonas* yellow mucoid bacterial colonies were re-isolated from the symptoms and confirmed by sequencing analysis of *gyrB* and *fyuA*.

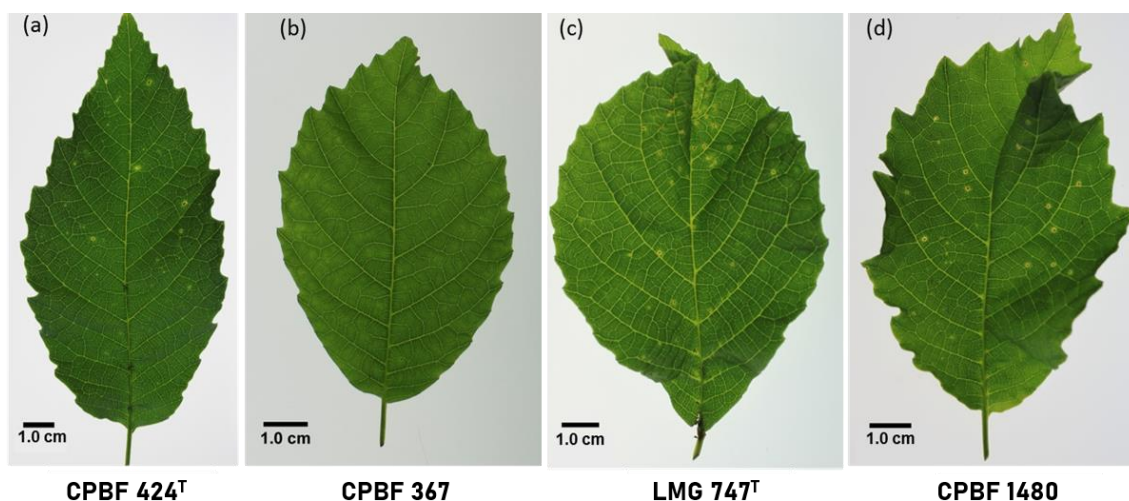


Figure II.1. Walnut plantlets leaves seven days post-inoculation with (a) isolate CPBF 424^T showing walnut bacteria blight (WBB) symptoms, (b) isolate CPBF 367 no WBB symptoms were observed, (c) *Xanthomonas arboricola* pv. *juglandis* LMG 747^T and (d) *X. arboricola* pv. *juglandis* strain CPBF 1480, used as positive controls, show characteristic WBB symptoms.

Phenotypic characterization

Phenotypic characterization was carried out with Biolog GEN III Microplates (Biolog) according to manufacturer's instruction, on walnut isolates CPBF 367, CPBF 424^T, CPBF 427, CPBF 1521 and reference strains *X. arboricola* pv. *juglandis* LMG 747^T and *X. arboricola* pv. *populi* CFBP 3123^{PT}. For all strains except CFBP 3123^{PT}, three MicroPlates were assayed on different dates. Selected strains were first grown on solid NYGA medium (5 g/L peptone, 3 g/L yeast extract, 20 g/L glycerol and 15 g/L agar) for 48h at 28 °C. The strains were subsequently grown on solid Biolog Dehydrated Growth agar for 24h at 28°C. Fresh colonies were then transferred into Inoculating Fluid A vials using a cotton-tipped swab. The density of the inoculum was checked and adjusted to a transmittance of 95-98% using a turbidimeter. A total of 100 µl prepared inoculum was then dispensed into each well of the Biolog MicroPlate. MicroPlates were then incubated at 28 °C and read using a MicroStation 2 Reader (Biolog) at 24h, 48h, 72h, 120h, 192h and 240h.

Considering compound metabolism, values 160% superior to the negative control (water) value were considered positive, while those below 130% the negative control were taken as negative. Values in between were considered borderline (Garita-Cambronero *et al.*, 2016). The data measured at 72h was shown to provide the most consistent results between replicates and thus was selected for further analysis. Isolates CPBF 367 and CPBF 424^T seem to metabolize mucic acid and D-saccharic acid contrary to *X. arboricola* pv. *juglandis* CPBF 427, CPBF 1521 and LMG 747^T. Additionally, comparison with results from other *Xanthomonas* strains [4-6] revealed differences in metabolizing D-salicin, sucrose, acetic acid, melibiose, formic acid, D-arabitol, D-glucose-6-phosphate, maltose, and gentiobiose (Table II.2). The metabolization of some substrates was inconclusive due to the inconsistencies observed between replicates.

Table II.2. Biolog profiles of *anthomonas euroxanthea* strains in comparison with members of *Xanthomonas arboricola* pv. *juglandis* and other related *Xanthomonas* species.

Test	<i>X. euroxanthea</i> CPBF367	<i>X. euroxanthea</i> CPBF424 ^T	<i>X. arboricola</i> pv. <i>juglandis</i> CPBF427	<i>X. arboricola</i> pv. <i>juglandis</i> CPBF1521	<i>X. arboricola</i> pv. <i>juglandis</i> LMG 747 ^T	<i>X. arboricola</i> pv. <i>populi</i> CFBP 3123 ^{PT}	<i>X. floridensis</i> ¹	<i>X. nasturtii</i> ¹	<i>X. campestris</i> pv. <i>campestris</i> ¹	<i>X. axonopodis</i> pv. <i>phaseoli</i> ¹	<i>X. prunicola</i> ²	<i>X. arboricola</i> ³
D-salicin	-	-	-	-	-	+	+	-	-	-	n/a	n/a
Mucic acid	+	+	-	-	-	+	+	-	+	+	n/a	n/a
D-saccharic acid	+	+	-	-	-	+	+	-	+	+	n/a	-
Vancomycin	+ / BL	+ / BL	-	-	- / BL	-	-	-	-	-	n/a	n/a
Sucrose	+	+	+	+	-	+	+	+	+	+	+	n/a
Acetic acid	+	+	+	+	+	+	+	+	+	+	-	+
D-melibiose	+	+	+	+	+	+	+	(+) / +	+	+	-	n/a
Formic acid	+	+	+	+	+	+	(+)	-	-	-	-	-
D-arabitol	-	-	-	-	-	BL	+	+	-	(+)	n/a	n/a
D-glucose-6-phosphate	-	-	-	-	-	+	-	-	-	-	-	n/a
D-maltose	+	+	+	+	+	+	+	+	-	-	+	+
Gentiobiose	+	+	+	+	+	+	+	-	(+)	-	+	+

¹ Vicente *et al.*, 2017 (*X. floridensis*: WHRI 8844, WHRI 8846A, WHRI 8848^T, WHRI 8851; *X. nasturtii*: WHRI 8853^T, WHRI 8930A); *X. campestris* pv. *campestris*: WHRI 1279A); *X. axonopodis* pv. *phaseoli*: WHRI 1925C)

² López *et al.*, 2018 (*X. prunicola*: CFBP 8353^T, CFBP 8354, CFBP 8355)

³ Vauterin *et al.*, 1995 (*X. arboricola* pv. *corylina*: LMG 688, LMG 689^T, LMG 8658, LMG 8660; *X. arboricola* pv. *juglandis*: LMG 747^T, LMG 8047; *X. arboricola* pv. *poinsetticola*: LMG 5403; *X. arboricola* pv. *populi*: LMG 12141^{PT}; *X. arboricola* pv. *pruni*: LMG 852^T, LMG 8680)

+, positive; -, negative; BL, borderline; (+), weak positive; n/a, not analysed.

Multilocus sequence analysis

MLSA based on the concatenated partial sequences of seven genes *fyuA* (684 bp), *gyrB* (735 bp), *rpoD* (586 bp), *atpD* (750 bp), *dnaK* (759 bp), *efp* (339 bp) and *glnA* (675 bp) was performed. Sequences were retrieved from NCBI database from 118 strains, including eight different *X. arboricola* pathovars previously used to describe the diversity of *X. arboricola* (Fischer-Le Saux *et al.*, 2015), and another 15 *Xanthomonas* species. The analysis was carried out using Geneious version 9.1.7 software (Biomatters) and a Maximum Likelihood tree based on the General Time Reversible (GTR+G+I) model in MEGA 7.0 was reconstructed (Figure II.2). The novel strains CPBF 424^T CPBF 367 and CPBF 426 formed a separate cluster from all other walnut-associated strains. Interestingly, although isolated from walnut trees, these strains are more closely related to *X. prunicola* and *X. arboricola* pv. *pruni* strains, which are known phytopathogens of *Prunus* trees.

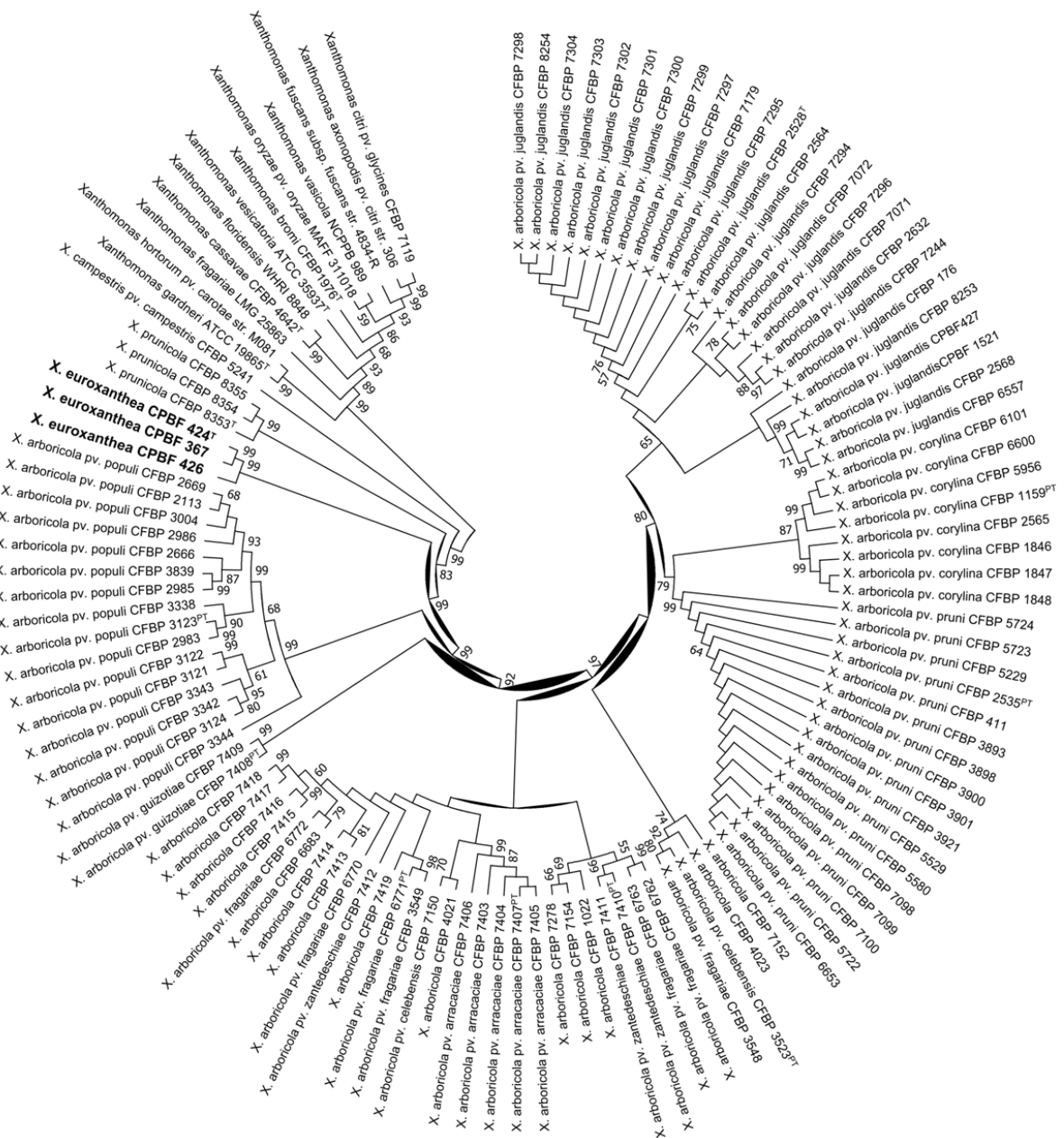


Figure II.2 . Maximum likelihood tree based on the nucleotide alignments of 118 concatenated sequences of *atpD*, *dnaK*, *efp*, *fyuA*, *glnA*, *gyrB* and *rpoD* partial sequences (total length of 4528 bp), emphasizing the phylogenetic relatedness of *Xanthomonas euroxanthea* strains CPBF 367, CPBF 424^T and CPBF 426 (highlighted in bold) within the *Xanthomonas* genus. Bootstrap values >50 from 1000 iterations are shown over and below the branches.

Genomic characterization

The draft genome sequence of CPBF 424^T was previously published (Fernandes *et al.*, 2018a). Whole genome sequencing of strains CPBF 367 and CPBF 426 was carried out as described for CPBF 424^T (Fernandes *et al.*, 2018a). Genomes of strains CPBF 367 and CPBF 424^T had a size of approximately 4.96 and 4.90 Mbp respectively, and comprised 22 and 10 contigs, respectively. High-quality draft genomes of strains CPBF 367 and CPBF 424^T were deposited in GenBank under accession numbers UNRN000000000.1 and UIHB000000000.1, respectively.

ANI and digital DNA-DNA hybridization (dDDH) values between CPBF 367 and CPBF 424^T, and genomes of *X. arboricola* and *Xanthomonas* species were calculated using the EDGAR version 2.0 platform (Blom *et al.*, 2009, 2016) and the formula 2 of Genome-to-Genome Distance Calculator (GGDC 2.1; <http://ggdc.dsmz.de/ggdc.php>) (Meier-Kolthoff *et al.*, 2013). ANI analysis included the genomes of 10 *X. arboricola* strains, belonging to eight distinct pathovars, and 10 other *Xanthomonas* species (Supplementary Table II.S1). ANI values of strains CPBF 367 and CPBF 424^T were less than 93% with *X. arboricola* species and less than 89% with non-arboricola *Xanthomonas* species (Table II.3). Those values reside below the threshold of 95-96% commonly considered for representatives of the same species (Richter & Rosselló-Móra, 2009).

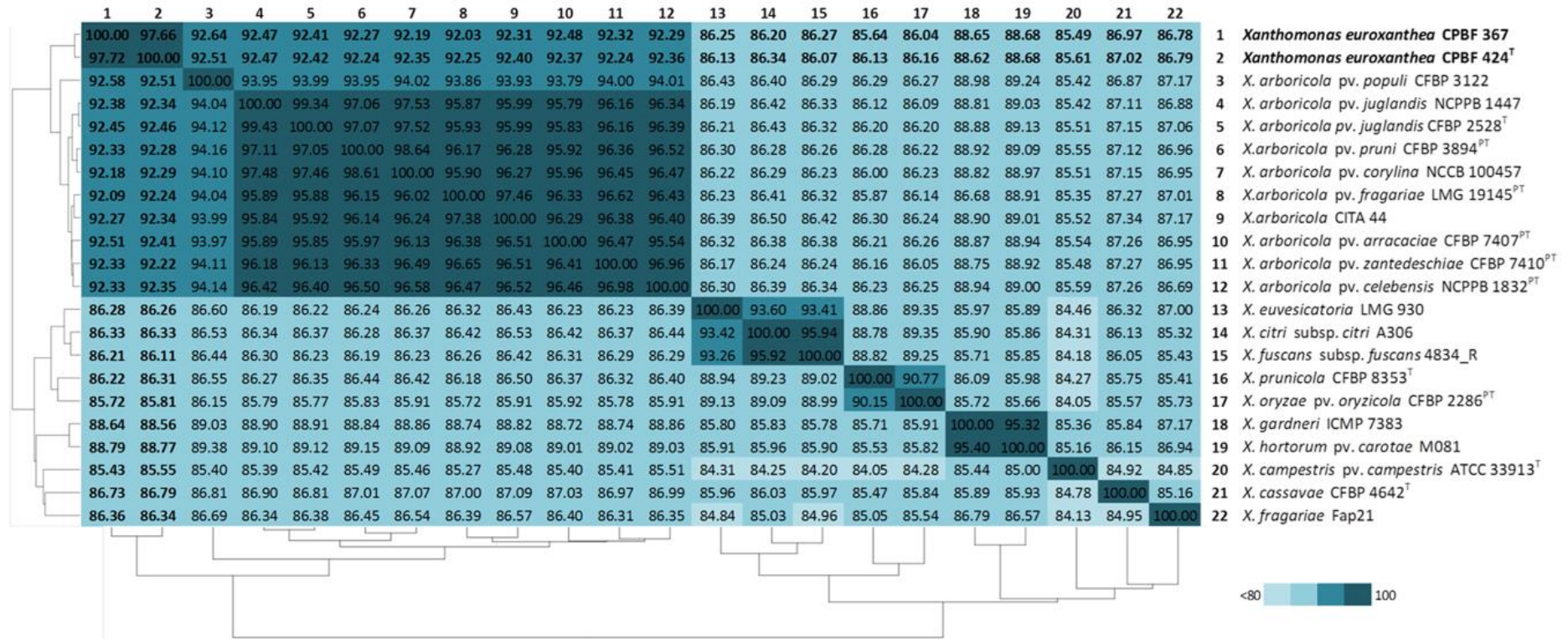


Figure II.3 Average nucleotide identity (ANI) between novel isolates CPBF 367 and CPBF 424^T (in bold) and 20 different *Xanthomonas* genomes using EDGAR. Pairwise ANI values are shown in percentages and indicated within the cells of the heat-map. Accession numbers of sequences are available in Supplementary Table II.S1.

Furthermore, these CPBF 367 and CPBF 424^T strains share 98% ANI with each other (Figure II.3). The dDDH analysis considered 20 strains representing different species of the genus *Xanthomonas* and revealed similarities values below the 70% threshold (Supplementary Table II.S2) considered for delimitation of new species (Meier-Kolthoff *et al.*, 2013). Taken together, ANI and dDDH analysis suggests that these strains belong to a new *Xanthomonas* species.

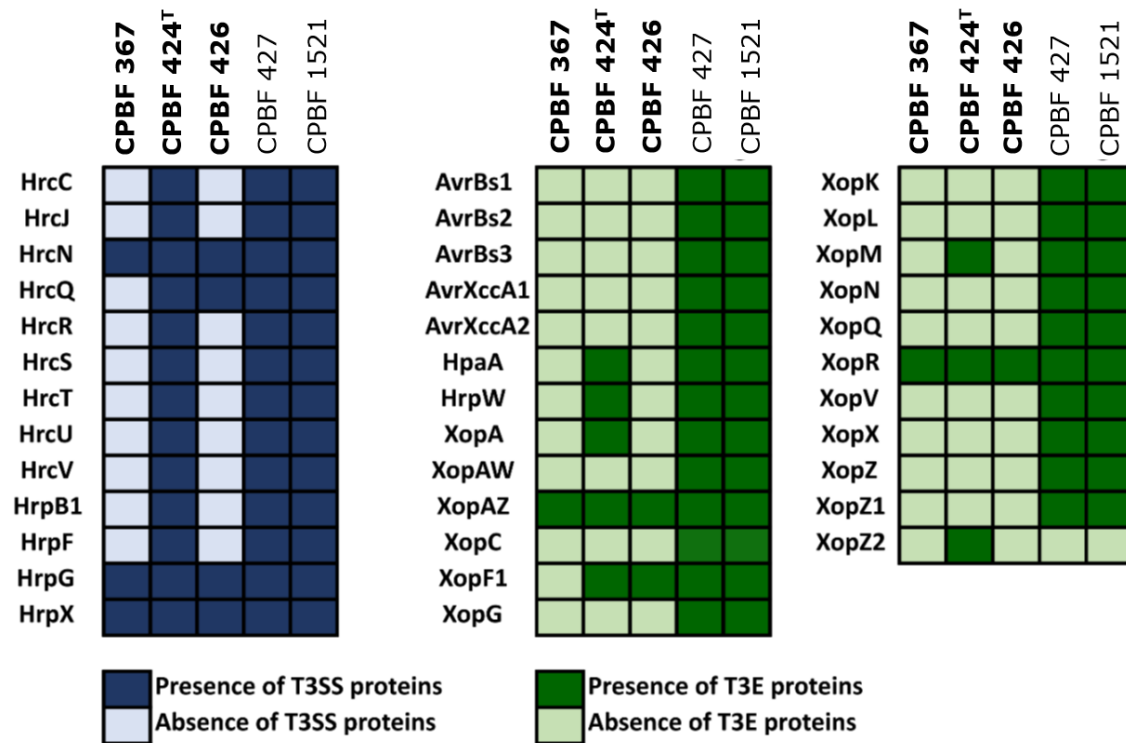


Figure II.4 Distribution of protein homologs of the type III secretion system (T3SS) and type III effectors (T3Es) among the *Xanthomonas euroxanthea* strains (CPBF 367, CPBF 424^T, CPBF 426) and *Xanthomonas arboricola* pv. *juglandis* (CPBF 427 and CPBF 1521) strains based on tBLASTn analysis. Dark blue corresponds to the presence of T3SS protein homologous, and dark green corresponds to the presence of T3E homologous. The light blue and green indicate absence of T3SS and T3E homologous, respectively. The threshold criteria applied were minimal e-value of 1e-10, query length similarity cut-off $\geq 40\%$, and sequence identity cut-off $\geq 70\%$.

Type III Secretion System (T3SS) and Type III Effectors (T3E)

The prediction of type III secretion system (T3SS) and type III effectors (T3E) homologs was carried out with tBLASTn analysis against previously characterized T3SS and T3Es of the *Xanthomonas* group (White *et al.*, 2009; Hajri *et al.*, 2011; Ryan *et al.*, 2011). The criteria applied as threshold was a minimal e-value of $1e^{-10}$, query length similarity cut-off $\geq 40\%$, and sequence identity with a cut-off $\geq 70\%$ (Figure II.4). The walnut pathogenic strain CPBF 424^T has a T3SS gene profile similar to the strains *X. arboricola* pv. *juglandis* CPBF 427 and CPBF 1521 and possess gene homologs described for pathogenic strains of *X. arboricola* (Cesbron *et al.*, 2015; Garita-Cambronero *et al.*, 2016, 2017, 2018). In contrast, strains CPBF 367 and CPBF 426 lack the majority of the T3SS and T3E encoding genes (Figure II.4), resembling the T3SS and T3E profile of atypical non-pathogenic strains of *Xanthomonas* species described previously (Cesbron *et al.*, 2015; Garita-Cambronero *et al.*, 2017). The absence of some of the components from the Hrp2 family and of genes that are part of the macromolecular structure of the T3SS may explain the non-pathogenic phenotype of CPBF 367, even while harbouring regulators genes of T3SS, as *hrpX* and *hrpG* (Guo *et al.*, 2011; Jacobs *et al.*, 2015). Additionally, the fact that strain CPBF 426 is deficient for most T3SS genes, suggests that this strain may not be pathogenic as shown for CPBF 367. Intriguingly, CPBF 424^T seems to have a functional T3SS and holds a narrow gene repertoire of T3E homologs in comparison with other *X. arboricola* pv. *juglandis* (Hajri *et al.*, 2011; Essakhi *et al.*, 2015; Cesbron *et al.*, 2015). This differs markedly from *X. arboricola* pv. *juglandis* CPBF 427 and CPBF 1521 and strains CPBF 367 and CPBF 426. Although the pathogenic strain CPBF 424^T possesses homologs for eight known effectors, from which only three were identified in non-pathogenic *X. euroxanthea* (CPBF 367 and CPBF 426), no homologs were found for most of the genes assigned as important for pathogenicity of *Xanthomonas* species, namely *avrbs2*, *avrbs2*, *avrbs3*, *avrXccA1*, *avrXccA2*, *xopAW*, *xopC*, *xopG*, *xopK*, *xopL*, *xopN*, *xopQ*, *xopV*, *xopX*, *xopZ* and *xopZ1*. All the aforementioned genes, with the exception of *xopZ2*, were present in both *X. arboricola* pv. *juglandis* CPBF 427 and CPBF 1521 strains (Figure II.4). The genomic features of this new species and the pathogen phenotype of strain CPBF 424^T in walnut make these strains particularly appealing to elucidate the evolutionary hypothesis of pathogenicity in walnut, to uncover the genetic footprints of adaptation and to address speciation in *Xanthomonas*.

Description of *Xanthomonas euroxanthea* sp. nov.

Xanthomonas euroxanthea (eu.ro.xan.the'a. N.L. fem. adj. *euroxanthea* referring to the EuroXanth COST Action CA 16107, the EU funded network in which the isolates of this species were characterized).

Cells are Gram-stain-negative straight rods and form colonies that are yellow, circular, smooth, mucoid and slightly convex with entire margins when grown on YDC for 2 days. Regarding carbon source metabolism *X. euroxanthea* was positive for 34 substrates (dextrin, maltose, trehalose, cellobiose, gentiobiose, sucrose, meliobiose, *N*-acetyl-D-glucosamine, α -D-glucose, D-mannose, D-fructose, D-galactose, L-fucose, glycerol, gelatin, glycyl-L-proline, L-alanine, L-glutamic acid, L-serine, mucic acid, D-saccharic acid, methyl pyruvate, L-lactic acid, citric acid, α -keto-glutaric acid, L-malic acid, bromo-succinic acid, Tween 40, α -hydroxy-butyric acid, α -keto-butyric acid, acetoacetic acid, propionic acid, acetic acid, formic acid) and negative for 25 compounds (turanose, stachyose, raffinose, β -methyl-D-glucoside, D-salicin, *N*-acetyl- β -D-mannosamine, *N*-acetyl-D-galactosamine, *N*-acetyl-neuraminic acid, 3-methyl glucose, L-rhamnose, D-sorbitol, D-arabitol, myo-inositol, D-glucose-6-phosphate, D-aspartic acid, D-serine, L-arginine, L-pyroglutamic acid, D-gluconic acid, quinic acid, p-hydroxy-phenylacetic acid, D-lactic acid methyl ester, D-malic acid, γ -amino-butyric acid, β -hydroxy-D,L-butyric acid).

This species includes non-pathogenic (CPBF 367) and pathogenic (CPBF 424^T) strains. The type strain is CPBF 424^T (LMG 31037^T = CCOS 1891^T = NCPPB 4675^T), isolated from an asymptomatic bud of an isolated ornamental diseased walnut tree (*J. regia*) growing in a municipal garden of Loures, Portugal. The high-quality draft genome of the type strain is characterized by a 4.8 Mbp and a G+C content of 65.9 mol%. The GenBank accession number of the 16S rRNA gene sequence of strain CPBF 424^T is MT036365 and its draft genome sequence accession number is GCA_900476395.

Funding information

This research was co-financed by the European Structural & Investment Funds (ESIFs) through COMPETE 2020, and by National Funds through FCT- Fundação para a Ciência e Tecnologia, within the framework of the project EVOXANT (PTDC/BIA-EVF/3635/2014-POCI-01-0145-FEDER-016600). Camila Fernandes and Leonor Martins were supported by fellowships from FCT (SFRH/BD/95913/2013; SFRH/BD/137079/2018, respectively). Nay C. Dia was supported by grant no IZCOZ0_177064 from the Swiss National Foundation (SNF) for Scientific Research. The EDGAR platform is financially supported by the BMBF grant FKZ031A533 within the de.NBI network. This article is based upon work from COST Action CA16107 EuroXanth, supported by COST (European Cooperation in Science and Technology).

Acknowledgments

We thank Prof. Dr. Jean-Louis Charlet, Prof. Dr. Bernhard Schink and Prof. Aharon Oren for advice regarding the nomenclature, and Christian Kunkel from the Culture Collection of Switzerland (CCOS AG, Wädenswil, Switzerland) for his support during the phenotypic characterization.

References

- Belisario A, Santori A, Potente G *et al.*, 2010. Brown apical necrosis (BAN): a fungal disease causing fruit drop of english walnut. In: McNeil, ed. *Vi International Walnut Symposium (Acta Horticulturae. Leuven 1: International Society of Horticultural Science*. 449–452.
- Blom J, Albaum SP, Doppmeier D *et al.*, 2009. EDGAR: A software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics* **10**.
- Blom J, Kreis J, Spänig S *et al.*, 2016. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic acids research* **44**, W22–W28.
- Cesbron S, Briand M, Essakhi S *et al.*, 2015. Comparative Genomics of Pathogenic and Nonpathogenic Strains of *Xanthomonas arboricola* Unveil Molecular and Evolutionary Events Linked to Pathoadaptation. *Frontiers in Plant Science* **6**, 1126.
- Constantin EC, Cleenwerck I, Maes M *et al.*, 2016. Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. *Plant Pathology* **65**, 792–806.
- Essakhi S, Cesbron S, Fischer-Le Saux M, Bonneau S, Jacques M-A, Manceau C, 2015. Phylogenetic and Variable-Number Tandem-Repeat Analyses Identify Nonpathogenic *Xanthomonas arboricola* Lineages Lacking the Canonical Type III Secretion System. *Applied and Environmental Microbiology* **81**, 5395–5410.
- Fernandes C, Albuquerque P, Cruz L, Tavares F Genotyping and epidemiological metadata provides new insights into population structure of *Xanthomonas* isolated from walnut trees. *bioRxiv*: 397703. <https://doi.org/10.1101/397703>.
- Fernandes C, Albuquerque P, Sousa R, Cruz L, Tavares F, 2017. Multiple DNA Markers for Identification of *Xanthomonas arboricola* pv. *juglandis* Isolates and its Direct Detection in Plant Samples. *Plant Disease* **101**, 858–865.
- Fernandes C, Blom J, Pothier JF, Tavares F, 2018a. High-Quality Draft Genome Sequence of *Xanthomonas* sp. Strain CPBF 424, a Walnut-Pathogenic Strain with Atypical Features. *Microbiology Resource Announcements* **7**.
- Fernandes C, Sousa R, Tavares F, Cruz L, 2018b. First Report of *Xanthomonas arboricola* Causing Bacterial Blight on Pecan Trees in Portugal. *Plant Disease* **102**, 2632.
- Fischer-Le Saux M, Bonneau S, Essakhi S, Manceau C, Jacques M-A, 2015. Aggressive

- Emerging Pathovars of *Xanthomonas arboricola* Represent Widespread Epidemic Clones Distinct from Poorly Pathogenic Strains, as Revealed by Multilocus Sequence Typing. *Applied and Environmental Microbiology* **81**, 4651–4668.
- da Gama MAS, Mariano R de LR, da Silva Júnior WJ *et al.*, 2018. Taxonomic Repositioning of *Xanthomonas campestris* pv. *viticola* (Nayudu 1972) Dye 1978 as *Xanthomonas citri* pv. *viticola* (Nayudu 1972) Dye 1978 comb. nov. and Emendation of the Description of *Xanthomonas citri* pv. *anacardii* to Include Pigmented Isolates Pa. *Phytopathology*® **108**, 1143–1153.
- Garita-Cambronero J, Palacio-Bielsa A, Cubero J, 2018. *Xanthomonas arboricola* pv. *pruni*, causal agent of bacterial spot of stone fruits and almond: its genomic and phenotypic characteristics in the *X. arboricola* species context. *Molecular Plant Pathology* **19**, 2053–2065.
- Garita-Cambronero J, Palacio-Bielsa A, López MM, Cubero J, 2016. Comparative genomic and phenotypic characterization of pathogenic and non-pathogenic strains of *Xanthomonas arboricola* reveals insights into the infection process of bacterial spot disease of stone fruits. *PLoS ONE* **11**, e0161977.
- Garita-Cambronero J, Palacio-Bielsa A, López MM, Cubero J, 2017. Pan-genomic analysis permits differentiation of virulent and non-virulent strains of *Xanthomonas arboricola* that cohabit *Prunus* spp. and elucidate bacterial virulence factors. *Frontiers in Microbiology* **8**, 1–17.
- Guo Y, Figueiredo F, Jones J, Wang N, 2011. HrpG and HrpX Play Global Roles in Coordinating Different Virulence Traits of *Xanthomonas axonopodis* pv. *citri*. *Molecular Plant-Microbe Interactions* **24**, 649–661.
- Hajri A, Meyer D, Delort F, Guillaumès J, Brin C, Manceau C, 2010. Identification of a genetic lineage within *Xanthomonas arboricola* pv. *juglandis* as the causal agent of vertical oozing canker of Persian (English) walnut in France. *Plant Pathology* **59**, 1014–1022.
- Hajri A, Pothier JF, Saux MF Le *et al.*, 2011. Type three effector gene distribution and sequence analysis provide new insights into the pathogenicity of plant-pathogenic *Xanthomonas arboricola*. *Applied and Environmental Microbiology* **78**, 371–384.
- International Seed Testing Association, 1999. International rules for seed testing rules 1999. *Seed Science & Technology* **27** **supplem.**
- Jacobs JM, Pesce C, Lefeuvre P, Koebnik R, 2015. Comparative genomics of a cannabis

pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*. *Frontiers in Plant Science* **6**.

Jacques M-A, Arlat M, Boulanger A *et al.*, 2016. Using Ecology, Physiology, and Genomics to Understand Host Specificity in *Xanthomonas*. *Annual Review of Phytopathology* **54**, 163–187.

Lamichhane JR, 2014. *Xanthomonas arboricola* diseases of stone fruit, almond, and walnut trees: Progress toward understanding and management. *Plant Disease* **98**, 1600–1610.

Leslie C, Uratsu S, McGranahan G, Dandekar A, 2006. *Agrobacterium Protocols Volume 2*. Springer.

López MM, Lopez-Soriano P, Garita-Cambronero J *et al.*, 2018. *Xanthomonas prunicola* sp. nov., a novel pathogen that affects nectarine (*Prunus persica* var. *nectarina*) trees. *International journal of systematic and evolutionary microbiology* **68**, 1857–1866.

Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M, 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**, 60.

Merda D, Bonneau S, Guimbaud JF *et al.*, 2016. Recombination-prone bacterial strains form a reservoir from which epidemic clones emerge in agroecosystems. *Environmental Microbiology Reports* **8**, 572–581.

Moragrega C, Matias J, Aletà N, Montesinos E, Rovira M, 2011. Apical necrosis and premature drop of Persian (English) walnut fruit caused by *Xanthomonas arboricola* pv. *juglandis*. *Plant Disease* **95**, 1565–1570.

Richter M, Rosselló-Móra R, 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences USA* **106**, 19126–19131.

Ryan RP, Vorhölter FJ, Potnis N *et al.*, 2011. Pathogenomics of *Xanthomonas*: Understanding bacterium-plant interactions. *Nature Reviews Microbiology* **9**, 344–355.

Schulze-Lefert P, Panstruga R, 2011. A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. *Trends in Plant Science* **16**, 117–125.

- Stolp' H, Starr MP, 1964. Bacteriophage Reactions and Speciation of Phytopathogenic *Xanthomonads*. *Journal of Phytopathology* **51**, 442–478.
- Triplett LR, Verdier V, Campillo T et al., 2015. Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali and proposal of *Xanthomonas maliensis* sp. nov. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* **107**, 869–881.
- Vauterin L, Hoste B, Kersters K, Swings J, 1995. Reclassification of *Xanthomonas*. *International Journal of Systematic Bacteriology* **45**, 472–489.
- Vauterin L, Rademaker J, Swings J, 2000. Synopsis on the taxonomy of the genus *Xanthomonas*. *Phytopathology* **90**, 677–682.
- Vicente JG, Rothwell S, Holub EB, Studholme DJ, 2017. Pathogenic, phenotypic and molecular characterisation of *Xanthomonas nasturtii* sp. nov. and *Xanthomonas floridensis* sp. nov., new species of *Xanthomonas* associated with watercress production in Florida. *International Journal of Systematic and Evolutionary Microbiology* **67**, 3645–3654.
- White FF, Potnis N, Jones JB, Koebnik R, 2009. The type III effectors of *Xanthomonas*. *Molecular Plant Pathology* **10**, 749–766.
- Yeates GW, 1978. A proposed nomenclature and classification for plant pathogenic bacteria. *New Zealand Journal of Agricultural Research* **21**, 153–177.

Supplementary material

One supplementary figure and two supplementary tables are available with the online version of this article:

https://www.microbiologyresearch.org/docserver/fulltext/ijsem/70/12/004386_1.pdf?expires=1674001714&id=id&accname=guest&checksum=22E19CD00D9540BB476FDC1AC8F498E5. Table S1. List of genomes used in ANI – EDGAR. Table II.S2. Percentage of digital

DNA-DNA hybridization values. Figure II. S1. Ornamental walnut tree host from where isolates were recovered from. Detail of an infected leaf with a close -up of. bacterial blight symptoms.



Chapter III

Comparative Genomics of
Xanthomonas euroxanthea and
Xanthomonas arboricola pv.
juglandis Strains Isolated from a
Single Walnut Host Tree

Chapter III include the following publication:

Fernandes C*, Martins L*, Teixeira M*, Blom J, Pothier JF, Fonseca NA, Tavares F (2021) Comparative Genomics of *Xanthomonas euroxanthea* and *Xanthomonas arboricola* pv. *juglandis* Strains Isolated from a Single Walnut Host Tree. *Microorganisms* 2021, Vol 9, Page 624

* These authors contributed equally to this work.

Part of the results from this chapter were presented at:

Fernandes C, Martins L, Teixeira M, Blom J, Pothier JF, Fonseca NA, Tavares F. Comparative genomics of walnut associated *Xanthomonas euroxanthea* and *Xanthomonas arboricola* pv. *juglandis* strains isolated from a single host tree. II Plant Pests and Diseases Forum: Redefining Concepts, Mechanisms and Management Practices - Faculdade de Ciências da Universidade do Porto, Portugal, 24th March 2021 (Oral presentation)

Fernandes C, Martins L, Blom J, Pothier JF, Tavares F. Comparative genomics highlight putative pathoadaptations of a new *Xanthomonas* sp. isolated from walnut and *Xanthomonas arboricola* pv. *juglandis*. 3rd Annual EuroXanth Conference, Faculty of Horticulture, Lednice, Czech republic, 9-11 September 2019 (Poster and oral presentation)

Fernandes C, Martins L, Blom J, Pothier JF, Tavares F. Insights into sympatric lifestyle of walnut *Xanthomonas* isolates through comparative genomics. Trends in Biodiversity and Evolution (TIBE), CIBIO-InBio, Vairão, Portugal, 5-7 December 2018 (Oral presentation)

Comparative Genomics of *Xanthomonas euroxanthea* and *Xanthomonas arboricola* pv. *juglandis* Strains Isolated from a Single Walnut Host Tree

Abstract

The recent report of distinct *Xanthomonas* lineages of *Xanthomonas arboricola* pv. *juglandis* and *Xanthomonas euroxanthea* within the same walnut tree revealed that this consortium of walnut-associated *Xanthomonas* includes both pathogenic and nonpathogenic strains. As the implications of this co-colonization are still poorly understood, in order to unveil niche-specific adaptations, the genomes of three *X. euroxanthea* strains (CPBF 367, CPBF 424^T, and CPBF 426) and of an *X. arboricola* pv. *juglandis* strain (CPBF 427) isolated from a single walnut tree in Loures (Portugal) were sequenced with two different technologies, Illumina and Nanopore, to provide consistent single scaffold chromosomal sequences. General genomic features showed that CPBF 427 has a genome similar to other *X. arboricola* pv. *juglandis* strains, regarding its size, number, and content of CDSs, while *X. euroxanthea* strains show a reduction regarding these features comparatively to *X. arboricola* pv. *juglandis* strains. Whole genome comparisons revealed remarkable genomic differences between *X. arboricola* pv. *juglandis* and *X. euroxanthea* strains, which translates into different pathogenicity and virulence features, namely regarding type 3 secretion system and its effectors and other secretory systems, chemotaxis-related proteins, and extracellular enzymes. Altogether, the distinct genomic repertoire of *X. euroxanthea* may be particularly useful to address pathogenicity emergence and evolution in walnut-associated *Xanthomonas*.

1. Introduction

Xanthomonas is a genus of gammaproteobacteria (Dowson, 1939), which include numerous species acknowledged as important plant-associated bacteria with the capacity to cause disease in a wide range of plant species, including important agricultural crops (Leyns *et al.*, 1984; Swings & Civerolo, 1993). At the current date, the genus comprises 32 species (with validly published and correct names) (Parte *et al.*, 2020), some of which are subdivided into distinct pathogenicity groups, also known as pathovars, according to their high degree of host specificity, disease symptoms, and infection mechanisms (Young *et al.*, 1978; Vauterin *et al.*, 2000; Meyer & Bogdanove,

2009; Ryan *et al.*, 2011). Within this genus, *Xanthomonas arboricola* (Vauterin *et al.*, 1995) is responsible for severe diseases in important stone fruits and nut trees (Lamichhane, 2014). Pathogenic strains of *X. arboricola* belonging to the pathovar *juglandis* (Pierce NB, 1901) cause serious diseases on walnut, namely walnut bacterial blight (WBB), brown apical necrosis (BAN), and vertical oozing canker (VOC), altogether responsible for major drops in walnut production and extended damages in walnut nurseries, leading to major economic losses (Smith *et al.*, 1912; Belisario *et al.*, 2002; Hajri *et al.*, 2010).

Compared to other *X. arboricola* pathovars, *X. arboricola* pv. *juglandis* is characterized by a larger genetic diversity recorded throughout geographically distinct walnut cultivation regions (Ivanović *et al.*, 2014; Kaluzna *et al.*, 2014; Fischer-Le Saux *et al.*, 2015; Giovanardi *et al.*, 2016; Fernandes *et al.*, 2021). Genotyping and population studies of walnut-associated xanthomonads identified a high diversity of *X. arboricola* pv. *juglandis* pathogenic strains responsible for walnut bacterial diseases (Hajri *et al.*, 2010) and nonpathogenic strains of *X. arboricola* shown to be asymptomatic on walnut (Essakhi *et al.*, 2015). Based on a large body of evidence, carefully reviewed by Büttner and Bonas (2010) (Büttner & Bonas, 2010) and more recently by An *et al.* (2020) (An *et al.*, 2019), it is currently acknowledged that *Xanthomonas* pathogenesis is dependent on an array of pathogenicity and virulence factors, including type 3 secretion system (T3SS) and type 3 effectors (T3E), but also other type secretion systems, adhesins, extracellular polysaccharides (EPS), lipopolysaccharides (LPS), and plant cell wall hydrolytic enzymes, among others. Furthermore, it has been suggested that these genetic determinants of pathogenicity and virulence are under the control of tight transcriptional and post-transcriptional regulatory mechanisms that allow infective bacteria to adhere, invade the plant host tissues, and to multiply and overcome the plant defense mechanisms (Büttner & Bonas, 2010).

Comparative genomics of pathogenic and nonpathogenic strains of *X. arboricola* has been instrumental to unveil genetic determinants of pathogenicity and virulence (Hajri *et al.*, 2011; Cesbron *et al.*, 2015; Garita-Cambronero *et al.*, 2018) and to provide insights into evolutionary events linked to pathoadaptation, as emphasized by Cesbron *et al.* (2015) (Cesbron *et al.*, 2015). Still, in order to gain more insight regarding adaptations of nonpathogenic strains to new environments, comparative genomics of pathogenic and nonpathogenic strains isolated from the same host is required (An *et al.*, 2019). Although T3SS and T3E have been pointed out as major players of pathogenicity in *X. arboricola*, the different repertoires of virulence-associated genes between pathogenic and nonpathogenic strains of *X. arboricola* suggests that pathogenicity and virulence are

determined by a complex network of genetic determinants, which is not fully understood (Cesbron *et al.*, 2015).

Most of the *X. arboricola* whole genomes currently available are scattered in multiple contigs assembled from millions of short-reads (typically < 300 bp) obtained by sequencing-by-synthesis technologies (such as Illumina) incapable of resolving repetitive domains and impairing accurate comparisons across wide genomic domains (Kchouk *et al.*, 2017). More recently, single-molecule-sequencing technologies (such as Oxford Nanopore Technologies (ONT)) have emerged. These, are capable of producing long reads (up to several thousand bp) that may span over entire repetitive regions, achieving a complete structural resolution of genetic elements, regardless the higher error rates (above 10% of mismatches and indels, compared to ~1% of mismatches for short reads) that blur the sequence resolution (Lu *et al.*, 2016; Kchouk *et al.*, 2017). Hybrid assembling approaches, capable to conciliate the accuracy of Illumina short-reads and the gap-free long-read sequences obtained by ONT, result in single chromosome scaffolds with high confidence per base call, and particularly suitable for high accuracy comparative genomics to unveil genetic and structural variants (Oxford Nanopore Technologies; Goldstein *et al.*, 2019).

The recent description of pathogenic and nonpathogenic strains isolated from walnut belonging to the new species *Xanthomonas euroxanthea* (Martins *et al.*, 2020) raises further questions regarding new pathoadaptations to this host (Fernandes *et al.*, 2018a, 2021; Teixeira *et al.*, 2020). Furthermore, the fact that these *X. euroxanthea* strains were isolated from the same walnut host tree, together with characteristic *X. arboricola* pv. *juglandis* strains (CPBF 427, CPBF 1521) (Fernandes *et al.*, 2018b, 2021; Martins *et al.*, 2020; Teixeira *et al.*, 2021), suggest a sympatric lifestyle that may contribute to unveil events of genetic recombination and trade-offs related to pathogenicity and virulence of *Xanthomonas* in walnut. The present work discloses a comprehensive comparative genomics study between the complete genome sequences of three *X. euroxanthea* strains (CPBF 367, CPBF 424^T, CPBF 426) and one *X. arboricola* pv. *juglandis* strain (CPBF 427) isolated from the same diseased walnut tree. To better understand the ecology and evolution of these two walnut-associated *Xanthomonas* species, emphasis has been given to the genetic determinants of pathogenicity and virulence and putative niche-specific adaptations.

2. Materials and Methods

2.1. *Xanthomonas* Strains Used in This Study and Genome Sequencing

our isolates (CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427) obtained from a single walnut tree host located in the region of Loures, Portugal (Fernandes *et al.*, 2021), with distinct genotypes and pathogenicity in walnut (while CPBF 367 was nonpathogenic, CPBF 424^T and CPBF 427 were pathogenic) (Martins *et al.*, 2020), were chosen for genome sequencing. A previous MLSA of *acnB*, *fyuA*, *gyrB*, and *rpoD* genes grouped the strains CPBF 427 along with other *X. arboricola* pv. *juglandis* strains (Fernandes *et al.*, 2021), whereas CPBF 367, CPBF 424^T, and CPBF 426 were recently described as belonging to the new species *X. euroxanthea* (Martins *et al.*, 2020).

Bacteria culture, DNA extraction for sequencing, library preparation, and genome sequencing and annotation were carried out as previously described by Teixeira *et al.* (Teixeira *et al.*, 2020, 2021). Briefly, bacteria were recovered on culture medium M2 at 28 °C and 100 rpm for 48 h. DNA was extracted using an E.Z.N.A. bacterial DNA purification kit (Omega Bio-tek, Norcross, GA, USA) and sequenced with Illumina and Oxford Nanopore Technologies (ONT) MinION platforms. Illumina sequencing was outsourced to GATC Biotech, AG (Konstanz, Germany), using an Illumina HiSeq instrument with a standard 2 × 150 bp paired-end library protocol, while ONT sequencing was performed on a MinION sequencer using an R9.4.1 flow cell. Reads were base called and demultiplexed using Guppy v3.4.1 (high accuracy base-calling mode) and assembled de novo following a hybrid Nanopore-Illumina approach using Unicycler v. 0.4.8 (Wick *et al.*, 2017) and annotated with PGAP v2020-03-30.build4489 (Tatusova *et al.*, 2016). The complete genome sequences of *Xanthomonas* strains CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427 have been deposited in the European Nucleotide Archive (ENA). Illumina reads, ONT reads, and the assembled genomes are respectively available under the following accession numbers: ERX2780809, ERX4296808, and GCA_903989455 for CPBF 367; ERR2767968, ERX4911540, and GCA_905187425 for CPBF 424^T; ERX2780811, ERX4296809, and GCA_903989465 for CPBF 426; and ERX2780812, ERX4296810, and GCA_903989475 for CPBF 427.

2.2. Average Nucleotide Identity

The average nucleotide identity (ANI), based on BLASTn, was carried out with OrthoANI v1.40 (Lee *et al.*, 2016) to determine the genetic distance between each of the four bacterial genomes (CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427) sequenced within the framework of this study and 40 genomes of *Xanthomonas* spp., including 29 genomes of *X. arboricola* and eight distinct pathovars (Supplementary Table III.S1), all available at the NCBI genome database.

2.3. Comparative Genome Analysis

A total of 1149 genes from BUSCO (v. 4.0.6, database xanthomonadales_odb10) (Simão *et al.*, 2015; Waterhouse *et al.*, 2018), present in all the 44 mentioned *Xanthomonas* genomes, were aligned with TCoffee (v11.00.8cbe486, mcoffee mode) (Notredame *et al.*, 2000) to generate a tree, using RAxML (raxmlGUI 2.0, model GTR+FO) (Stamatakis, 2006; Edler *et al.*, 2021) with 500 bootstrap replicates. The corresponding phylogram was visually represented with the R package “ggtree” (Yu *et al.*, 2017). In order to unveil strain-specific features and pools of genes shared between the CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427, a pairwise comparison of genomes based on SRVs (Lerat *et al.*, 2003; Blom *et al.*, 2009) was performed in the EDGAR platform (EDGAR v. 2.3.(Blom *et al.*, 2009, 2016) and represented in a Venn diagram.

2.4. Homologous of Pathogenicity and Virulence-Associated Proteins Inferred by tBLASTn Analysis

The genome sequences of the four *Xanthomonas* strains used in this study were scrutinized for the presence of protein homologs by tBLASTn analysis (tBLASTn v. 2.10.1 (Altschul *et al.*, 1990)) against a created database of protein sequences previously reported to be involved in the pathogenesis and virulence of *Xanthomonas* (Supplementary Table III.S3). To ensure that only closely-related orthologous proteins were selected, the tBLASTn cut-offs criteria to identify protein homologs were $\geq 40\%$ identity and $\geq 75\%$ query sequence length. The search for homologs was performed using as query sequences proteins of xanthan biosynthesis identified by Lee *et al.* (Lee *et al.*, 2005) and Vorhölter *et al.* (Vorhölter *et al.*, 2008), a list of proteins of the flagellar system from *X. campestris* pv. *vesicatoria* 85-10 (Darrasse *et al.*, 2013) and from *X. fragariae* LMG 25863 (AJRZ000000000.1, NCBI database), as well as proteins of the *rpf*

gene cluster for regulation of pathogenicity factors in *X. campestris*, (Crossman & Dow, 2004) and *X. fragariae* LMG 25863, AJRZ000000000.1, NCBI database). Homologs of chemotaxis and methyl-accepting chemotaxis proteins, proteins involved in the biosynthesis of quorum sensing signals and non-fimbrial adhesins, were identified using as query the protein sequences previously used for *X. arboricola* genomes by Garita-Cambronero et al. (Garita-Cambronero et al., 2018). The presence or absence of homologs associated with components of the different secretion systems were also predicted, using as query sequences proteins of the type II secretion system (T2SS) (Filloux, 2004; Szczesny et al., 2010) and related hemicellulolytic, cellulolytic, and pectolytic enzymes, lipases, and proteases (Garita-Cambronero et al., 2018); proteins of the type IV secretion system and type IV pilus (Dunger et al., 2016; Garita-Cambronero et al., 2018); and proteins of the type VI secretion system (Shrivastava & Mande, 2008).

3. Results

3.1. General Features of *X. euroxanthea* and *X. arboricola* pv. *juglandis* Genome Assemblies

The complete genome sequences obtained by hybrid assemblies of Illumina and Nanopore reads for the four studied strains (CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427), allowed a single chromosomal scaffold to be achieved for all four strains and revealed the presence of one plasmid in strains CPBF 367 and CPBF 426 (Table III.1). When comparing the genome properties, while some features are similar between the four genomes, namely a G + C content around 65%, a number of coding genes proportional to genome size, and the number of rRNA operons and 5S tRNA genes, some other features are clearly distinct between the three *X. euroxanthea* strains (CPBF 367, CPBF 424^T, and CPBF 426) and the *X. arboricola* pv. *juglandis* strain CPBF 427. These differences are particularly underlined by a higher genome size for the *X. arboricola* pv. *juglandis* strain CPBF 427 (5.23 Mb), in comparison with the slightly smaller genomes of *X. euroxanthea* strains CPBF 367, CPBF 424^T, and CPBF 426, i.e., less than 5.0 Mb, the higher number of RNA genes, non-coding RNAs (ncRNA), and pseudogenes observed for strain CPBF 427 comparatively with the three *X. euroxanthea* strains (Table III.1).

Table III.1. General genomic features of the four strains analyzed in this study.

General Features	<i>Xanthomonas euroxanthea</i> strains			<i>Xanthomonas arboricola</i> pv. <i>juglandis</i> Strain
	CPBF 367	CPBF 424 ^T	CPBF 426	CPBF 427
Genome size (bp)	4,968,459	4,900,930	4,900,648	5,228,174
Contigs	2	1	2	1
N ₅₀ (bp)	4,923,218	4,900,930	4,883,254	5,228,174
G+C content (%)	65.81	65.88	65.85	65.38
Plasmids	1	0	1	0
Total genes	4157	4119	4143	4465
Total CDSs	4077	4040	4066	4367
Coding genes	4012	3993	4003	4237
RNA genes	80	79	77	98
rRNA (5S, 16S, 23S)	2, 2, 2	2, 2, 2	2, 2, 2	2, 2, 2
ncRNA	18	17	18	38
tRNA	56	56	53	54
Pseudogenes	65	47	63	130
ENA/GenBank accession number	GCA_903989455	CGA_905187425	GCA_903989465	GCA_903989475
Reference	(Teixeira <i>et al.</i> , 2020)	This study	(Teixeira <i>et al.</i> , 2020)	(Teixeira <i>et al.</i> , 2021)

3.2. Genomic Distance Assessed by ANI and Phylogenetic Analysis

The ANI values determined for a sampling of 44 *Xanthomonas* strains, including the four walnut-associated *Xanthomonas* isolates characterized in the current study, assigned strains CPBF 367, CPBF 424^T, and CPBF 426 to the recently described new species *X. euroxanthea* (Martins *et al.*, 2020), and strain CPBF 427 as a member of *X. arboricola* pv. *juglandis*.

In fact, the high ANI similarity values of $\geq 97.9\%$ shared between strains CPBF 367, CPBF 424^T, and CPBF 426 and the ANI values of ≤ 93.6 and $\leq 89.4\%$ observed between these three strains and *X. arboricola* and CPBF 427 or other *Xanthomonas* species strains, respectively (Figure III.1, Supplementary Table III.S2), are clearly above and below the high stringent threshold of $> 95\%$ to separate different species (Richter & Rosselló-Móra, 2009). Interestingly, two nonpathogenic *X. arboricola* strains (CFBP 7635 and CFBP 7653), previously described (Essakhi *et al.*, 2015), share ANI values of $\geq 97.8\%$ with the three *X. euroxanthea* strains (CPBF 367, CPBF 424^T, and CPBF 426), strongly suggesting that these two strains are misclassified and likely belong to *X. euroxanthea*. The phylogenetic tree obtained for the *Xanthomonas* 44 strains using the 1149 single-copy orthologous genes from BUSCO placed *X. euroxanthea* strains (CPBF 367, CPBF 424^T, and CPBF 426) and CFBP 7635 and CFBP 7653 in a cluster well separated from all the other xanthomonads considered in the analysis, including the CPBF 427 and other *X. arboricola* pv. *juglandis* strains with which they share the same plant host (Figure III.2).

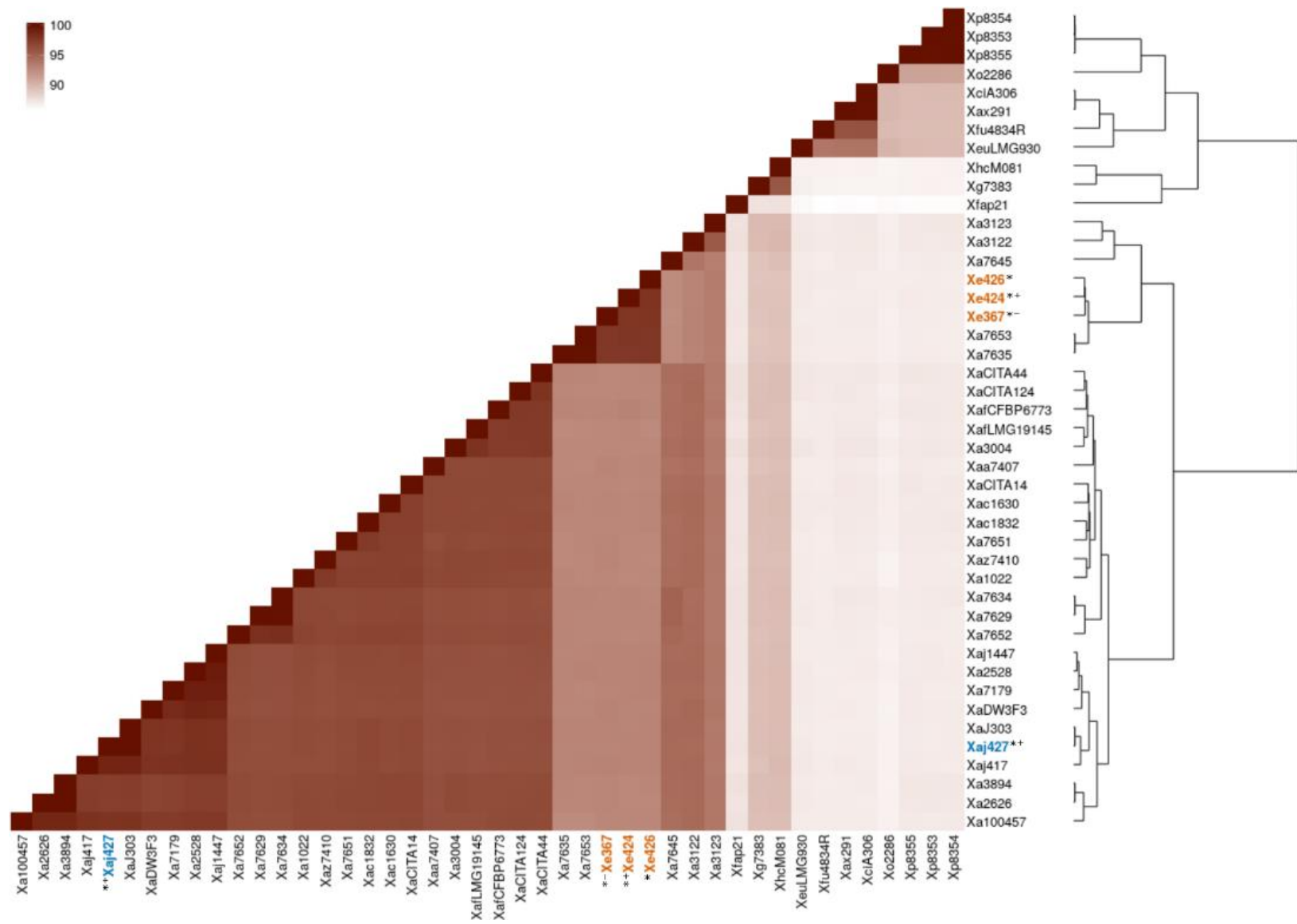


Figure III.1. Heatmap representing average nucleotide identity (ANI) (disclosed in Supplementary Table III.S2) and the respective distance tree cladogram, determined for 44 *Xanthomonas* genomes including the four genomes disclosed in this study, indicated with (*) and highlighted in orange (CPBF 367, CPBF 424^T, CPBF 426) or blue (CPBF 427). Pathogenic and nonpathogenic on walnut are marked with (+) or (–), respectively. The color scale ranging from white to dark red depicts lower to higher similarity values, respectively. The strain names refer to the code field from Supplementary Table III.S1

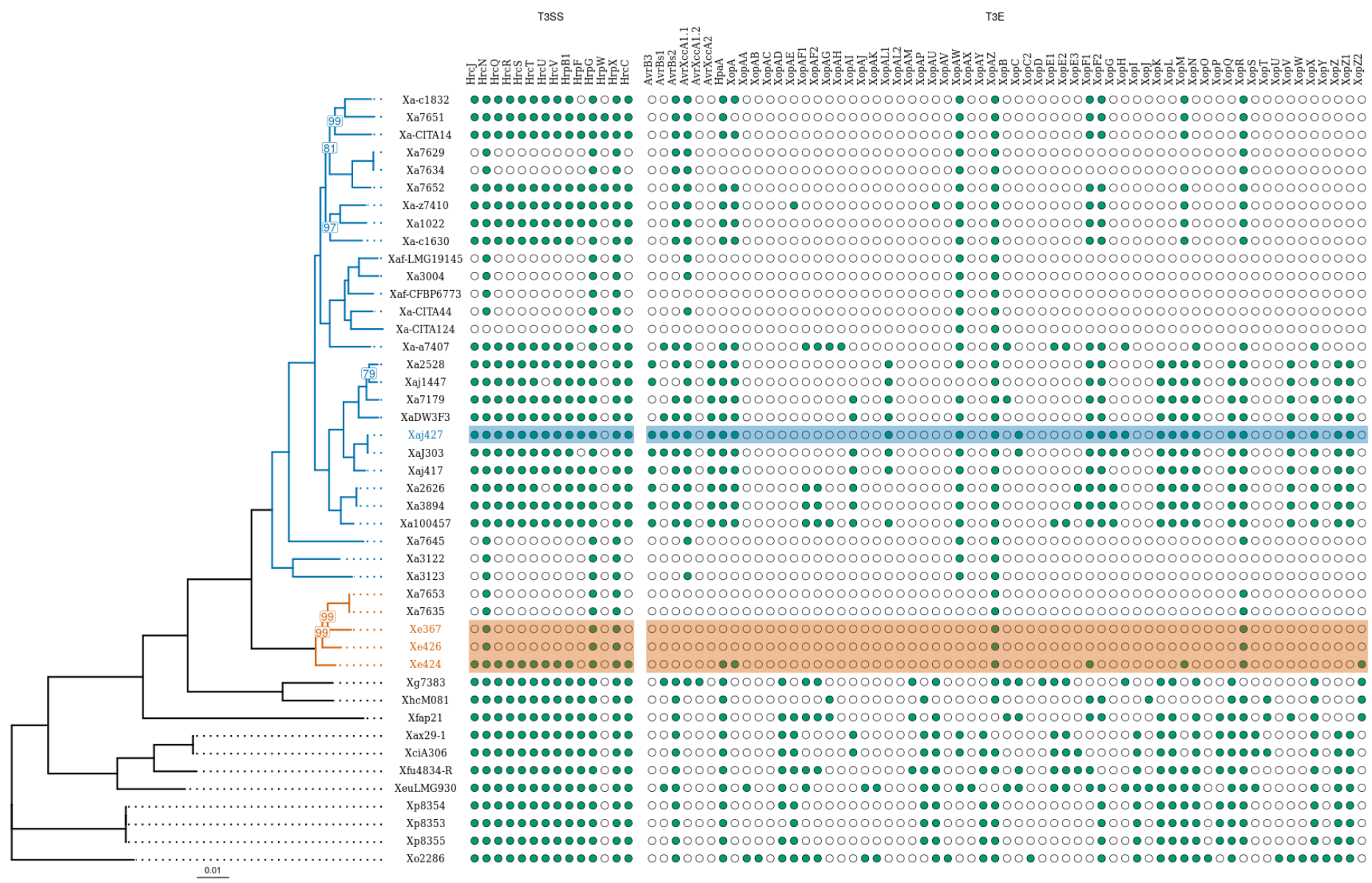


Figure III.2. Maximum likelihood tree based on concatenated sequences of the 1149 core genes of 44 *Xanthomonas* genomes, including the *Xanthomonas euroxanthea* strains (branches in orange) and *Xanthomonas arboricola* strains (branches in blue). Phylogenetic relations were inferred using RaxML and the phylogram represented with the R package "ggtree". Supporting values from 500 bootstrap replicates are indicated near nodes. Dot plot scheme represents the presence/absence scheme for type 3 secretion system (T3SS) and effectors (T3E) putative homologs; ●, present; ○, not present; considering a tBLASTn hit with a query coverage threshold $\geq 75\%$, and sequence identity with $\geq 40\%$ cut-off. Results for genomes disclosed in this study are marked with (*); also, *X. euroxanthea* strains CPBF 367, CPBF 424T, and CPBF 426 are highlighted in orange and for *X. arboricola* pv. *juglandis* CPBF 427 in blue. Pathogenic and nonpathogenic on walnut are marked with (+) or (–), respectively. The strain names refer to the code field from Supplementary Table III.S1. Best BLAST results and accession numbers of sequences used as query are disclosed in Supplementary Table III.S3

3.3. Genetic Patrimony Retrieved from the *X. euroxantha* and *X. arboricola* pv. *juglandis* strains Isolated from a Single Walnut Host Tree

To disclose the gene pool of xanthomonads found in a single walnut tree host, the total number of CDSs corresponding to non-redundant genes retrieved from the four strains studied (CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427) was determined. The results highlighted in a Venn diagram showed that the core genome, i.e., the set of genes common to the four strains (Figure III.3). The remaining CDSs, which corresponds to the accessory genome, represent the differential gene content of these strains; that is, they include the strain-specific CDSs and genes shared by two or more strains. When focusing on genomic sub-sets, 212 CDSs are shared exclusively by *X. euroxantha* strains (CPBF 367, CPBF 424^T, and CPBF 426), and 52 CDSs are shared exclusively by the two pathogenic strains (CPBF 424^T, and CPBF 427). Among the genes shared by *X. euroxantha*, a great number of regulator proteins and hypothetical proteins were found, with the latter constituting 24% of the total genes shared by *X. euroxantha*. Regarding strain-specific genomic contents, 621 strain-specific CDSs were retrieved for *X. arboricola* pv. *juglandis*, while for *X. euroxantha* strains, 208 unique CDSs were identified for CPBF 367, 213 unique CDSs for CPBF 426, and 186 unique CDSs for CPBF 424^T. From these singletons, 60%, 52%, and 45% were assigned as hypothetical proteins for CPBF 367, CPBF 424^T, and CPBF 426, respectively.

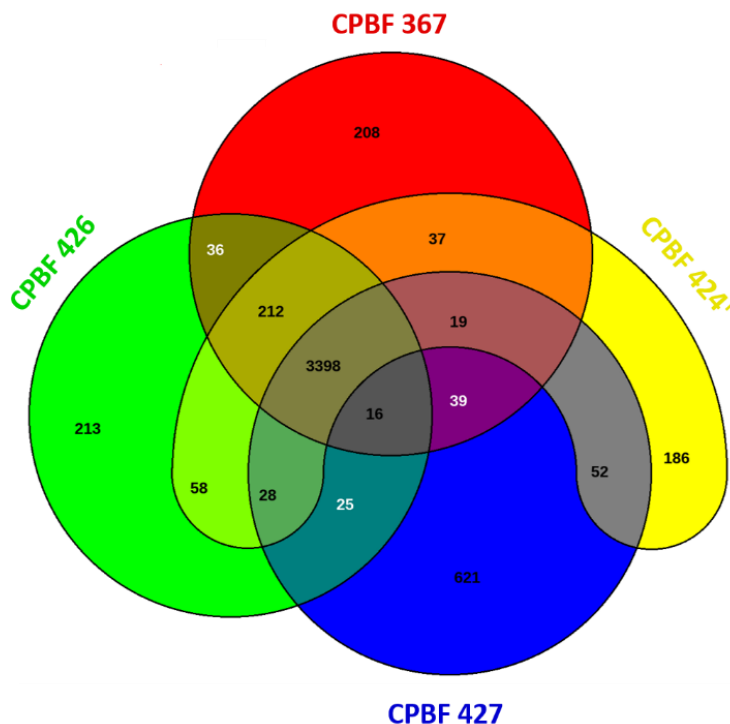


Figure III.3. Venn diagram highlighting the genomic patrimony from CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427 genomes. The core genome is given by the interception of the four strains and corresponds to the number of orthologous CDSs shared (3398 CDSs). Strain specific CDSs are also represented in the periphery of the diagram (208 for CPBF 367, 213 for CPBF 426, 186 for CPBF 424^T, and 621 for CPBF 427). The remaining combinations represent the number of orthologous shared between two to three genomes. The strain names refer to the code field from Supplementary Table III.S1.

3.4. Pathogenic and Virulence-Related Factor Prediction

In addition to the functional analysis, the profile of *Xanthomonas* pathogenicity and virulence factors characterized in previous studies was determined, unravelling notable differences between the consortium strains isolated from the same walnut host (Supplementary Figures III.S1–III.S5, Supplementary Table III.S3). All four xanthomonad strains were found to share numerous genes associated with pathogenesis, namely genes for the biosynthesis of the xanthan polysaccharide biosynthesis (operon *gumBCDEFGHIJKLMN*), lipopolysaccharide biosynthesis, the flagellar system, the regulatory *rpf* cluster of pathogenicity factor synthesis, the *xps* gene of T2SS (*xpsD*, *E*, *F*, *G*, *H*, *I*, *J*, *K*, *L*, and *M*), homologs of the type IV pilus (T4P), several T4SS genes (*virB1*, *virB2*, *virB3*, *virB4*, *virB6*, *virB8*, *virB9*, *virB10*, *virB11*, and *virD4*), as well as genes from the T6SS (Supplementary Table III.S3). Regarding the T2SS, the *xpsN* gene was only present in the *X. euroxantha* strains (CPBF 367, CPBF 424^T, and CPBF 426) (Table III.2, Supplementary Figure III.S1), and five T4P-related genes (*pilY1*, *pilX*, *pilW*, *pilV*, and *fimT*) were found to be exclusively present in the pathogenic CPBF 424T (Table III.2, Supplementary Figure III.S2). Concerning the presence of non-fimbrial adhesins, it is noticeable that the *X. euroxantha* CPBF 424^T and CPBF 426 strains did not harbor homologous genes for *fhaB1* and *fhaB2*, which encode filamentous hemagglutinin-related proteins, regardless of the fact that all four strains share at least three homologs to non-fimbrial adhesins (Table III.2, Supplementary Figure III.S3). Additionally, no major differences were observed between the four strains studied regarding genes encoding proteins associated with *Xanthomonas* sensing and chemotaxis mechanisms, with the exception of a methyl-accepting chemotaxis protein and a chemotaxis protein of which homologs were present in all three *X. euroxantha* strains but not in *X. arboricola* pv. *juglandis* strain CPBF 427 (Table III.2, Supplementary Figure III.S4). The main differences between *X. arboricola* pv. *juglandis* and *X. euroxantha* strains were observed in the profile of pectolytic enzymes associated with T2SS (Table III.2, Supplementary Figure III.S5). Homologs for a pectate lyase E and a pectinesterase were only identified in the genomes of the three *X. euroxantha* strains. On the contrary, homologs of a degenerated pectate lyase, an endoglucanase, a rhamnogalacturonase B, and a polygalacturonase were present in CPBF 427 and not in *X. euroxantha* strains. Furthermore, homologs of xylosidase/arabinosidase (*xylB*) were found in the two *X. arboricola* pv. *juglandis* strains CPBF 427 and in the pathogenic *X. euroxantha* strain CPBF 424^T.

Table III.2. Comparison of the occurrence of homologs associated with chemotaxis, non-fimbrial adhesins, T2SS, extracellular enzymes, and T4P, between *Xanthomonas euroxanthea* strains (CPBF 367, CPBF 424^T, and CPBF 426) and the *Xanthomonas arboricola* pv. *juglandis* (*Xaj*) strain (CPBF 427).

			<i>X. euroxanthea</i>			<i>Xaj</i>
Protein/Gene Name	Label	GenBank Accession Number	CPBF 367	CPBF 424 ^T	CPBF 426	CPBF 427
Chemotaxis-related proteins:						
Methyl-accepting chemotaxis protein	XCV1938	CAJ23615.1	●*	●	●	○
Chemotaxis protein	XAC3768	AAM38611.1	●	●	●	○
Non-fimbrial adhesins:						
Filamentous hemagglutinin-related protein (<i>fhaB1</i>)	FhaB1	CAJ23537.1	●	○	○	●
Filamentous hemagglutinin-related protein (<i>fhaB2</i>)	FhaB2	CAJ23538.1	●	○	○	●
Type II secretion system (T2SS):						
General secretion pathway protein XpsN (<i>xpsN</i>)	XpsN	WP_011035909.1	●	●	●	○
Extracellular enzymes:						
Xylosidase/arabinosidase (<i>xylB</i>)	XylB.1	WP_011036375.1	○	●	○	●
Xylosidase/arabinosidase (<i>xsa</i>)	Xsa.1	WP_011037540.1	○	○	○	●
Xylosidase/arabinosidase (<i>xylB</i>)	XylB.2	WP_011039174.1	○	●	○	●
Endoglucanase (<i>bcsZ</i>)	bcsZ	AAM38359.1	○	○	○	●
Polygalacturonase (<i>pglA</i>)	pglA	WP_011037410.1	○	○	○	●
Pectate lyase E (<i>peIA</i>)	peIA	WP_011035380.1	●	●	●	○
Degenerated pectate lyase (<i>pel</i>)	pel	AAM37225.1	○	○	○	●
Pectinesterase	XCC0121	WP_011035379.1	●	●	●	○
Type IV pilus (T4P):						
PilY1 protein (<i>pilY1</i>)	pilY1	WP_011051753.1	○	●	○	○
PilX protein (<i>pilX</i>)	pilX	WP_011051754.1	○	●	○	○
PilW protein (<i>pilW</i>)	XAC2667	WP_040107776.1	○	●	○	○
Pre-pilin leader sequence (<i>pilV</i>)	pilV	WP_011051756.1	○	●	○	○
Pre-pilin like leader sequence (<i>fimT</i>)	fimT	WP_011051757.1	○	●	○	○

present; ○, not present.

3.5. Type 3 Secretion System and Its Effectors

The presence of homologous proteins for the type 3 secretion system and its effectors were evaluated in CPBF 367, CPBF 424^T, CPBF 426, CPBF 427, and other *Xanthomonas* genomes, including known pathogenic and nonpathogenic *X. arboricola* strains (Figures III.2 and Supplementary Figure III.S6). For both T3SS and T3E proteins, *X. euroxantha* CPBF 367 and CPBF 426 presented the same profiles as nonpathogenic *X. arboricola* strains CFBP 7653 and CFBP 7635 and a reduced number of homologs than pathogenic *X. euroxantha* strains CPBF 424^T. Particularly for T3SS, *X. euroxantha* strains CPBF 367 and CPBF 426 revealed a similar profile to nonpathogenic CFBP 7653, CFBP 7635, CFBP 7645, CFBP 7634, CFBP 7629, CITA 124, and CITA 44 with the presence of only HrpG, HrpX, and HrcN. Conversely, pathogenic *X. euroxantha* strain CPBF 424^T demonstrated a more complete profile that includes HrpG, HrpX, HrcN, HrcJ, HrcQ, HrcR, HrcS, HrcT, HrcV, HrpB1, HrcC, and HrcU, similarly to *X. arboricola* pathogenic strains including CPBF 427 (strain isolated from the same walnut tree host, at the same sampling event) and nonpathogenic *X. arboricola* strains CFBP 7652, CFBP 7651, and CITA 14 (Figures III.2 and Supplementary Figure III.S6). Regardless of the arsenal of T3SS proteins observed for CPBF 424^T, this strain includes fewer T3 effector proteins (XopAZ, XopR, HpaA, XopM, XopF1, XopA, XopZ2), similarly to nonpathogenic *X. arboricola* strains CFBP 7651, CFBP 7652, CITA 14. CPBF 367, and CPBF 426, which only possess XopAZ and XopR, as observed for nonpathogenic CFBP 7653 and CFBP 7635.

4. Discussion

The occurrence of distinct *Xanthomonas* populations colonizing the same host plant has been previously documented (Jacques *et al.*, 2016b). In walnut and stone fruit trees, besides the presence of *X. arboricola* strains belonging to pathovar *juglandis* and *pruni*, characteristic of these two host species, respectively, the isolation of distinct yellow-pigmented xanthomonads has been reported, mostly represented by nonpathogenic lineages that do not form a phylogenetically coherent group with the pathogenic strains of *X. arboricola* pathovars (Essakhi *et al.*, 2015; Garita-Cambronero *et al.*, 2016b; Jacques *et al.*, 2016a). This raises the need to understand the role played in the pathosystems by these bacteria characterized by distinct genotypes.

In this study, four *Xanthomonas* strains (CPBF 367, CPBF 424^T, CPBF 426, CPBF 427) isolated from the same walnut tree were sequenced to provide an in-depth characterization of these co-colonizing strains to disclose differential genomic contents putatively related to pathogenicity, virulence, and other specific niche adaptations. ANI

and a core-genome phylogenetic analysis disclosed the presence of two different *Xanthomonas* species in one disease walnut tree, with the confirmation that CPBF 427 belongs to *X. arboricola* pv. *juglandis*, whilst the other three strains, CPBF 367, CPBF 424^T, and CPBF 426, were already assigned to the recently described species *X. euroxanthea* (Martins *et al.*, 2020). Moreover, two of the atypical strains described by Essakhi *et al.* (Essakhi *et al.*, 2015) as *X. arboricola*, CFBP 7635 and CFBP 7653, were now confirmed to belong to *X. euroxanthea*. The role of nonpathogenic strains in *Xanthomonas* evolution and its potential for pathogenicity emergence is often neglected due to their unvalued direct agro-economic impact. However, a recent genomics study on nonpathogenic strains reinforces our lack of knowledge regarding the lifestyle of these strains. In fact, a nonpathogenic isolate from citrus (LMG 8993) was revealed to belong to *X. arboricola* species, being placed phylogenetically closest to nonpathogenic *X. arboricola* isolates from walnut (CFBP 7634 and CFBP 7651), than with other nonpathogenic citrus isolates (Bansal *et al.*, 2020). Currently, the ecological, evolutionary, and pathogenicity implications of this co-colonization are not understood, but it is hardly refutable that this knowledge is needed for the improvement of efficient phytosanitary practices and the design of appropriate management strategies (Quibod *et al.*, 2016).

The genome size of CPBF 427 (5.23 Mbp) is roughly equal to the genome size reported for other sequenced *X. arboricola* pv. *juglandis* genomes, namely *Xaj* 417 (Pereira *et al.*, 2015), NCPPB 1447 (Vauterin *et al.*, 1995), J303 (Retamales *et al.*, 2017), DW3F3, (Fu *et al.*, 2018), and CFBP 2528^T and CFBP 7179 (Cesbron *et al.*, 2015) or CFSAN033077 and CFSAN033080 (Higuera *et al.*, 2015). The genomes of *X. euroxanthea* strains CPBF 367, CPBF 424^T, and CPBF 426 were smaller (ranging from $\approx$ 4.90 to 4.97 Mbp), presenting values closer to the nonpathogenic or avirulent strains of *X. arboricola*, i.e., with uncertain pathogenicity or belonging to non-*juglandis*, non-*pruni*, and non-*corylina* pathovars, with genome sizes inferior to 5 Mb (Ignatov *et al.*, 2015; Garita-Cambronero *et al.*, 2016a; Harrison *et al.*, 2016). Despite the presence of one plasmid in strains CPBF 367 and CPBF 426, all the virulence-related homologs were chromosomal. The genetic patrimony of the studied strains illustrated by a Venn diagram encompasses genes associated with basic biological aspects of the *Xanthomonas* genus, as phenotypic traits (Medini *et al.*, 2005). Among these, it was possible to discern 212 genes shared exclusively by *X. euroxanthea* strains. The analysis of these gene sets specific to *X. euroxanthea* or *X. arboricola* pv. *juglandis* suggests the presence of genes encoding for proteins associated with biochemical functions that may confer selective advantages, such as adaptation to different niches, pathogenicity, or colonization of a new host

(Medini *et al.*, 2005). Some genetic determinants of virulence were shown to be group-specific or even strain-specific. For example, *X. euroxanthea* strains evidenced an exclusive presence of the *xpsN* gene, which encodes the XpsN protein of the type II secretion system. Proteins of the xps system were shown to be associated with virulence of *Xanthomonas* species such as *X. campestris*, *X. oryzae*, and *X. euvesicatoria* (Ryan *et al.*, 2011). Furthermore, the *X. euroxanthea* pathogenic strain CPBF 424^T harbors a set of genes that encodes proteins associated with type IV pilus (PilY1, PilX, PilW, PilV, FimT), some of which are considered primary structures of the T4P pilin subunits. Indeed, T4P could play an important role in the pathogenesis of various species of *Xanthomonas*, and in some cases it is thought that this system has a role in plant colonization (Dunger *et al.*, 2016; Herseemann *et al.*, 2017). Comparative studies between pathogenic *X. arboricola* pv. *juglandis* and nonpathogenic *X. arboricola* strains have shown a differential repertoire of genes encoding chemotaxis-related proteins and proteins related to type I, II, and IV secretion systems (Cesbron *et al.*, 2015). Furthermore, a differential repertoire of non-fimbrial adhesins involved in different functions related to bacterial attachment to the host surface were found in all strains, whereas homologs of non-fimbrial adhesins *fhaB* probably associated with the bacteria colonization were only identified in pathogenic strains (Cesbron *et al.*, 2015). In the same way, homologs of proteins involved in the biogenesis of type IV pilus were observed, but the absence of PilA, PilX, and/or PilV proteins in the genomes of pathogenic and nonpathogenic strains may point to the absence of bacterial surface filaments in all strains (Cesbron *et al.*, 2015). Moreover, differences can be pinpointed between pathogenic and nonpathogenic *X. euroxanthea* strains, particularly, the five type IV pilus proteins exclusively present in CPBF 424^T, and the xylosidases (XylB.1 and XylB.2) only present in CPBF 424^T and also in CPBF 427. It was also possible to identify homologs encoding for proteins specific to *X. arboricola* pv. *juglandis* strain CPBF 427 that are missing in *X. euroxanthea*, such as the extracellular enzymes xylosidase (Xsa.1), endoglucanase, polygalacturonase, and degenerate pectate lyase.

Furthermore, major genomic differences between the strains analyzed in this study were observed for T3SS and related T3E homologs. In *Xanthomonas* spp., T3SS is crucial for translocating effector proteins that have a key role in bacterial proliferation in host tissues and the development of disease symptoms (Büttner & Bonas, 2010). The majority of pathogenic strains from the 44 analyzed genomes, including CPBF 427 and *X. euroxanthea* CPBF 424^T, displayed a T3SS profile comprised of most homologous genes for highly conserved T3SS of the Hrp2 family (Hajri *et al.*, 2011; Cesbron *et al.*, 2015; Garita-Cambronero *et al.*, 2016b, 2017, 2018). Interestingly, a similar pattern was

spotted for nonpathogenic *X. arboricola* strains CFBP 7652, CFBP 7651, and CITA 14 and for the pathogenic *X. euroxanthea* strain CPBF 424^T, with the exception for T3SS homologs HrpF and HrpW, which were slightly below the 75% query coverage threshold used to filter for the most conserved homologs. Conversely, nonpathogenic *X. euroxanthea* CPBF 367 and CPBF 426, and nonpathogenic *X. arboricola* CFBP 7653, CFBP 7645, CFBP 7635, CFBP 7634, CFBP 7629, CITA 44, and CITA 124 lacked most of the genes coding for the macromolecular structure of T3SS, as well as the *hrpF* gene, involved in the translocation of T3E (Cesbron *et al.*, 2015), despite harboring regulators genes of T3SS, as *hrpX* and *hrpG* (Guo *et al.*, 2011; Jacobs *et al.*, 2015). Interestingly CPBF 424^T lacks homologs for several pathogenicity genes thought to be essential for *X. arboricola* pv. *juglandis* strains. In fact, CPBF 424^T harbors homologs for seven known effectors, i.e., less than the nine to ten T3Es found in nonpathogenic *X. arboricola* CFBP 7651, CFBP 7652, and CITA 14. Only two T3Es were identified in nonpathogenic *X. euroxanthea* stains (CPBF 367 and CPBF 426), suggesting that *X. euroxanthea* may also make use of other virulence and pathogenicity-related proteins to trigger infection. Still, the intricate mechanism for successful pathogenicity of *X. euroxanthea* CPBF 424^T can only be disclosed with further investigation and dedicated functional assays.

Cesbron *et al.* (Cesbron *et al.*, 2015) and Garita-Cambronero *et al.* (Garita-Cambronero *et al.*, 2017) started to elucidate the mechanisms associated with the emergence of *X. arboricola* pv. *juglandis* and *X. arboricola* pv. *pruni* pathogenic strains. This was achieved by comparative genomics of *X. arboricola* strains differing on their pathogenicity, including the nonpathogenic *X. arboricola* strains isolated from walnut and evaluated in this study (CFBP 7634 and CFBP 7651), and *X. arboricola* strains (CFBP 14, CFBP 44, and CFBP 124) isolated from *Prunus*. Regardless of the contributions of these comparative genomics studies in highlighting the importance of T3SS and T3E genes, dedicated functional studies are still required to identify essential genes for successful infection. Furthermore, these nonpathogenic strains are particularly valuable to address questions regarding pathogenicity evolution in *Xanthomonas*.

5. Conclusions

The extensive genomic comparison of four walnut-associated strains isolated from the same walnut specimen and belonging to two species (*X. arboricola* pv. *juglandis* and *X. euroxanthea*), including pathogenic (CPBF 427 and CPBF 424^T) and nonpathogenic strains (CPBF 367 and CPBF 426), provides insights about niche-specific adaptations that could inform on the role played by each of these strains in the co-colonization of walnut. Comprehensive genomics analysis, which also includes previously reported nonpathogenic *X. arboricola* strains, shows that two of these strains, CFBP 7635 and CFBP 7653, belong to *X. euroxanthea* species. The data gathered suggest a pattern of homologous genes putatively associated with pathogenicity, virulence, and niche-specific adaptations that need to be addressed in future functional studies to determine their importance in walnut diseases caused by *Xanthomonas*.

Funding information

Funding for this work was provided by the EU Horizon 2020 Research and Innovation program (grant agreement 668981 and 857251), European Structural and Investment Funds (ESIFs) through COMPETE 2020, National Funds through FCT (project EVOXANT-PTDC/BIA-EVF/3635/2014-POCI-01-0145-FEDER-016600), and the Operational Thematic Program for Competitiveness and Internationalization (POCI) under the PORTUGAL 2020 Partnership Agreement through the European Regional Development Fund (FEDER) project PORBIOTA—POCI-01-0145-FEDER-022127. C.F. and L.M. were supported by fellowships from FCT (SFRH/BD/95913/2013 and SFRH/BD/137079/2018, respectively). This article is based upon work from COST Action CA16107 EuroXanth, supported by COST (European Cooperation in Science and Technology). C.F. was granted a Short-Term Scientific Mission by this COST Action to perform genomic assemblies in Wädenswil (CH). The EDGAR platform is financially supported by the BMBF grant FKZ031A533 within the de.NBI network. The APC was funded COST Action CA16107 EuroXanth.

Data Availability Statement

The assembled genome sequences have been deposited in the European Nucleotide Archive (ENA) under the accession numbers GCA_903989455, CGA_905187425, GCA_903989465, and GCA_903989475 for CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427, respectively.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.
- An SQ, Potnis N, Dow M *et al.*, 2019. Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen *Xanthomonas*. *FEMS Microbiology Reviews* **44**, 1–32.
- Bansal K, Kumar S, Patil PB, 2020. Phylogenomic Insights into Diversity and Evolution of Nonpathogenic *Xanthomonas* Strains Associated with Citrus . *mSphere* **5**.
- Belisario A, Maccaroni M, Corazza L, Balmas V, Valier A, 2002. Occurrence and etiology of brown apical necrosis on Persian (English) walnut fruit. *Plant Disease* **86**, 599–602.
- Blom J, Albaum SP, Doppmeier D *et al.*, 2009. EDGAR: A software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics* **10**.
- Blom J, Kreis J, Spänig S *et al.*, 2016. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic acids research* **44**, W22–W28.
- Büttner D, Bonas U, 2010. Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiology Reviews* **34**, 107–133.
- Cesbron S, Briand M, Essakhi S *et al.*, 2015. Comparative Genomics of Pathogenic and Nonpathogenic Strains of *Xanthomonas arboricola* Unveil Molecular and Evolutionary Events Linked to Pathoadaptation. *Frontiers in Plant Science* **6**, 1126.
- Crossman L, Dow JM, 2004. Biofilm formation and dispersal in *Xanthomonas campestris*. *Microbes and Infection* **6**, 623–629.
- Darrasse A, Carrère S, Barbe V *et al.*, 2013. Genome sequence of *Xanthomonas fuscans* subsp . *fuscans* strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. *BMC Genomics* **14**, 1–30.
- Dowson DW, 1939. On the systematic position and generic names of the Gram negative bacterial plant pathogens. *Zentralblatt fur Bakteriologie, Parasitenkunde und Infektionskrankheiten*, 2 **100**, 177–193.
- Dunger G, Llontop E, Guzzo CR, Farah CS, 2016. The *Xanthomonas* type IV pilus. *Current Opinion in Microbiology* **30**, 88–97.
- Edler D, Klein J, Antonelli A, Silvestro D, 2021. raxmlGUI 2.0: A graphical interface and

toolkit for phylogenetic analyses using RAxML (M Matschiner, Ed.). *Methods in Ecology and Evolution* **12**, 373–377.

Essakhi S, Cesbron S, Fischer-Le Saux M, Bonneau S, Jacques M-A, Manceau C, 2015. Phylogenetic and Variable-Number Tandem-Repeat Analyses Identify Nonpathogenic *Xanthomonas arboricola* Lineages Lacking the Canonical Type III Secretion System. *Applied and Environmental Microbiology* **81**, 5395–5410.

Fernandes C, Albuquerque P, Mariz-Ponte N, Cruz L, Tavares F, 2021. Comprehensive diversity assessment of walnut-associated xanthomonads reveal the occurrence of distinct *Xanthomonas arboricola* lineages and of a new species (*Xanthomonas euroxanthea*) within the same tree. *Plant Pathology* **70**, 5.

Fernandes C, Blom J, Pothier JF, Tavares F, 2018a. High-Quality Draft Genome Sequence of *Xanthomonas* sp. Strain CPBF 424, a Walnut-Pathogenic Strain with Atypical Features. *Microbiology Resource Announcements* **7**.

Fernandes C, Blom J, Pothier JF, Tavares F, 2018b. High-Quality Draft Genome Sequence of *Xanthomonas arboricola* pv. *juglandis* CPBF 1521, Isolated from Leaves of a Symptomatic Walnut Tree in Portugal without a Past of Phytosanitary Treatment . *Microbiology Resource Announcements* **7**.

Filloux A, 2004. The underlying mechanisms of type II protein secretion. *Biochimica et Biophysica Acta - Molecular Cell Research* **1694**, 163–179.

Fischer-Le Saux M, Bonneau S, Essakhi S, Manceau C, Jacques M-A, 2015. Aggressive Emerging Pathovars of *Xanthomonas arboricola* Represent Widespread Epidemic Clones Distinct from Poorly Pathogenic Strains, as Revealed by Multilocus Sequence Typing. *Applied and Environmental Microbiology* **81**, 4651–4668.

Fu B, Chen Q, Wei M et al., 2018. Complete Genome Sequence of *Xanthomonas arboricola* pv. *juglandis* Strain DW3F3, Isolated from a *Juglans regia* L. Bacterial Blighted Fruitlet . *Genome Announcements* **6**, e00023-18.

Garita-Cambronero J, Palacio-Bielsa A, Cubero J, 2018. *Xanthomonas arboricola* pv. *pruni*, causal agent of bacterial spot of stone fruits and almond: its genomic and phenotypic characteristics in the *X. arboricola* species context. *Molecular Plant Pathology* **19**, 2053–2065.

Garita-Cambronero J, Palacio-Bielsa A, López MM, Cubero J, 2016a. Draft genome sequence for virulent and avirulent strains of *Xanthomonas arboricola* isolated from *Prunus* spp. in Spain. *Standards in Genomic Sciences* **11**, 1–10.

- Garita-Cambronero J, Palacio-Bielsa A, López MM, Cubero J, 2016b. Comparative genomic and phenotypic characterization of pathogenic and non-pathogenic strains of *Xanthomonas arboricola* reveals insights into the infection process of bacterial spot disease of stone fruits. *PLoS ONE* **11**, e0161977.
- Garita-Cambronero J, Palacio-Bielsa A, López MM, Cubero J, 2017. Pan-genomic analysis permits differentiation of virulent and non-virulent strains of *Xanthomonas arboricola* that cohabit *Prunus* spp. and elucidate bacterial virulence factors. *Frontiers in Microbiology* **8**, 1–17.
- Giovanardi D, Bonneau S, Gironde S, Saux MF Le, Manceau C, Stefani E, 2016. Morphological and genotypic features of *Xanthomonas arboricola* pv. *juglandis* populations from walnut groves in Romagna region, Italy. *European Journal of Plant Pathology* **145**.
- Goldstein S, Beka L, Graf J, Klassen JL, 2019. Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. *BMC Genomics* **20**, 23.
- Guo Y, Figueiredo F, Jones J, Wang N, 2011. HrpG and HrpX Play Global Roles in Coordinating Different Virulence Traits of *Xanthomonas axonopodis* pv. *citri*. *Molecular Plant-Microbe Interactions* **24**, 649–661.
- Hajri A, Meyer D, Delort F, Guillaumès J, Brin C, Manceau C, 2010. Identification of a genetic lineage within *Xanthomonas arboricola* pv. *juglandis* as the causal agent of vertical oozing canker of Persian (English) walnut in France. *Plant Pathology* **59**, 1014–1022.
- Hajri A, Pothier JF, Saux MF Le *et al.*, 2011. Type three effector gene distribution and sequence analysis provide new insights into the pathogenicity of plant-pathogenic *Xanthomonas arboricola*. *Applied and Environmental Microbiology* **78**, 371–384.
- Harrison J, Grant MR, Studholme DJ, 2016. Draft genome sequences of two strains of *Xanthomonas arboricola* pv. *celebensis* isolated from banana plants. *Genome Announcements* **4**.
- Hersemann L, Wibberg D, Blom J *et al.*, 2017. Comparative genomics of host adaptive traits in *Xanthomonas translucens* pv. *graminis*. *BMC Genomics* **18**, 35.
- Higuera G, González-Escalona N, Véliz C, Vera F, Romero J, 2015. Draft Genome Sequences of Four *Xanthomonas arboricola* pv. *juglandis* Strains Associated with Walnut Blight in Chile: TABLE 1 . *Genome Announcements* **3**.

- Ignatov AN, Kyrova EI, Vinogradova S V., Kamionskaya AM, Schaad NW, Luster DG, 2015. Draft genome sequence of *Xanthomonas arboricola* strain 3004, a causal agent of bacterial disease on barley. *Genome Announcements* **3**, 1572–1586.
- Ivanović Ž, Popović T, Janse J *et al.*, 2014. Molecular assessment of genetic diversity of *Xanthomonas arboricola* pv. *juglandis* strains from Serbia by various DNA fingerprinting techniques. *European Journal of Plant Pathology* **141**, 133–145.
- Jacobs JM, Pesce C, Lefeuvre P, Koebnik R, 2015. Comparative genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*. *Frontiers in Plant Science* **6**.
- Jacques M-A, Arlat M, Boulanger A *et al.*, 2016a. Using Ecology, Physiology, and Genomics to Understand Host Specificity in *Xanthomonas*. *Annual Review of Phytopathology* **54**, 163–187.
- Jacques MA, Guimbaud JF, Briand M, Indiana A, Darrasse A, 2016b. Flagellar Motility and Fitness in Xanthomonads. In: *Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria*. Wiley Blackwell, 1265–1273.
- Kaluzna M, Pulawska J, Waleron M, Sobiczewski P, 2014. The genetic characterization of *Xanthomonas arboricola* pv. *juglandis*, the causal agent of walnut blight in Poland. *Plant Pathology* **63**, 1404–1416.
- Kchouk M, Gibrat J-F, Elloumi M, 2017. Generations of Sequencing Technologies: From First to Next Generation. *Biology and Medicine* **9**.
- Lamichhane JR, 2014. *Xanthomonas arboricola* diseases of stone fruit, almond, and walnut trees: Progress toward understanding and management. *Plant Disease* **98**, 1600–1610.
- Lee I, Kim YO, Park SC, Chun J, 2016. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *International Journal of Systematic and Evolutionary Microbiology* **66**, 1100–1103.
- Lee BM, Park YJ, Park DS *et al.*, 2005. The genome sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice. *Nucleic Acids Research* **33**, 577–586.
- Lerat E, Daubin V, Moran NA, 2003. From Gene Trees to Organismal Phylogeny in Prokaryotes: The Case of the γ -Proteobacteria (David Penny, Ed.). *PLoS Biology* **1**, e19.

- Leyns F, De Cleene M, Swings JG, De Ley J, 1984. The host range of the genus *Xanthomonas*. *The Botanical Review* **50**, 308–356.
- Lu H, Giordano F, Ning Z, 2016. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics and Bioinformatics* **14**, 265–279.
- Martins L, Fernandes C, Blom J, Dia NC, Pothier JF, Tavares F, 2020. *Xanthomonas euroxanthea* sp. nov., a new xanthomonad species including pathogenic and non-pathogenic strains of walnut. *International Journal of Systematic and Evolutionary Microbiology* **70**, 9.
- Medini D, Donati C, Tettelin H, Massignani V, Rappuoli R, 2005. The microbial pan-genome. *Current Opinion in Genetics and Development* **15**, 589–594.
- Meyer DF, Bogdanove AJ, 2009. Genomics-driven advances in *Xanthomonas* biology. In: Jackson R, ed. *Genomics-driven advances in Xanthomonas biology*. Norfolk: Caister Academic Press, 147–161.
- Notredame C, Higgins DG, Heringa J, 2000. T-coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* **302**, 205–217.
- Oxford Nanopore Technologies Large insights into microorganisms.
- Parte AC, Carbasse JS, Meier-Kolthoff JP, Reimer LC, Göker M, 2020. List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology* **70**, 5607–5612.
- Pereira UP, Gouran H, Nascimento R, Adaskaveg JE, Goulart LR, Dandekar AM, 2015. Complete genome sequence of *Xanthomonas arboricola* pv. *juglandis* 417, a copper-resistant strain isolated from *Juglans regia* L. Genome. *Genome Announcements* **3**.
- Pierce NB, 1901. Walnut bacteriosis. *Botanical Gazette* **31**, 272–273.
- Quibod IL, Perez-Quintero A, Booher NJ et al., 2016. Effector Diversification Contributes to *Xanthomonas oryzae* pv. *Oryzae* Phenotypic Adaptation in a Semi-Isolated Environment. *Scientific Reports* **6**, 1–11.
- Retamales J, Segovia C, Alvarado R, Nuñez P, Santander J, 2017. Draft Genome Sequence of *Xanthomonas arboricola* pv. *juglandis* J303, Isolated from Infected Walnut Trees in Southern Chile . *Genome Announcements* **5**.
- Richter M, Rosselló-Móra R, 2009. Shifting the genomic gold standard for the prokaryotic

- species definition. *Proceedings of the National Academy of Sciences USA* **106**, 19126–19131.
- Ryan RP, Vorhölter FJ, Potnis N *et al.*, 2011. Pathogenomics of *Xanthomonas*: Understanding bacterium-plant interactions. *Nature Reviews Microbiology* **9**, 344–355.
- Shrivastava S, Mande SS, 2008. Identification and functional characterization of gene components of type VI secretion system in bacterial genomes. *PLoS ONE* **3**, 2955.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM, 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212.
- Smith RE., Smith CO., Ramsey HJ., 1912. *Walnut culture in California : walnut blight*. Berkeley, California: California Agricultural Experiments Station Publications.
- Stamatakis A, 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690.
- Swings JG, Civerolo EL, 1993. *Xanthomonas*. Chapman & Hall.
- Szczesny R, Jordan M, Schramm C *et al.*, 2010. Functional characterization of the Xcs and Xps type II secretion systems from the plant pathogenic bacterium *Xanthomonas campestris* pv *vesicatoria*. *New Phytologist* **187**, 983–1002.
- Tatusova T, Dicuccio M, Badretdin A *et al.*, 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research* **44**, 6614–6624.
- Teixeira M, Fernandes C, Chaves C, Pinto J, Tavares F, Fonseca NA, 2021. Complete Genome Sequence Obtained by Nanopore and Illumina Hybrid Assembly of *Xanthomonas arboricola* pv. *juglandis* CPBF 427, Isolated from Buds of a Walnut Tree (SR Gill, Ed.). *Microbiology Resource Announcements* **10**.
- Teixeira M, Martins L, Fernandes C *et al.*, 2020. Complete Genome Sequences of Walnut-Associated *Xanthomonas euroxanthea* Strains CPBF 367 and CPBF 426 Obtained by Illumina/Nanopore Hybrid Assembly (DA Baltrus, Ed.). *Microbiology Resource Announcements* **9**.
- Vauterin L, Hoste B, Kersters K, Swings J, 1995. Reclassification of *Xanthomonas*. *International Journal of Systematic Bacteriology* **45**, 472–489.
- Vauterin L, Rademaker J, Swings J, 2000. Synopsis on the taxonomy of the genus

Xanthomonas. *Phytopathology* **90**, 677–682.

- Vorhölter FJ, Schneiker S, Goesmann A *et al.*, 2008. The genome of *Xanthomonas campestris* pv. *campestris* B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis. *Journal of Biotechnology* **134**, 33–45.
- Waterhouse RM, Seppey M, Simao FA *et al.*, 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* **35**, 543–548.
- Wick RR, Judd LM, Gorrie CL, Holt KE, 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads (AM Phillippy, Ed.). *PLOS Computational Biology* **13**, e1005595.
- Young JM, Dye DW, Bradbury JF, Panagopoulos CG, Robbs CF, 1978. A proposed nomenclature and classification for plant pathogenic bacteria. *Journal of Agricultural Research* **21**, 153–177.
- Yu G, Smith DK, Zhu H, Guan Y, Lam TT, 2017. `ggtree` : an `r` package for visualization and annotation of phylogenetic trees with their covariates and other associated data (G McInerny, Ed.). *Methods in Ecology and Evolution* **8**, 28–36.

Supplementary material

The following are available online at <https://www.mdpi.com/2076-2607/9/3/624/s1>, Figure III.S1. (a) Scheme representing presence/absence of type 2 secretion system (T2SS) putative homologs in 44 *Xanthomonas* spp. genomes; (b) BLAST identity (%) and length (i.e., % of query coverage) for type 2 secretion system (T2SS) putative homologs in CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427, Figure III.S2. (a) Scheme representing presence/absence of type 4 pilus (T4P) putative homologs in 44 *Xanthomonas* spp. genomes; (b) BLAST identity (%) and length (i.e., % of query coverage) for type 4 pilus (T4P) putative homologs in CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427, Figure III.S3. (a) Scheme representing presence/absence of non-fimbrial adhesin putative homologs in 44 *Xanthomonas* spp. genomes; (b) BLAST identity (%) and length (i.e., % of query coverage) of non-fimbrial adhesins putative homologs in CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427, Figure III.S4. (a) Scheme representing presence/absence of chemotaxis-related protein putative homologs in 44 *Xanthomonas* spp. genomes; (b) BLAST identity (%) and length (i.e., % of query coverage) of chemotaxis-related protein putative homologs in CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427, Figure III.S5. (a) Scheme representing presence/absence of extracellular enzyme putative homologs in 44 *Xanthomonas* spp. genomes; (b) BLAST identity (%) and length (i.e., % of query coverage) of extracellular enzyme putative homologs in CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427, Figure III.S6. (a) BLAST identity (%) and length (i.e., % of query coverage) of type 3 secretion system (T3SS) putative homologs in CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427; Figure III.S6. (b) BLAST identity (%) and length (i.e., % of query coverage) of type 3 effector (T3E) putative homologs in CPBF 367, CPBF 424^T, CPBF 426, and CPBF 427. Table III.S1: List of *Xanthomonas* spp. genomes used in this study, Table III.S2: Average nucleotide identity results for the genomes of the 44 strains analyzed in this study, Table III.S3: Best BLAST hit results of putative homologs.



Chapter IV

Comparative Genomics of
Xanthomonas arboricola and
X. euroxanthea: two
phytopathogens, same
plant host species, similar
pathoadaptations

Chapter IV include the following publication:

Martins L, Teixeira M, Fernandes C, Pothier JF, Koebnik R, Fonseca NA, Tavares F. Genomic features of *Xanthomonas arboricola* and *X. euroxanthea* sharing the same plant host species preferences (walnut, pecan and tomato) seems unbiased by the host species. *In preparation*

Part of the results from this chapter were presented at:

Martins L, Teixeira M, Fernandes C, Chaves C, Pinto J, Fonseca NA, Tavares F. *X. arboricola* and *X. euroxanthea* strains isolated from pecan trees: comparative genomics and pathogenicity on walnut. 4th Annual EuroXanth Conference- University of Belgrade, Faculty of Agriculture, Belgrade, Serbia, 28-30 June 2021 (Oral presentation)

Genomics of *Xanthomonas arboricola* and *X. euroxantha*: two phytopathogens, same plant host species, similar pathoadaptations

Abstract

Xanthomonas euroxantha is a walnut-associated bacterium, previously isolated from the same diseased tree colonized by *Xanthomonas arboricola* pv. *juglandis* strains, an aggressive walnut tree pathogen. The diversity within *X. arboricola* species is intriguing, with its epidemic population structure subdivided into three groups, which two of them include highly recombinant, non-pathogenic, or less virulent strains isolated from different hosts, including pecan and tomato. In this study the genomes of four strains isolated from diseased pecan trees showing WBB-like symptoms were sequenced. Average Nucleotide Identity of these genomes and an assortment of available *Xanthomonas* sp. genomes revealed six new *X. euroxantha* strains, two isolated from pecan tree and four from tomato. Given the host diversity and niche sharing of *X. arboricola* and *X. euroxantha*, comparative genomics analysis was carried out between *X. arboricola* and *X. euroxantha* strains isolated together from walnut, pecan and tomato for assessing possible variation in genomic traits, with special focus on T3SS and T3E. No major differences were detected, regarding general genomic features, apart from genome size, and indeed the only differences coincided with regions of predicted genomic islands (considering walnut and pecan isolated strains). Although differing in pathogenicity potential and hosts, some *X. euroxantha* strains show a similar T3E profile, suggesting a non-host related pattern. Finally, this work could contribute to shed some light on the lifestyle behaviour of *X. euroxantha* species and synergistic effect with *X. arboricola*.

1. Introduction

Xanthomonas arboricola is a bacterial plant pathogen of numerous plant species, namely stone fruit trees (peach, apricot, and plum), nut trees (almond, hazelnut, and walnut), fruit trees (banana), ornamental (niger, arum lily, and poplar) and herbaceous plants (arracacha and strawberry with nine pathovars currently acknowledged: *pruni*, *corylina*, *juglans*, *celebensis*, *guizotiae*, *zantedeschiae*, *populi*, *arracaciae*, *fragariae* (Kałużna et al., 2021; Lamichhane, 2014). Particularly, the pathovar *juglandis* is the causal agent of three walnut diseases, namely bacterial blight (Vauterin et al., 1995), brown apical

necrosis (Moragrega et al., 2011) and vertical oozing canker (Hajri et al., 2010) and can be responsible for decreases in production of about 70% (Lamichhane, 2014). It infects *Juglans* sp., predominantly the *Juglans regia* species, also known as Persian (English) walnut (Kałużna et al., 2021). *X. arboricola* species are characterized by displaying high genetic diversity and extensive host range, with an epidemics-scale population structure segmented into three groups: epidemic clones- limited to *juglandis*, *corylina* and *pruni* pathovars (group A); highly recombinant strains (group B) and highly recombinant divergent lineages (group C), which also included the walnut-associated species *X. euroxanthea*, *X. arboricola* pv. *populi* and *X. arboricola* strains isolated from different hosts. While group A is composed exclusively by pathogenic/competent pathovars linked by a monophyletic origin, groups B and C encompass generalist strains isolated from a heterogeneous assortment of tree and herbaceous plant species of different pathogenicity potential (Kałużna et al., 2021; Merda et al., 2016). In fact, these strains can range from non-pathogenic to weakly pathogenic. Particularly, while *X. arboricola* strains isolated from tomato (BRIP 62410, BRIP 62412, BRIP 62414, BRIP 62416 and BRIP 62432) are nonpathogenic on their host of isolation (Roach et al., 2020), *X. arboricola* strains CPBF 765 and CPBF 1494 were recently reported for the first time on *Carya illinoensis* (pecan nut) evidencing symptoms of bacterial blight similar to walnut bacterial blight in *Juglans regia* (Fernandes et al., 2018). It's also worth mentioning the co-occurrence of members from two of these distinct groups with the finding of *X. arboricola* pv. *juglandis* (group A) and walnut-associated species, *X. euroxanthea* (group C) in the same tree during an extensive assessment of the diversity of *X. arboricola* pv. *juglandis* in walnut trees, recently carried out in Portugal (Fernandes, Albuquerque, et al., 2021; Fernandes, Martins, et al., 2021; Martins et al., 2020).

Carya illinoensis stands out as the crop with the highest economically importance within the *Carya* genus (Thompson & Conner, 2012), being susceptible to a great number of diseases, including insect and mites pests, that are altogether responsible for harvest and economical losses. However, according to EPPO Global database (accessed July 2022), until now no member of *Xanthomonas* species was described as Pecan pests/pathogens. This corroborates the cosmopolitan distribution of this species and points to an underestimation of pathogenic potential or a rapid host adaptability of *X. arboricola* which may lead to a reassessment of current *pathovar* denominations. Additionally, a correction to the current simplistic concept of host appears to be necessary. Indeed, a thorough review on the concept of host range by (Morris & Moury, 2019) proposed an update from the current notion to obtain a broader concept and take into consideration potential pathogenic adaptations. For clarity, this study considers

nonpathogenic strains those that are not able to develop disease symptoms when inoculated into the host they were isolated from (Merda et al., 2016).

The type III secretion system is an essential element *Xanthomonas* pathogens' arsenal to overcome host plant defenses and lead to infection. Interestingly, *X. euroxanthea* strain isolated from walnut CPBF 424^T include a reduced number of type III effectors comparatively to *X. arboricola* pv. *juglandis* CPBF 427, isolated from the same tree specimen, while harboring a complete type III secretion system (Fernandes, Martins, et al., 2021) Considering the discovery of *X. euroxanthea* strains on other hosts, it will be interesting to see if these exhibit the same genetic profile discern distinctive host- specific adaptations.

This study proposed six new strains belonging to *X. euroxanthea* species, from two new hosts, tomato and pecan. In the light of *Xanthomonas* diversity and bacterial speciation, the study delivered a comprehensive genomic characterization of *X. arboricola* and *X. euroxanthea* strains recovered from three different host species, *Juglans regia*, *Carya illinoensis* and *Solanum lycopersicum* in order to elucidate on host-specific adaptations of these two species and the tendency to often be found co-colonizing the same host species.

2. Methods

2.1. Genome sequencing

The genomes of four xanthomonads (CPBF 761, CPBF 765, CPBF 766 and CPBF 1494) isolated from a *Carya illinoensis* tree specimen showing symptoms of bacterial blight (Fernandes et al., 2018) were sequenced within the framework of this study (Table IV.1). Bacteria culturing, DNA extraction, library preparation, genome sequencing and annotation were carried out as previously described (Teixeira et al., 2020, 2021).

2.2. *Xanthomonas* genomes used in this study and taxonomic affiliation determined by the Average Nucleotide Identity

To determine the taxonomic affiliation, the genetic distances between *X. euroxanthea* strains *Xanthomonas* strains for which the species remained undetermined, or have been identified as atypical *X. arboricola* were estimated by the Average nucleotide identity (ANI) based on BLASTn applying OrthoANI v1.40 (Lee et al., 2016). The analysis was extended to 51 genomes of *Xanthomonas* spp. available from NCBI genome database, including 11 different *Xanthomonas* species with five *X. euroxanthea* representative strains and eight distinct pathovars (Supplementary Table IV.S1)

Table IV.1. General genomic features

Strain	Species	Host of isolation	Genome size (Mbp)	Contigs	CDSs	Genes	Source of genomes ¹
CPBF 424^T	<i>X. euroxanthea</i>	<i>Juglans regia</i>	4.90	1	4035	4114	Fernandes <i>et al.</i> , 2021
CPBF 367	<i>X. euroxanthea</i>	<i>Juglans regia</i>	4.97	2	4089	4169	Teixeira <i>et al.</i> , 2020
CPBF 426	<i>X. euroxanthea</i>	<i>Juglans regia</i>	4.90	2	4066	4143	Teixeira <i>et al.</i> , 2020
CPBF 427	<i>X. arboricola</i>	<i>Juglans regia</i>	5.23	1	4367	4465	Teixeira <i>et al.</i> , 2021
CPBF 761	<i>X. euroxanthea</i>	<i>Carya illinoensis</i>	4.85	1	4015	4092	This study
CPBF 766	<i>X. euroxanthea</i>	<i>Carya illinoensis</i>	4.81	1	4013	4092	This study
CPBF 765	<i>X. arboricola</i>	<i>Carya illinoensis</i>	5.11	1	4294	4388	This study
CPBF 1494	<i>X. arboricola</i>	<i>Carya illinoensis</i>	5.00	1	4205	4296	This study
BRIP62409	<i>X. euroxanthea</i>	<i>Solanum lycopersicum</i>	4.92	37	4052	4129	Roach <i>et al.</i> , 2019
BRIP62411	<i>X. euroxanthea</i>	<i>Solanum lycopersicum</i>	4.81	46	4047	4122	Roach <i>et al.</i> , 2019
BRIP62415	<i>X. euroxanthea</i>	<i>Solanum lycopersicum</i>	4.85	49	4032	4108	Roach <i>et al.</i> , 2019
BRIP62418	<i>X. euroxanthea</i>	<i>Solanum lycopersicum</i>	5.09	67	4318	4393	Roach <i>et al.</i> , 2019
BRIP62410	<i>X. arboricola</i>	<i>Solanum lycopersicum</i>	4.90	36	4123	4204	Roach <i>et al.</i> , 2019
BRIP62412	<i>X. arboricola</i>	<i>Solanum lycopersicum</i>	5.14	53	4362	4452	Roach <i>et al.</i> , 2019
BRIP62416	<i>X. arboricola</i>	<i>Solanum lycopersicum</i>	5.19	63	4324	4416	Roach <i>et al.</i> , 2019
BRIP62432	<i>X. arboricola</i>	<i>Solanum lycopersicum</i>	4.95	49	4110	4201	Roach <i>et al.</i> , 2019

2.3. Phylogenetic tree of core genome

A total of 1149 single copy orthologous genes present in a total of 51 *Xanthomonas* genomes (Supplementary Table IV.S1) were extracted from BUSCO (v. 4.0.6, database xanthomonadales_odb10) (Simão et al., 2015; Waterhouse et al., 2018) and sequences were aligned using TCOFFEE (v11.00.8cbe486, mcoffee mode) (Notredame et al., 2000). All single copy orthologous were used to infer a maximum likelihood phylogenetic tree, using RAXML (raxmlGUI 2.0, model GTR+FO) (Edler et al., 2021; Stamatakis, 2006) with 100 bootstrap replicates. The corresponding phylogram was visually represented with the R package “ggtree” (Yu et al., 2017).

2.4. General genomic characteristics

Towards performing a reliable comparative analysis of genomic features from *X. arboricola* and *X. euroxanthea* strains isolated from distinct plant host species (*Juglans regia* (walnut), *Carya illinoensis* (pecan) and *Solanum lycopersicum* (tomato)), the genomes were re-annotated (Table IV.1) using the NCBI Prokaryotic Genome Annotation Pipeline (v. 2021-07-01.build5508) (Tatusova et al., 2016). Whole genomic comparisons between genomes of *X. euroxanthea* (*Xea*) and *X. arboricola* (*Xa*) strains from walnut (*Xea* CPBF 367, *Xaj* CPBF 427) and pecan (*Xea* CPBF 761, CPBF 766, *Xa* CPBF 765, CPBF 1494) were carried out and visualized through a circular plot with Edgar 3.0 (Dieckmann et al., 2021), using *X. euroxanthea* type strain (CPBF 424^T) as reference. The genome of CPBF 424^T was also inspected for genomic islands, using IslandViewer4 (Bertelli et al., 2017).

2.5. Pathogenicity assays and hypersensitive response (HR)

The pathogenicity assays on *Carya illinoensis* and *Juglans regia* ‘Hartley’ were carried out as described previously (Fernandes et al., 2018). Briefly, three plantlets with at least four fully expanded leaves were sprayed with bacterial suspension of 10⁸ CFU/ml. Plantlets were maintained under a 16h photoperiod (24°C day and 18°C night). For the hypersensitivity reaction assay, bacterial suspensions of OD₆₀₀=0.2 were infiltrated into *Nicotiana benthamiana* leaves using a needleless syringe. HR symptoms were registered 72h after inoculation. For both assays, strain *X. arboricola* pv. *juglandis* LMG 747^T was used as positive control and water as a negative control.

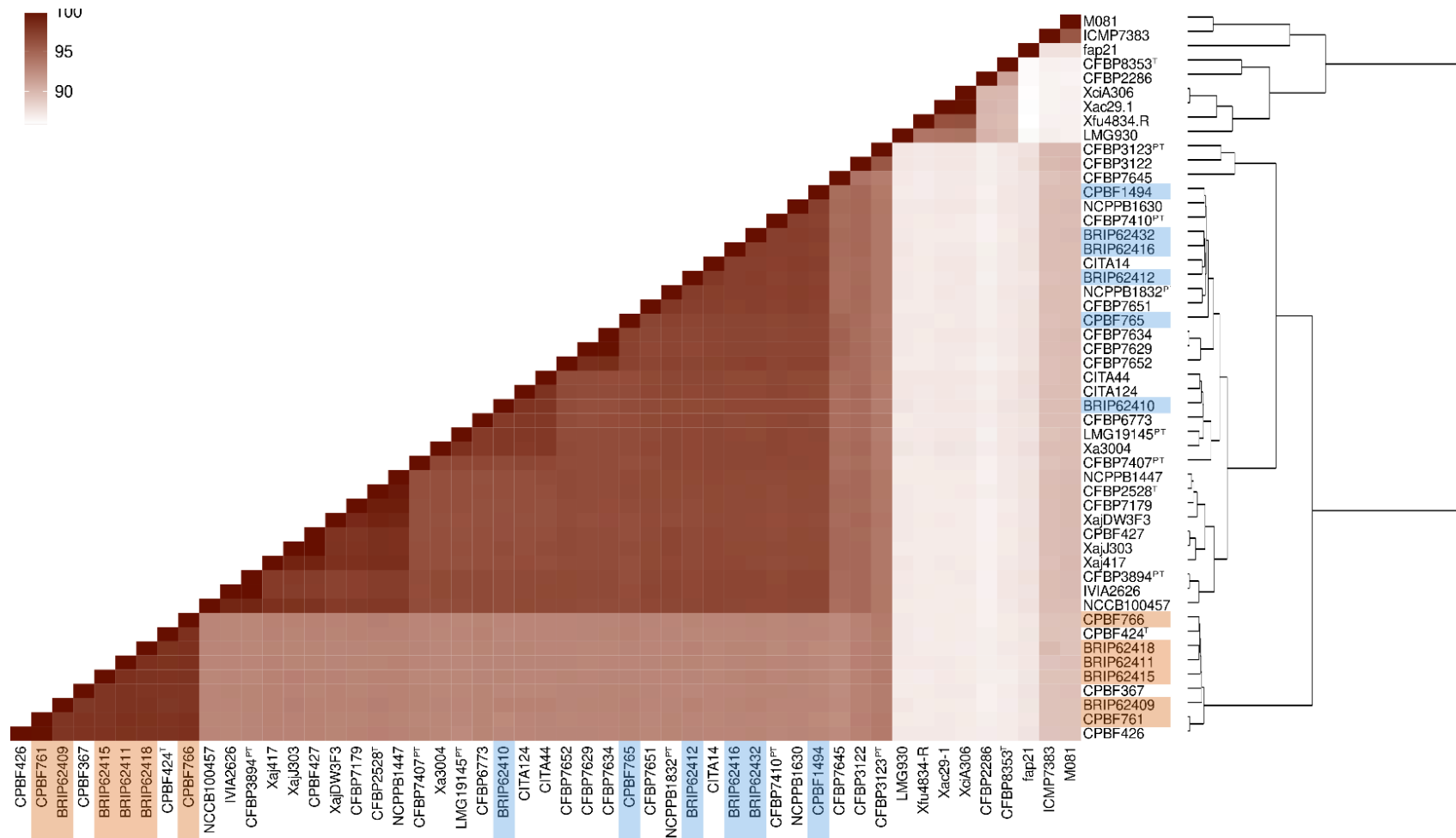


Figure IV.1. Heatmap representing average nucleotide identity (ANI) (disclosed in Supplementary Table IV.S1) and the respective distance tree cladogram determined for 51 *Xanthomonas* genomes. *Carya illinoensis* and *Solanum lycopersicum* isolates with *X. euroxanthea* strains are highlighted in blue, while isolates with ANI values >95% for *X. arboricola* are highlighted in orange. The colour scale ranging from white to dark red depicts lower to higher similarity values, respectively.

2.6. Pathogenicity and virulence factors prediction

The presence of effector/virulence protein homologs in the genomes of *X. arboricola* and *X. euroxanthea* isolated from the three different host plants was identified using reciprocal tBLASTn (tBLASTn v. 2.10.1 (Altschul et al., 1990) with cut-offs of $\geq 40\%$ identity and $\geq 75\%$ query sequence length, performed against a database of protein sequences previously used for *X. arboricola* (Fernandes, Martins, et al., 2021) also considering the *Xanthomonas* resource type III effectors database (<http://xanthomonas.org/t3e.html>) and genes that were previously automatically annotated (Supplementary Table IV.S2). Furthermore, Easyfig (Sullivan et al., 2011) was used to compare and represent the type 3 secretion system gene cluster of *X. euroxanthea* (CPBF 424^T and CPBF 766) and *X. arboricola* (CPBF 765 and CPBF 427) from two different hosts.

3. Results

3.1. *Xanthomonas* sp. isolates are evidenced as belonging to *X. euroxanthea* species

Taxonomic affiliation of uncharacterized *Xanthomonas* and *X. euroxanthea* strains CPBF 367, CPBF 424^T and CPBF 426 carried out through ANI revealed that two of the Portuguese strains CPBF 1494 and CPBF 765, isolated from *Carya illinoensis*, belong to the *Xanthomonas arboricola* group (Figure IV.1) (supported by the $>95\%$ ANI threshold that recognizes strains as the same species (Richter & Rosselló-Móra, 2009). Meanwhile the other two strains isolated from *Carya illinoensis*, CPBF 761 and CPBF 766, are included in the *Xanthomonas euroxanthea* group, (ANI value of 97.9%) with previously described *X. euroxanthea* strains (Martins et al., 2020). Additionally, four *Xanthomonas* strains (BRIP 62409, BRIP 62411, BRIP 62415, BRIP 62418) isolated from tomato in Stanthorpe in 2015 (Roach et al., 2018) reveal ANI values of $> 97.7\%$ with *X. euroxanthea* strains, strongly indicating their taxonomic affiliation to *X. euroxanthea* species

Genomic features of *Xanthomonas arboricola* and *X. euroxanthea* sharing the same plant host preferences (walnut pecan and tomato) seems unbiased by the host species

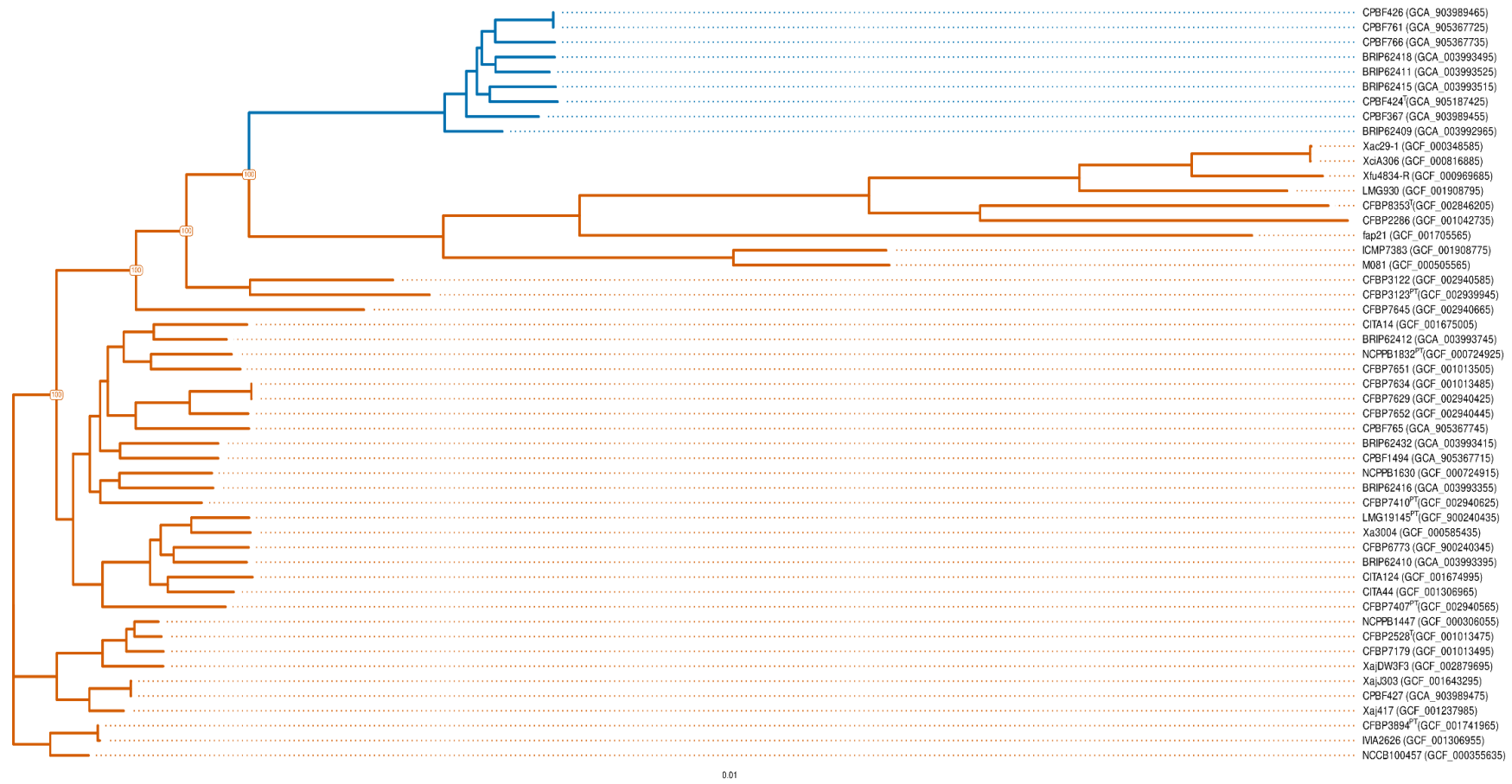


Figure IV.2. Phylogenomic maximum likelihood tree using RaxML from concatenated sequences of 1149 core genes from 51 genomes including the *Xanthomonas euroxanthea* strains (branches in blue) and *Xanthomonas arboricola* and *Xanthomonas* sp. strains (branches in orange). The phylogram was constructed with the R package "ggtree" with supporting values from 100 bootstrap replicates indicated near nodes.

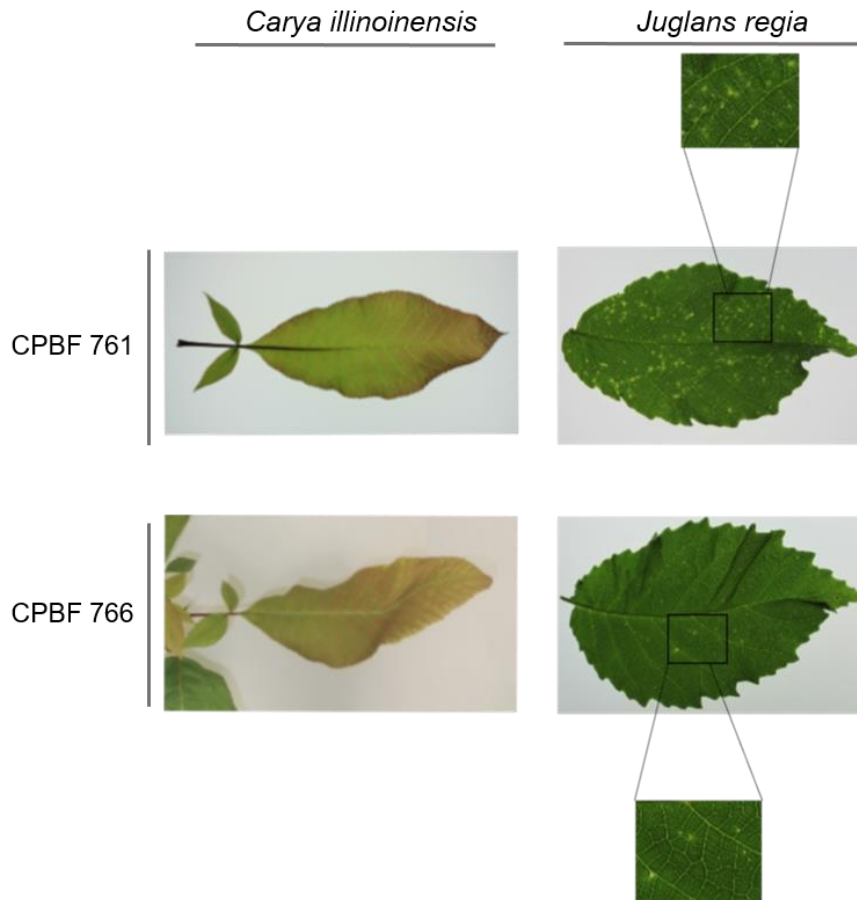
3.2. Phylogenomics tree denotes a cluster with *X. euroxantha* strains isolated from three different hosts

The ANI-confirmed *X. euroxantha* strains isolated from *Carya illinoensis* (CPBF 761, CPBF 766) and from tomato (BRIP 62409, BRIP 624011, BRIP 62415, BRIP 62418) group in a restricted cluster with *X. euroxantha* strains found in *Juglans regia* (CPBF 367, CPBF 424^T and CPBF 426). Meanwhile, the other strains isolated from *Carya illinoensis* (CPBF 765 and CPBF 1494) cluster with the other *X. arboricola* strains including the ones isolated from *Solanum lycopersicum* (BRIP 62410, BRIP 62412, BRIP 62416, BRIP 62432) (Figure IV.2).

3.3. *X. euroxantha* CPBF 761 and CPBF 766 are pathogenic on walnut and nonpathogenic on pecan

Pathogenicity assays revealed that *X. euroxantha* strains CPBF 761 and CPBF 766 caused small leaf lesions of necrotic spots with chlorotic halos on *Juglans regia*, while no symptoms were registered on *Carya illinoensis* (Figure IV.3). Meanwhile, both *X. euroxantha* (CPBF 761 and CPBF 766) and *X. arboricola* (CPBF 765 and CPBF 1494) were positive for hypersensitive response in *Nicotiana benthamiana*.

(A)



(B)

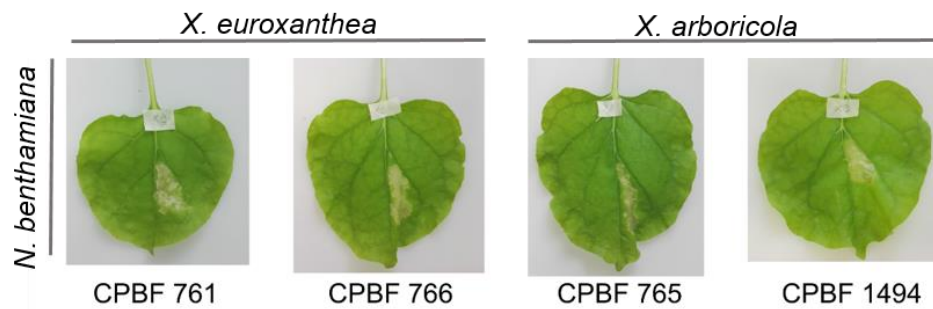


Figure IV.3. (A) Pathogenicity tests results four weeks after inoculation. No symptoms were observed on pecan leaflets for CPBF 761 and CPBF 766 (on the left), while small necrotic spots with chlorotic halos were produced by CPBF 761 and CPBF 766 on walnut (on the right); (B) Hypersensitivity response symptoms on leaves of *Nicotiana benthamiana* after 72 h after inoculation with the *X. euroxanthea* strains CPBF 761, CPBF 766 and *X. arboricola* CPBF 765 and CPBF 1494 (from left to right).

3.4. General genomic characteristics of *X. euroxantha* and *X. arboricola* from different hosts

The analysis of general genomic features of the *X. euroxantha* and *X. arboricola* genomes from three different hosts denotes great level of similarity between them, only noticing that within walnut and pecan isolated strains, *X. arboricola* genomes are slightly bigger than *X. euroxantha* (Table IV.1). Concordantly, CDSs and gene number of the above-mentioned genomes follows the same trend. Regarding the genomes of tomato-isolated strains, there are no major differences in genome size, CDSs and gene number between the two bacterial species. The fact that genomes of strains isolated from *Solanum lycopersicum* are scattered throughout a great number of contigs (from 36 to 67 contigs per genome) precluded their use in some analyses due to the difficulty of working with them for some of the tools used.

The overall whole genome comparative analysis of *X. arboricola* and *X. euroxantha* strains from walnut and pecan was showcased in a circular plot (Figure IV.4A) to depict major genomic regions of variation that could be specific to bacterial species and/or host. No differences were detected related to each species in particular or host of isolation, but it's interesting to notice that the breaks in core genome content between the seven analyzed strains seem to match to genomic areas where genomic islands were predicted for CPBF 424^T (Figure IV.4B). The type III secretion cluster does not seem to be embedded in any genomic island.

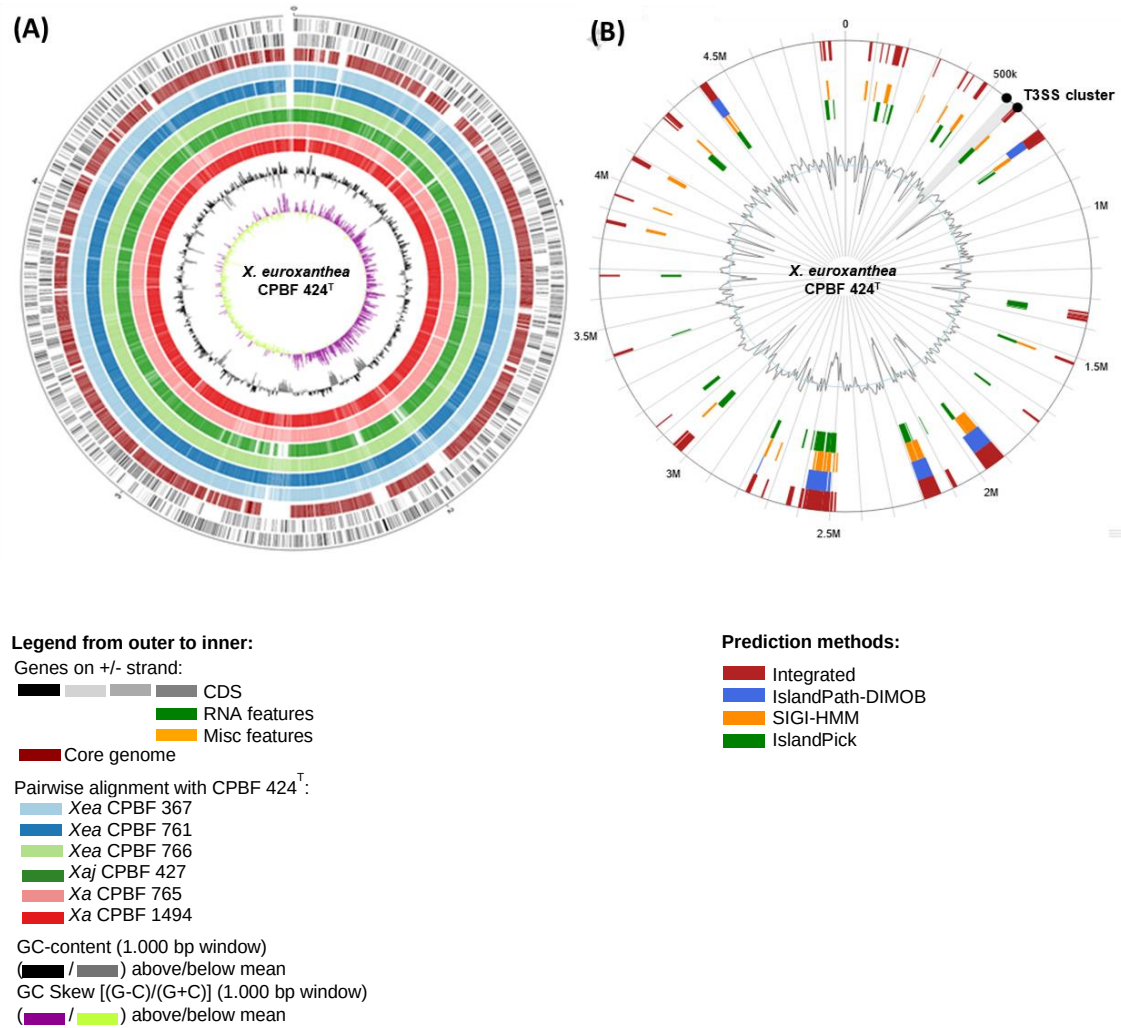


Figure IV.4 (A) Circular plot for genome comparison of *X. euroxanthae* (Xea) and *X. arboricola* (Xa) strains from two different hosts, walnut (Xea CPBF 367, Xaj CPBF 427) and pecan (Xea CPBF 761, CPBF 766, Xa CPBF 765, CPBF 1494), using *X. euroxanthae* CPBF 424^T as the reference genome. The analysis and plot were carried out with EDGAR 3.0. (B) Genomic island prediction of *X. euroxanthae* CPBF 424^T using IslandViewer4, which integrates data from three tools: IslandPath-DIMOB, SIGI-HMM and IslandPick. The T3SS of CPBF 424^T is highlighted in light grey (560320-592275bp)

3.5. *X. euroxanthae* type III effector (T3E) profiles follow a non-host related pattern

The *X. euroxanthae* strain CPBF 766 isolated from pecan, CPBF 424^T from walnut and BRIP62409, BRIP 62411, BRIP 62415, BRIP 62418 from tomato harbor a similar arsenal of T3 effectors, mainly composed of XopAZ, XopR, AvrXccA1.1, XopF1, XopF2, XopM, XopA/hpa1 and XopZ2, with the absence of AvrXccA1.1 in BRIP 62409 (Figure IV.5). Strikingly, XopZ2 is exclusive to these *X. euroxanthae* strains. And distinctively, *X. euroxanthae* CPBF 761 from pecan and CPBF 367 and CPBF 426 from walnut lack the majority of effectors, possessing only XopAZ, XopR and AvrXccA1.1.

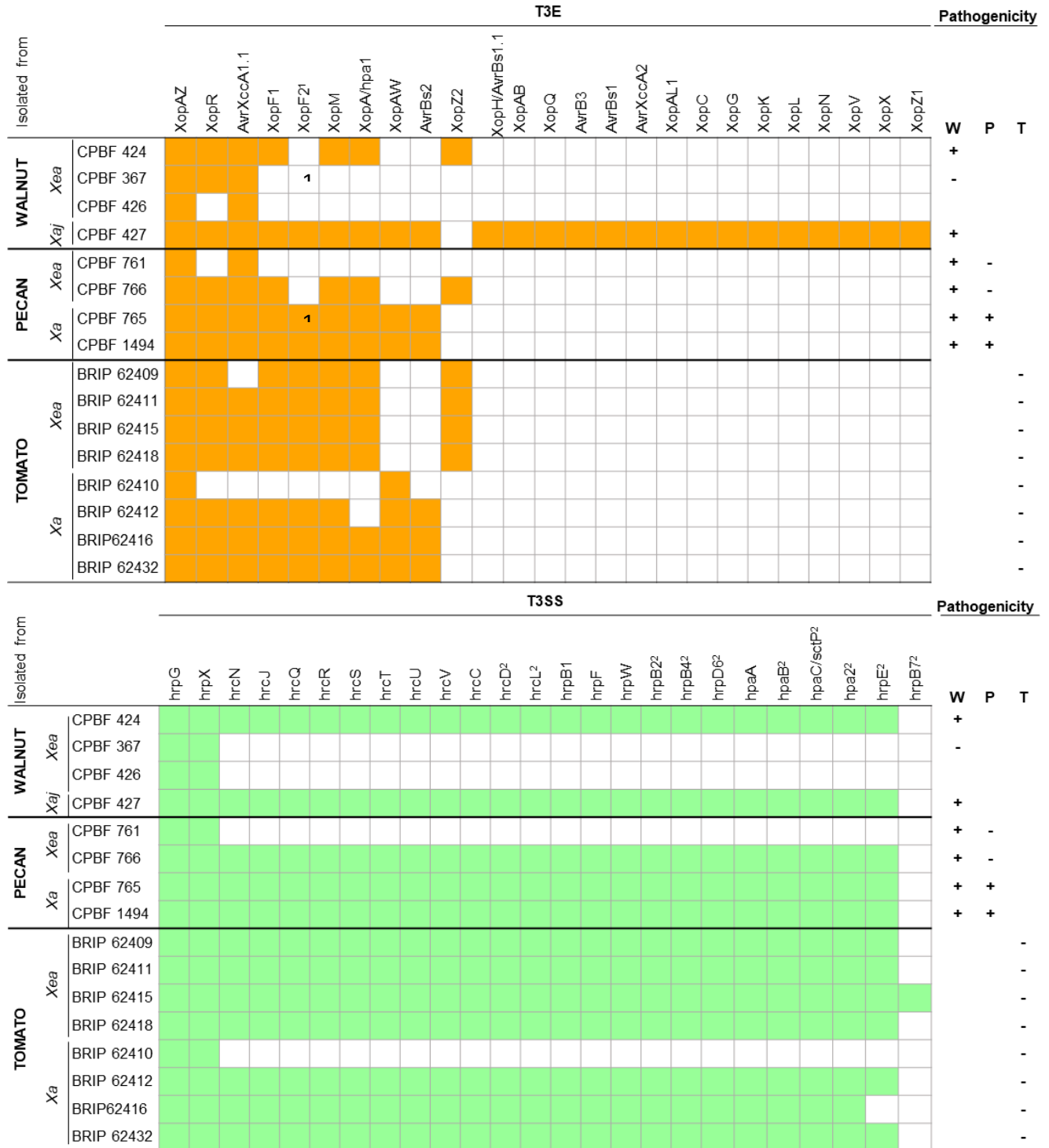


Figure IV.5. Dot plot scheme represents the presence/absence scheme for type 3 secretion system (T3SS) and effectors (T3E) putative homologs; ■ present; □ not present; considering a tBLASTn hit with a query coverage threshold $\geq 75\%$, and sequence identity with $\geq 40\%$ % cut-off. Pathogenicity tests results in walnut (W), pecan (P) and tomato (T) are displayed on the right (this study, Martins et al. 2020, Roach et al. 2020). ¹Coverage and identity values are borderline and within the defined threshold, respectively; ²genes automatically annotated in CPBF 424^T genome that were confirmed by blast using the threshold of $\geq 40\%$ identity and $\geq 75\%$ query sequence length.

Overall, it is possible to denote two distinct profiles of T3E in *X. euroxantha*: the “virulent” profile composed of XopAZ, XopR, AvrXccA1.1, XopF1, XopF2, XopM, XopA and XopZ2 and the “non-virulent” that encompass XopR, XopAZ and AvrXccA1.1. *X. arboricola* strains CPBF 765 and CPBF 1494 isolated from pecan and BRIP 62412, BRIP 62416, BRIP 62432 from tomato display a T3E profile similar to pathogenic *X. euroxantha* but additionally contain effectors XopAW and AvrBs2, which are exclusively present in the *X. arboricola* analyzed. Remarkably, these strains lack *xopH/avrBs1.1*, *xopAB*, *xopQ*, *avrB3*, *avrBs1*, *avrXccA2*, *xopAL1*, *xopC*, *xopG*, *xopK*, *xopL*, *xopN*, *xopV*, *xopX* and *xopZ1*, that are only present in one *X. arboricola* strain from walnut, CPBF 427. *X. arboricola* BRIP 62410 lacks the majority of effectors. Type III effectors *xopAA*, *xopAC*, *xopAD*, *xopAE* (*hpaF*), *xopAF1* (*hopAF1*), *xopAF2* (*xopAF*), *xopAG* (*avrGf1*), *xopAH*, *xopAI*, *xopAJ*, *xopAK*, *xopAL2*, *xopAM*, *xopAP*, *xopAU*, *xopAV*, *xopAX*, *xopAY*, *xopB*, *xopC2*, *xopD*, *xopE1*, *xopE2*, *xopE3*, *xopl*, *xopJ*, *xopO*, *xopP*, *xopS*, *xopT*, *xopU* and *xopW* were absent on all *X. arboricola* and *X. euroxantha* strains analyzed (data not shown/ not included in the image).

Accordingly, every strain has a complete type III secretion system (T3SS), encoding all T3SS proteins, with the exception of strains that are depleted of most T3 effectors, which are *X. euroxantha* CPBF 761, CPBF 367, CPBF 426 and *X. arboricola* BRIP 62410. Comparing the T3SS cluster structural organization of *X. euroxantha* and *X. arboricola* from walnut and pecan it is evidenced that *X. euroxantha* strains CPBF 424^T and CPBF 766 from different hosts (walnut and pecan, respectively) are completely syntenic (Figure IV.6). Comparing with *X. arboricola*, *xopZ2* is absent in CPBF 765 and in CPBF 427, as well as a gene downstream of *hrpF* in *X. euroxantha* (not considered as T3E/T3SS), different from one in the same location present in the *X. arboricola* strains. Moreover, the cluster of CPBF 427 includes a gene that is predicted to belong to the IS5 family (data not shown).

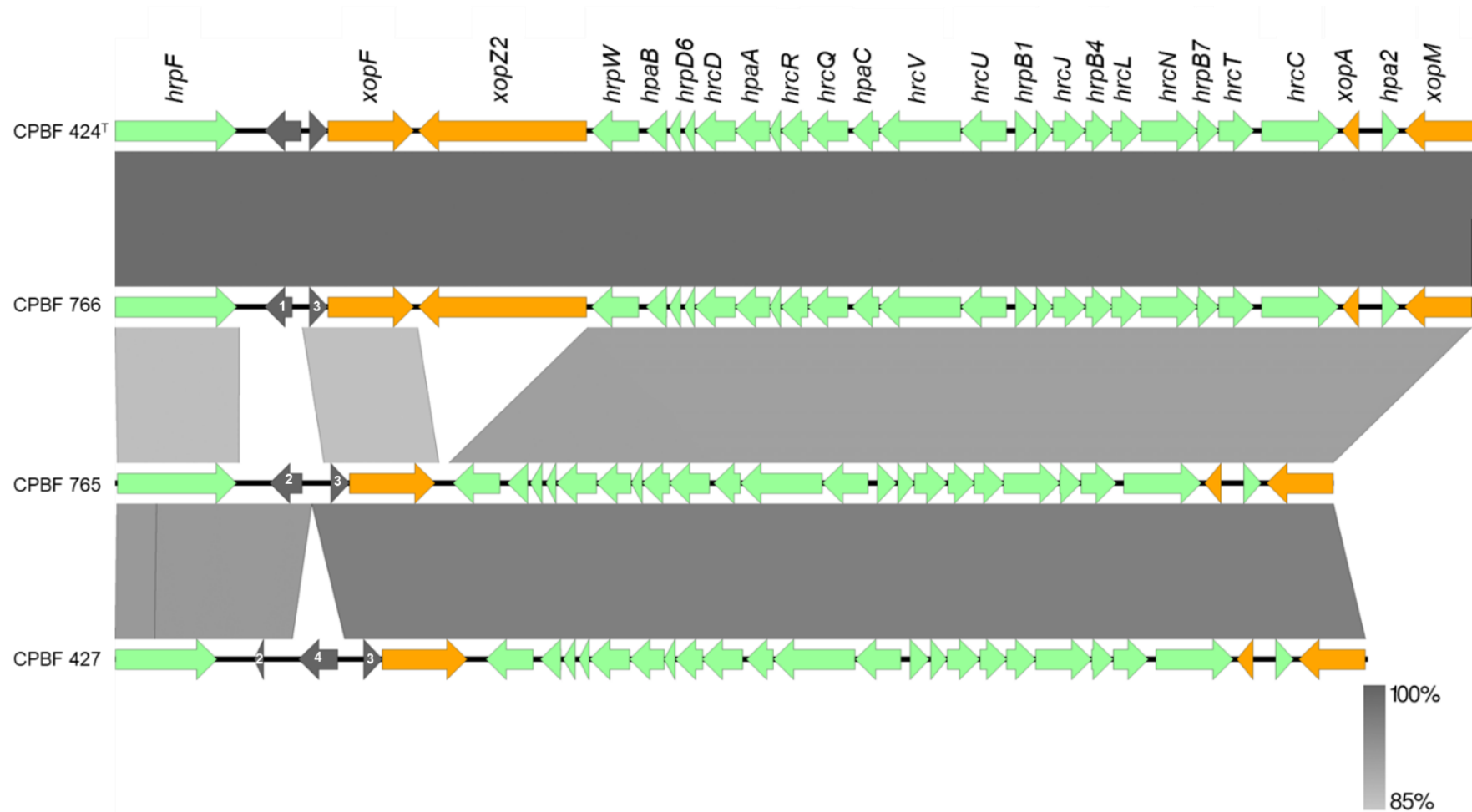


Figure IV.6 Synteny analysis of genes included in type 3 secretion system (T3SS) cluster of *X. euroxanthea* strains CPBF 424^T and CPBF 766 and *X. arboricola* CPBF 765 and CPBF 427 using Easyfig. T3SS and T3E genes and depicted as green and orange arrows, respectively. ^{1,2} not characterized as effectors, multiple sequence alignments and blast results of protein sequences corresponding to each strain are disclosed in Supplementary Figure IV.S1; ¹accession numbers XTG000468 (CPBF 424^T), XCY000470 (CPBF 766); ²accession numbers XCY000604 (CPBF 765), XSP003898 (CPBF 427); amino acid sequence alignment between ¹ and ² is shown in Supplementary Figure IV.S1; ³accession numbers XTG000469 (CPBF 424^T), XCY000471 (CPBF 766), XCY000605 (CPBF 765), XSP003896 (CPBF 427); ⁴XSP003897 is only present in CPBF 427.

4. Discussion

The co-colonization between *X. arboricola* and *X. euroxanthea* has been previously observed on *Juglans regia*, with the report of three *X. euroxanthea* strains CPBF 367, CPBF 424^T, CPBF 426 and *X. arboricola* pv. *juglandis* CPBF 427 and CPBF 1521 within the same tree (Fernandes, Albuquerque, et al., 2021; Fernandes, Martins, et al., 2021). Correspondingly, the present study attributed *Carya illinoensis*-isolated strains CPBF 761 and CPBF 766 to *X. euroxanthea* and CPBF 765 and CPBF 1494 to *X. arboricola*, relying on ANI calculations. Taking into consideration that CPBF 1494 and CPBF 761 strains were isolated from the same *Carya illinoensis* tree and CPBF 765 and CPBF 766 were isolated together from a different *Carya illinoensis* tree (data not shown), it appears that *X. arboricola* and *X. euroxanthea* occur sympatrically within pecan trees, as it was also previously observed in walnut (Fernandes, Albuquerque, et al., 2021; Fernandes, Martins, et al., 2021). Moreover, *Xanthomonas* sp. strains BRIP 62409, BRIP 624011, BRIP 62415, BRIP 62418, also indicated as *X. euroxanthea* by ANI in this study, were isolated from tomato plants on the same region and in the same year as *X. arboricola* strains BRIP 62410, BRIP 62412, BRIP 62416, BRIP 62432 (Roach et al., 2018). Additionally, strain CFBP 7622, whose taxonomic classification was recently corrected from *X. arboricola* to *X. euroxanthea* (Zarei et al., 2022) was isolated from *Phaseolus vulgaris*, a plant where *X. arboricola* strains were also retrieved from (CFBP 8152, CFBP 8150, CFBP 8130, CFBP 8153, CFBP 8138 CFBP 8139, CFBP 8132, CFBP 8147, CFBP 8142 CFBP 8149, CFBP 8140, CFBP 7697) (Kałużna et al., 2021). These observations suggest that the occurrence of these two *Xanthomonas* species in the same plant host species is hardly circumstantial and demands for further studies, namely by addressing the pathogenicity phenotype of these strains and the scrutiny of species-specific genetic determinants through comparative genomics to unveil distinct pathoadaptations.

Concerning pathogenicity, while the four strains isolated from *Carya illinoensis*, *X. euroxanthea* strains CPBF 761 and CPBF 766 and *X. arboricola* strains CPBF 765 and CPBF 1494, were HR-positive on *Nicotiana benthamiana* and pathogenic on walnut, only the *X. arboricola* strains CPBF 765 and CPBF 1494 were shown to be additionally pathogenic on pecan (Fernandes et al., 2018). This data suggests that these strains possess a distinct genetic composition regarding pathogenicity and virulence. Strains lacking a functional T3SS, *X. euroxanthea* CPBF 367, CPBF 426, CPBF 761 and *X. arboricola* BRIP 62410, also lack most of the T3 effectors, suggesting the loss of these effectors as a form of adaptation. Our data also reveal that the T3SS in *X. euroxanthea*

CPBF 424^T does not seem to be located on a region of genomic instability as it is not predicted to be included on a genomic island.

Interestingly, *X. arboricola* strains from pecan (CPBF 765 and CPBF 1494) and from tomato (BRIP 62412, BRIP 62416, BRIP 62432) have same set of effectors from the pathogenic *X. euroxantha* CPBF 424^T, with the additional *xopAW* and *avrBs2*. The presence of these two effectors in *X. arboricola* and absence in *X. euroxantha* was observed previously (Assis et al., 2021; Fernandes, Martins, et al., 2021). These *X. arboricola* strains (CPBF 765, CPBF 1494, BRIP 62412, BRIP 62416, BRIP 62432) lack the effectors *xopH/avrBs1.1*, *xopAB*, *xopQ*, *avrB3*, *avrBs1*, *avrXccA2*, *xopAL1*, *xopC*, *xopG*, *xopK*, *xopL*, *xopN*, *xopV*, *xopX*, *xopZ1*, that are present in *X. arboricola* pv. *juglandis* strain from walnut, CPBF 427. Such an extended profile of T3E is typically represented in *X. arboricola* successful pathovars *juglandis*, *pruni* and *corylina* (Essakhi et al., 2015; Hajri et al., 2011; Kałużna et al., 2021; Merda et al., 2016). Furthermore, they lack the T3E set composed of *XopL*, *XopN*, *XopQ*, *XopX* and *XopZ* that have been shown to be key in suppressing initial immune defence responses of the host (Merda et al., 2017; Sinha et al., 2013) and thus crucial for pathogenicity. Possibly, in the absence of this set of genes, other functionally-redundant effectors may target host defences similarly. In fact, *XopF2* of *X. euvesicatoria* and *XopR* of *Xanthomonas oryzae* pv. *oryzae* and *X. axonopodis* pv. *manihotis* that have been pointed out to have an important role as suppressors of host immunity (Medina et al., 2018; Popov et al., 2016; Verma et al., 2018, 2019) are present in all strains except BRIP 62410 and CPBF 426 and CPBF 761 (where is annotated as a pseudogene).

Type three effector profiles have been suggested to shape host specificity and particularity pathovar-related specificity, with this specialization attributed to additional effectors over the smaller profile displayed by generalists (Hajri et al., 2009; Jacques et al., 2016). With *XopZ2* being exclusive to virulent strains of *X. euroxantha* (CPBF 424^T, CPBF 766, BRIP 62409, BRIP 62411, BRIP 62415, BRIP 62418), this T3E could be an interesting marker for discerning pathogenic and non-pathogenic strains of *X. euroxantha*. The T3SS cluster organization placed *X. euroxantha*'s *xopZ2* between *xopF1* (upstream) and *hrpW* (downstream), as previously reported for *X. gardneri* and *X. vesicatoria*, both tomato pathogens (Potnis et al., 2011).

Overall, it is possible to designate two distinct profiles of T3E in *X. euroxantha*: the “virulent” profile composed of eight effectors, *xopAZ*, *xopR*, *avrXccA1.1*, *xopF1*, *xopF2*, *xopM*, *xopA* and *xopZ2* and the “non-virulent” group that encompass only three effectors, *xopR*, *xopAZ* and *avrXccA1.1*. The strains CPBF 424^T, CPBF 766, BRIP 62409, BRIP 62411, BRIP 62415, BRIP 62418 fit into “virulent” group and CPBF 367, CPBF 426,

CPBF 761 fit into “non-virulent” group. Furthermore, CFBP 7622 and CFBP 7653, recently reclassified as *X. euroxanthae* (Fernandes, Martins, et al., 2021; Zarei et al., 2022) seem to follow the “virulent” group and “non-virulent” patterns, respectively (Assis et al., 2021; Fernandes, Martins, et al., 2021; Merda et al., 2016), while considering that the effectors XopAZ, AvrXccA1.1, XopF2 and XopM were not screened for these two strains (Merda et al., 2016).

Some type III effector proteins are recognized by receptors in the host subsequently triggering plant defense responses (Deb et al., 2022). As an adaptation to bypass host recognition, it is possible that these effectors will be pressured to be lost, possibly waning overall pathogen fitness (Deb et al., 2022). Still, as the majority of type III effectors have redundant functions regarding virulence (Song & Yang, 2010), these rearrangements triggered by host-pathogen adaptation to pathogenicity/defense may eventually lead to alterations in host specificity, as it has been previously hypothesized that the variable arsenal of type three effectors displayed by different xanthomonads may be the driver unlocking host specificity (Sinha et al., 2013).

The fact that *X. euroxanthae* strains with distinct T3E profiles are found in distinct plant hosts *Juglans regia*, *Solanum lycopersicum* and *Carya illinoensis* point to the absence of selection from the plant of isolation, indicating that these species may be host generalists. Sustaining this, the phylogenomic tree built with a comprehensive set of *Xanthomonas* strains reveals that *X. euroxanthae* isolated from these three hosts do not cluster according to the host of isolation. Likewise, the reduced T3E profile and phylogenetic grouping of *X. arboricola* CPBF 765 and CPBF 1494 place them within the group B of *Xanthomonas arboricola* epidemic population structure, corresponding to generalists and weakly pathogenic strains, while *X. euroxanthae* CPBF 761 and CPBF 766 are more likely to be included in group C, as the phylogenomics tree reveals they are part of the monophyletic group with the *X. euroxanthae* from the diverse lineages of group C (Kałużna et al., 2021; Merda et al., 2016).

As *X. euroxanthae* strains CPBF 766 and CPBF 424^T are both pathogenic on walnut and display the same T3E pattern as Australian strains isolated from tomato (BRIP 62409, BRIP 62411, BRIP 62415, BRIP 62418), it would be interesting to verify if these strains are pathogenic on *Juglans regia* and use tomato as a reservoir. Given *X. euroxanthae* cosmopolite distribution and sharing of the same plant host species as *X. arboricola*, it is important to assess non-host plants with reservoir potential for these bacteria. In fact, a recent study focused on evaluating the potential of non-host plants as reservoirs for *X. arboricola* pv. *pruni* pointed out that this bacterium is able to live epiphytically on a wider

range of plants, with the highest threat being a wide range of crop and herbaceous plants including common bean and tomato (Zarei et al., 2018).

It is still to be uncovered if *Carya illinoensis* and *Solanum lycopersicum* act as reservoirs for *X. euroxantha* and if these strains are able to colonise all these different habitats as alternative hosts. The proximity of these two bacterial species is puzzling, and raises a couple of questions, such as: Is *X. euroxantha* able to colonise these same hosts in the absence of *X. arboricola* strains? Does the presence of *X. euroxantha* influences in any way the aggressiveness of diseases caused by *X. arboricola*?

Funding

Funding for this work was provided by the EU Horizon 2020 Research and Innovation program (grant agreement 668981 and 857251), European Structural and Investment Funds (ESIFs) through COMPETE 2020, National Funds through FCT (in the scope of the projects EVOXANT-PTDC/BIA-EVF/3635/2014-POCI-01-0145-FEDER-016600; [UIDB/50027/2020](https://doi.org/10.1007/978-94-007-5002-7)), and the Operational Thematic Program for Competitiveness and Internationalization (POCI) under the PORTUGAL 2020 Partnership Agreement through the European Regional Development Fund (FEDER) project PORBIOTA—POCI-01-0145- FEDER-022127. L.M. was supported by fellowships from FCT (SFRH/BD/137079/2018). This article is based upon work from COST Action CA16107 EuroXanth, supported by COST (European Cooperation in Science and Technology). The APC was funded by EU's Horizon 2020 Research and Innovation program through BIOPOLIS (grant agreement 857251).

Authors' Contributions

F.T. was responsible for the conception and design of the work. Genomics analyses were carried out by M.T.; formal analysis, L.M. and F.T.; investigation L.M., and M.T.; supervision, N.A.F. and F.T.; validation, C.F., L.M., M.T., N.A.F., and F.T.; All authors contributed to the drafting and revision of the manuscript

References

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Assis, R. A. B., Varani, A. M., Sagawa, C. H. D., Patané, J. S. L., Setubal, J. C., Uceda-Campos, G., da Silva, A. M., Zaini, P. A., Almeida, N. F., Moreira, L. M., & Dandekar, A. M. (2021). A comparative genomic analysis of *Xanthomonas arboricola* pv. *juglandis* strains reveal hallmarks of mobile genetic elements in the adaptation and accelerated evolution of virulence. *Genomics*, 113(4), 2513–2525. <https://doi.org/10.1016/j.ygeno.2021.06.003>
- Bertelli, C., Laird, M. R., Williams, K. P., Lau, B. Y., Hoad, G., Winsor, G. L., & Brinkman, F. S. L. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Research*, 45(Web Server issue), W30. <https://doi.org/10.1093/NAR/GKX343>
- Deb, S., Gokulan, C. G., Nathawat, R., Patel, H. K., Sonti, R. v, & Ramesh Sonti, C. v. (2022). Suppression of XopQ–XopX-induced immune responses of rice by the type III effector XopG. *Molecular Plant Pathology*. <https://doi.org/10.1111/MPP.13184>
- Dieckmann, M. A., Beyvers, S., Nkouamedjo-Fankep, R. C., Hanel, P. H. G., Jelonek, L., Blom, J., & Goesmann, A. (2021). EDGAR3.0: comparative genomics and phylogenomics on a scalable infrastructure. *Nucleic Acids Research*, 49(W1), W185–W192. <https://doi.org/10.1093/NAR/GKAB341>
- Edler, D., Klein, J., Antonelli, A., & Silvestro, D. (2021). raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods in Ecology and Evolution*, 12(2), 373–377. <https://doi.org/10.1111/2041-210X.13512>
- Essakhi, S., Cesbron, S., Fischer-Le Saux, M., Bonneau, S., Jacques, M.-A., & Manceau, C. (2015). Phylogenetic and Variable-Number Tandem-Repeat Analyses Identify Nonpathogenic *Xanthomonas arboricola* Lineages Lacking the Canonical Type III Secretion System. *Applied and Environmental Microbiology*, 81(16), 5395–5410. <https://doi.org/10.1128/aem.00835-15>
- Fernandes, C., Albuquerque, P., Mariz-Ponte, N., Cruz, L., & Tavares, F. (2021). Comprehensive diversity assessment of walnut-associated xanthomonads reveal the occurrence of distinct *Xanthomonas arboricola* lineages and of a new species

- (*Xanthomonas euroxanthea*) within the same tree. *Plant Pathology*, 70(4), 5. <https://doi.org/10.1111/ppa.13355>
- Fernandes, C., Martins, L., Teixeira, M., Blom, J., Pothier, J. F., Fonseca, N. A., & Tavares, F. (2021). Comparative Genomics of *Xanthomonas euroxanthea* and *Xanthomonas arboricola* pv. *juglandis* Strains Isolated from a Single Walnut Host Tree. *Microorganisms* 2021, Vol. 9, Page 624, 9(3), 624. <https://doi.org/10.3390/microorganisms9030624>
- Fernandes, C., Sousa, R., Tavares, F., & Cruz, L. (2018). First Report of *Xanthomonas arboricola* Causing Bacterial Blight on Pecan Trees in Portugal. *Plant Disease*, 102(12), 2632. <https://doi.org/10.1094/pdis-03-18-0467-pdn>
- Hajri, A., Brin, C., Hunault, G., Lardeux, F., Lemaire, C., Manceau, C., Boureau, T., & Poussier, S. (2009). A «repertoire for repertoire» hypothesis: Repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS ONE*, 4(8). <https://doi.org/10.1371/journal.pone.0006632>
- Hajri, A., Meyer, D., Delort, F., Guillaumès, J., Brin, C., & Manceau, C. (2010). Identification of a genetic lineage within *Xanthomonas arboricola* pv. *juglandis* as the causal agent of vertical oozing canker of Persian (English) walnut in France. *Plant Pathology*, 59(6), 1014–1022. <https://doi.org/10.1111/j.1365-3059.2010.02362.x>
- Hajri, A., Pothier, J. F., Saux, M. F. Le, Bonneau, S., Poussier, S., Boureau, T., Duffy, B., & Manceau, C. (2011). Type three effector gene distribution and sequence analysis provide new insights into the pathogenicity of plant-pathogenic *Xanthomonas arboricola*. *Applied and Environmental Microbiology*, 78(2), 371–384. <https://doi.org/10.1128/AEM.06119-11>
- Jacques, M.-A., Arlat, M., Boulanger, A., Boureau, T., Carrère, S., Cesbron, S., Chen, N. W. G., Cociancich, S., Darrasse, A., Denancé, N., Fischer-Le Saux, M., Gagnevin, L., Koebnik, R., Lauber, E., Noël, L. D., Pieretti, I., Portier, P., Pruvost, O., Rieux, A., ... Vernière, C. (2016). Using Ecology, Physiology, and Genomics to Understand Host Specificity in *Xanthomonas*. *Annual Review of Phytopathology*, 54(1), 163–187. <https://doi.org/10.1146/annurev-phyto-080615-100147>
- Kałużna, M., Fischer-Le Saux, M., Pothier, J. F., Jacques, M., Obradović, A., Tavares, F., & Stefani, E. (2021). *Xanthomonas arboricola* pv. *juglandis* and pv. *corylina*: Brothers or distant relatives? Genetic clues, epidemiology, and insights for disease

- p>management.
- Molecular Plant Pathology*
- , 22, 1481–1499.
-
- <https://doi.org/10.1111/mpp.13073>
- Lamichhane, J. R. (2014). *Xanthomonas arboricola* diseases of stone fruit, almond, and walnut trees: Progress toward understanding and management. *Plant Disease*, 98(12), 1600–1610. <https://doi.org/10.1094/PDIS-08-14-0831-FE>
- Lee, I., Kim, Y. O., Park, S. C., & Chun, J. (2016). OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *International Journal of Systematic and Evolutionary Microbiology*, 66(2), 1100–1103. <https://doi.org/10.1099/ijsem.0.000760>
- Martins, L., Fernandes, C., Blom, J., Dia, N. C., Pothier, J. F., & Tavares, F. (2020). *Xanthomonas euroxanthea* sp. nov., a new xanthomonad species including pathogenic and non-pathogenic strains of walnut. *International Journal of Systematic and Evolutionary Microbiology*, 70(12), 9. <https://doi.org/10.1099/ijsem.0.004386>
- Medina, C. A., Reyes, P. A., Trujillo, C. A., Gonzalez, J. L., Bejarano, D. A., Montenegro, N. A., Jacobs, J. M., Joe, A., Restrepo, S., Alfano, J. R., & Bernal, A. (2018). The role of type III effectors from *Xanthomonas axonopodis* pv. *manihotis* in virulence and suppression of plant immunity. *Molecular Plant Pathology*, 19(3), 593–606. <https://doi.org/10.1111/mpp.12545>
- Merda, D., Bonneau, S., Guimbaud, J. F., Durand, K., Brin, C., Boureau, T., Lemaire, C., Jacques, M. A., & Fischer-Le Saux, M. (2016). Recombination-prone bacterial strains form a reservoir from which epidemic clones emerge in agroecosystems. *Environmental Microbiology Reports*, 8(5), 572–581. <https://doi.org/10.1111/1758-2229.12397>
- Merda, D., Briand, M., Bosis, E., Rousseau, C., Portier, P., Barret, M., Jacques, M. A., & Fischer-Le Saux, M. (2017). Ancestral acquisitions, gene flow and multiple evolutionary trajectories of the type three secretion system and effectors in *Xanthomonas* plant pathogens. *Molecular Ecology*, 26(21), 5939–5952. <https://doi.org/10.1111/mec.14343>
- Moragrega, C., Matias, J., Aletà, N., Montesinos, E., & Rovira, M. (2011). Apical necrosis and premature drop of Persian (English) walnut fruit caused by *Xanthomonas arboricola* pv. *juglandis*. *Plant Disease*, 95(12), 1565–1570. <https://doi.org/10.1094/PDIS-03-11-0259>

- Morris, C. E., & Moury, B. (2019). Revisiting the Concept of Host Range of Plant Pathogens. *Annual Review of Phytopathology*, 57(1), 63–90. <https://doi.org/10.1146/annurev-phyto-082718-100034>
- Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, 302(1), 205–217. <https://doi.org/10.1006/jmbi.2000.4042>
- Popov, G., Fraiture, M., Brunner, F., & Sessa, G. (2016). Multiple *Xanthomonas euvesicatoria* Type III Effectors Inhibit flg22-Triggered Immunity . *Molecular Plant-Microbe Interactions*, 29(8), 651–660. <https://doi.org/10.1094/mpmi-07-16-0137-r>
- Potnis, N., Krasileva, K., Chow, V., Almeida, N. F., Patil, P. B., Ryan, R. P., Sharlach, M., Behlau, F., Dow, J. M., Momol, M. T., White, F. F., Preston, J. F., Vinatzer, B. A., Koebnik, R., Setubal, J. C., Norman, D. J., Staskawicz, B. J., & Jones, J. B. (2011). Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. *BMC Genomics*, 12. <https://doi.org/10.1186/1471-2164-12-146>
- Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences USA*, 106(45), 19126–19131. <https://doi.org/10.1073/pnas.0906412106>
- Roach, R., Mann, R., Gambley, C. G., Shivas, R. G., Chapman, T., & Rodoni, B. (2020). Pathogenicity and copper tolerance in Australian *Xanthomonas* species associated with bacterial leaf spot. *Crop Protection*, 127, 104923. <https://doi.org/10.1016/j.cropro.2019.104923>
- Roach, R., Mann, R., Gambley, C. G., Shivas, R. G., & Rodoni, B. (2018). Identification of *Xanthomonas* species associated with bacterial leaf spot of tomato, capsicum and chilli crops in eastern Australia. *European Journal of Plant Pathology*, 150(3), 595–608. <https://doi.org/10.1007/s10658-017-1303-9>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Sinha, D., Gupta, M. K., Patel, H. K., Ranjan, A., & Sonti, R. v. (2013). Cell Wall Degrading Enzyme Induced Rice Innate Immune Responses Are Suppressed by the Type 3 Secretion System Effectors XopN, XopQ, XopX and XopZ of

- Xanthomonas oryzae pv. oryzae. *PLoS ONE*, 8(9), 1–19.
<https://doi.org/10.1371/journal.pone.0075867>
- Song, C., & Yang, B. (2010). Mutagenesis of 18 Type III Effectors Reveals Virulence Function of XopZ PXO99 in *Xanthomonas oryzae* pv. *oryzae*. / *893 MPMI*, 23(7), 893–902. <https://doi.org/10.1094/MPMI>
- Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21), 2688–2690. <https://doi.org/10.1093/bioinformatics/btl446>
- Sullivan, M. J., Petty, N. K., & Beatson, S. A. (2011). Easyfig: A genome comparison visualizer. *Bioinformatics*, 27(7), 1009–1010. <https://doi.org/10.1093/bioinformatics/btr039>
- Tatusova, T., Dicuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M., & Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*, 44(14), 6614–6624. <https://doi.org/10.1093/nar/gkw569>
- Teixeira, M., Fernandes, C., Chaves, C., Pinto, J., Tavares, F., & Fonseca, N. A. (2021). Complete Genome Sequence Obtained by Nanopore and Illumina Hybrid Assembly of *Xanthomonas arboricola* pv. *juglandis* CPBF 427, Isolated from Buds of a Walnut Tree. *Microbiology Resource Announcements*, 10(10). <https://doi.org/10.1128/MRA.00085-21>
- Teixeira, M., Martins, L., Fernandes, C., Chaves, C., Pinto, J., Tavares, F., & Fonseca, N. A. (2020). Complete Genome Sequences of Walnut-Associated *Xanthomonas euroxanthea* Strains CPBF 367 and CPBF 426 Obtained by Illumina/Nanopore Hybrid Assembly. *Microbiology Resource Announcements*, 9(45). <https://doi.org/10.1128/MRA.00902-20>
- Thompson, T. E., & Conner, P. J. (2012). Pecan. *Fruit Breeding*, 771–801. https://doi.org/10.1007/978-1-4419-0763-9_20
- Vauterin, L., Hoste, B., Kersters, K., & Swings, J. (1995). Reclassification of *Xanthomonas*. *International Journal of Systematic Bacteriology*, 45(3), 472–489. <https://doi.org/10.1099/00207713-45-3-472>
- Verma, G., Mondal, K. K., Kulshreshtha, A., & Sharma, M. (2019). XopR T3SS-effector of *Xanthomonas oryzae* pv. *oryzae* suppresses cell death-mediated plant defense

response during bacterial blight development in rice. 3 *Biotech*, 9(7), 272. <https://doi.org/10.1007/s13205-019-1802-9>

Verma, G., Sharma, M., & Mondal, K. K. (2018). XopR TTSS-effector regulates in planta growth, virulence of Indian strain of *Xanthomonas oryzae* pv. *oryzae* via suppressing reactive oxygen species production and cell wall-associated rice immune responses during blight induction. *Functional Plant Biology*, 45(5), 561. <https://doi.org/10.1071/FP17147>

Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, 35(3), 543–548. <https://doi.org/10.1093/molbev/msx319>

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. (2017). <sc>ggtree</sc> : an <sc>r</sc> package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28–36. <https://doi.org/10.1111/2041-210X.12628>

Zarei, S., Taghavi, S. M., Hamzehzarghani, H., Osdaghi, E., & Lamichhane, J. R. (2018). Epiphytic growth of *Xanthomonas arboricola* and *Xanthomonas citri* on non-host plants. *Plant Pathology*, 67(3), 660–670. <https://doi.org/10.1111/ppa.12769>

Zarei, S., Taghavi, S. M., Rahimi, T., Mafakheri, H., Potnis, N., Koebnik, R., Saux, M. F.-L., Pothier, J. F., Palacio-Bielsa, A., Cubero, J., Portier, P., Jacques, M.-A., & Osdaghi, E. (2022). Taxonomic Refinement of *Xanthomonas arboricola*. <https://doi.org/10.1094/PHTO-12-21-0519-R>. <https://doi.org/10.1094/PHTO-12-21-0519-R>

Supplementary material

Supplementary Table IV.S1. List of genomes used in this study

Strain	Label	Assembly accession	Reference
<i>Xanthomonas euroxanthea</i> CPBF 766	CPBF 766	GCA_905367735.1	This study
<i>Xanthomonas arboricola</i> CPBF 765	CPBF 765	GCA_905367745	This study
<i>Xanthomonas euroxanthea</i> CPBF 761	CPBF 761	GCA_905367725.1	This study
<i>Xanthomonas arboricola</i> CPBF 1494	CPBF 1494	GCA_905367715	This study
<i>Xanthomonas</i> sp. BRIP 62418	BRIP 62418	GCA_003993495.1	Roach <i>et al.</i> 2019
<i>Xanthomonas</i> sp. BRIP 62415	BRIP 62415	GCA_003993515.1	Roach <i>et al.</i> 2019
<i>Xanthomonas</i> sp. BRIP 62411	BRIP 62411	GCA_003993525.1	Roach <i>et al.</i> 2019
<i>Xanthomonas</i> sp. BRIP 62409	BRIP 62409	GCA_003992965.1	Roach <i>et al.</i> 2019
<i>Xanthomonas prunicola</i> CFBP 8353 ^T	CFBP 8353 ^T	GCF_002846205.1	López <i>et al.</i> 2020
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> CFBP 2286	CFBP 2286	GCF_001042735.1	Wilkins <i>et al.</i> 2015
<i>Xanthomonas hortorum</i> pv. <i>carotae</i> M081	Xhc M081	GCF_000505565.1	Kimbrel <i>et al.</i> 2011
<i>Xanthomonas hortorum</i> pv. <i>gardneri</i> ICMP 7383	ICMP 7383	GCF_001908775.1	Richard <i>et al.</i> 2017
<i>Xanthomonas fuscans</i> subsp. <i>fuscans</i> 4834-R	Xfu 4834-R	GCF_000969685.2	Darrasse <i>et al.</i> 2013
<i>Xanthomonas fragariae</i> fap21	Xf fap21	GCF_001705565.1	Henry <i>et al.</i> 2016
<i>Xanthomonas euvesicatoria</i> LMG 930	LMG 930	GCF_001908795.1	Richard <i>et al.</i> 2017
<i>Xanthomonas euroxanthea</i> CPBF 426	CPBF 426	GCA_903989465.1	Teixeira <i>et al.</i> 2020
<i>Xanthomonas euroxanthea</i> CPBF 424 ^T	CPBF 424 ^T	GCA_905187425.1	Fernandes <i>et al.</i> 2021
<i>Xanthomonas euroxanthea</i> CPBF 367	CPBF 367	GCA_903989455.1	Teixeira <i>et al.</i> 2020
<i>Xanthomonas citri</i> subsp. <i>citri</i> A306	Xci A306	GCF_000816885.1	da Silva <i>et al.</i> 2002
<i>Xanthomonas axonopodis</i> Xac29-1	Xac 29-1	GCF_000348585.1	Shanghai Jiao Tong University
<i>Xanthomonas arboricola</i> pv. <i>zantedeschiae</i> CFBP 7410 ^{PT}	CFBP 7410 ^{PT}	GCF_002940625.1	Merda <i>et al.</i> 2017
<i>Xanthomonas arboricola</i> pv. <i>pruni</i> IVIA 2626 1	IVIA 2626	GCF_001306955.1	Garita-Cambronero <i>et al.</i> 2016
<i>Xanthomonas arboricola</i> pv. <i>pruni</i> CFBP 3894 ^{PT}	CFBP 3894 ^{PT}	GCF_001741965.1	López-Soriano <i>et al.</i> 2016
<i>Xanthomonas arboricola</i> pv. <i>populi</i> CFBP 3123 ^{PT}	CFBP 3123 ^{PT}	GCF_002939945.1	Merda <i>et al.</i> 2017

<i>Xanthomonas arboricola</i> pv. <i>populi</i> CFBP 3122	CFBP 3122	GCF_002940585.1	Merda <i>et al.</i> 2017
<i>Xanthomonas arboricola</i> pv. <i>juglandis</i> NCPPB 1447	NCPPB 1447	GCF_000306055.1	Chungbuk National University
<i>Xanthomonas arboricola</i> pv. <i>juglandis</i> J303	Xaj J303	GCF_001643295.1	Retamales <i>et al.</i> 2017
<i>Xanthomonas arboricola</i> pv. <i>juglandis</i> DW3F3	Xaj DW3F3	GCF_002879695.1	Fu <i>et al.</i> 2013
<i>Xanthomonas arboricola</i> pv. <i>juglandis</i> CPBF 427	CPBF 427	GCA_903989475	Teixeira <i>et al.</i> 2021
<i>Xanthomonas arboricola</i> pv. <i>juglandis</i> CFBP 7179	CFBP 7179	GCF_001013495.1	Cesbron <i>et al.</i> 2015
<i>Xanthomonas arboricola</i> pv. <i>juglandis</i> CFBP 2528 ^T	CFBP 2528 ^T	GCF_001013475.1	Cesbron <i>et al.</i> 2015
<i>Xanthomonas arboricola</i> pv. <i>juglandis</i> 417	Xaj 417	GCF_001237985.1	Pereira <i>et al.</i> 2015
<i>Xanthomonas arboricola</i> pv. <i>fragariae</i> LMG 19145 ^{PT}	LMG 19145 ^{PT}	GCF_900240435.1	Vandroemme <i>et al.</i> 2013
<i>Xanthomonas arboricola</i> pv. <i>fragariae</i> CFBP 6773	CFBP 6773	GCF_900240345.1	Justus-Liebig-University Giessen
<i>Xanthomonas arboricola</i> pv. <i>corylina</i> NCCB 100457	NCCB 100457	GCF_000355635.2	Caballero <i>et al.</i> 2013
<i>Xanthomonas arboricola</i> pv. <i>celebensis</i> NCPPB 1832 ^{PT}	NCPPB 1832 ^{PT}	GCF_000724925.1	Harrison <i>et al.</i> 2016
<i>Xanthomonas arboricola</i> pv. <i>celebensis</i> NCPPB 1630	NCPPB 1630	GCF_000724915.1	Harrison <i>et al.</i> 2016
<i>Xanthomonas arboricola</i> pv. <i>arracaciae</i> CFBP 7407 ^{PT}	CFBP 7407 ^{PT}	GCF_002940565.1	Merda <i>et al.</i> 2017
<i>Xanthomonas arboricola</i> CITA 44	CITA44	GCF_001306965.1	Garita-Cambronero <i>et al.</i> 2016
<i>Xanthomonas arboricola</i> CITA 14	CITA14	GCF_001675005.1	Garita-Cambronero <i>et al.</i> 2017
<i>Xanthomonas arboricola</i> CITA 124	CITA 124	GCF_001674995.1	Garita-Cambronero <i>et al.</i> 2017
<i>Xanthomonas arboricola</i> CFBP 7652	CFBP 7652	GCF_002940445.1	Merda <i>et al.</i> 2017
<i>Xanthomonas arboricola</i> CFBP 7651	CFBP 7651	GCF_001013505.1	Cesbron <i>et al.</i> 2015
<i>Xanthomonas arboricola</i> CFBP 7645	CFBP 7645	GCF_002940665.1	Merda <i>et al.</i> 2017
<i>Xanthomonas arboricola</i> CFBP 7634	CFBP 7634	GCF_001013485.1	Cesbron <i>et al.</i> 2015
<i>Xanthomonas arboricola</i> CFBP 7629	CFBP 7629	GCF_002940425.1	Merda <i>et al.</i> 2017
<i>Xanthomonas arboricola</i> BRIP 62432	BRIP 62432	GCA_003993415.1	Roach <i>et al.</i> 2019
<i>Xanthomonas arboricola</i> BRIP 62416	BRIP 62416	GCA_003993355.1	Roach <i>et al.</i> 2019
<i>Xanthomonas arboricola</i> BRIP 62412	BRIP 62412	GCA_003993745.1	Roach <i>et al.</i> 2019
<i>Xanthomonas arboricola</i> BRIP 62410	BRIP 62410	GCA_003993395.1	Roach <i>et al.</i> 2019
<i>Xanthomonas arboricola</i> 3004	Xa 3004	GCF_000585435.1	Ignatov <i>et al.</i> 2015

Supplementary Table IV.S2. Query proteins used in T3E/T3SS prediction

	Query protein	GenBank accession number	Query organism	Genome GenBank accession number/WGS project
T3E	XopAZ	AAM40608.1	Xcc str. ATCC 33913	AE008922.1
	XopR	CAJ21916.1	Xcc str 85-10	AM039952.1
	AvrXccA1.1	AKU51772.1	Xaj 417	CP012251.1
	XopF1	CAJ22045.1	Xcv str. 85-10	AM039952.1
	XopF2	CAJ24621.1	Xcv str. 85-10	AM039952.1
	XopM	CAJ22073.1	Xcv str. 85-10	AM039952.1
	XopA	AAM40538.1	Xcc str. ATCC 33913	AE008922.1
	XopAW	CAJ24824.1	Xcv str. 85-10	AM039952.1
	AvrBs2	AKU51808.1	Xaj 417	CP012251.1
	XopY	BAE68243.1	Xoo MAFF 311018	AP008229.1
	XopZ2	ENZ96816.1	X. fragariae LMG 25863	AJRZ: Xanthomonas fragariae LMG 25863 (wgs)
	XopH/AvrBs1.1	AAM41387.1	Xcc str. ATCC 33913	AE008922.1
	XopAB	AEQ95831.1	Xoo BLS256	CP003057.2
	XopQ	AKU49288.1	Xaj 417	CP012251.1
	AvrB3	AKU50911.1	Xaj 417	CP012251.1
	AvrBs1	AAM41388.1	Xcc str. ATCC 33913	AE008922.1
	AvrXccA2	AKU52192.1	Xaj 417	CP012251.1
	XopAL1	AAM40544.1	Xcc str. ATCC 33913	AE008922.1
	XopC	CAJ24112.1	Xcv str. 85-10	AM039952.1
	XopG	BAE71013.1	Xoo MAFF311018	AP008229.1
	XopK	AKU50592.1	Xaj 417	CP012251.1
	XopL	AKU50613.1	Xaj 417	CP012251.1
	XopN	AKU50360.1	Xaj 417	CP012251.1
	XopV	AKU49758.1	Xaj 417	CP012251.1
	XopX	AKU49819.1	Xaj 417	CP012251.1
	XopZ1	AAM36871.1	X. citri pv. citri str. 306	AE008923.1
	XopAA	CAJ25516.2	Xcv str. 85-10	AM039952.1

XopAC	AAM41837.1	Xcc str. ATCC 33913	AE008922.1
XopAD	BAE70900.1	Xoo MAFF311018	AP008229.1
HpaF/XopAE	AAW73319.1	Xoo KACC 10331	PRJNA12931
HopAF1/XopAF1	ZP_02241516.1	Xo pv. oryzicola BLS256	PRJNA350867
XopAF/XopAF2	AEQ94681.1	Xo pv. oryzicola BLS256	PRJNA350867
AvrGf1/XopAG	ABB84189.1	X. citri	PRJNA255042
XopAH	AAV49067.1	Xcc str. 8004	PRJNA15
XopAI	AKU52422.1	Xaj 417	CP012251.1
XopAJ	CAJ26159.1	Xcv str. 85-10	AM039952.1
XopAK	CAJ25517.1	Xcv str. 85-10	AM039952.1
XopAL2	CAP49952.1	Xcc B100	PRJNA29801
XopAM	CAP52621.1	Xcc B100	PRJNA29801
XopAP	CAJ24869.1	Xcv str. 85-10	AM039952.1
XopAU	CAJ22827.1	Xcv str. 85-10	AM039952.1
XopAV	ACD60298.1	Xoo PXO99A	PRJNA28127
XopAX	CAJ19898.1	Xcv str. 85-10	AM039952.1
XopAY	AEQ95425.1	Xo pv. oryzicola BLS256	PRJNA350867
XopB	CAJ22212.1	Xcv str. 85-10	AM039952.1
XopC2	AAW76678.1	Xoo KACC 10331	PRJNA12931
XopD	CAJ22068.1	Xcv str. 85-10	AM039952.1
XopE1	CAJ21925.1	Xcv str. 85-10	AM039952.1
XopE2	WP_104613731.1*	Xanthomonas arboricola	not applicable
XopE3	AAM38068.1	X. citri pv. citri str. 306	AE008923.1
XopI	AAM35643.1	X. citri pv. citri str. 306	AE008923.1
XopJ	CAJ23833.1	Xcv str. 85-10	AM039952.1
XopO	CAJ22686.1	Xcv str. 85-10	AM039952.1
XopP	AAM40545.1	Xcc str. ATCC 33913	AE008922.1
XopS	CAJ21955.1	Xcv str. 85-10	AM039952.1
XopT	BAE68965.1	Xoo MAFF311018	AP008229.1

T3SS	XopU	BAE69632.1	Xoo MAFF311018	AP008229.1
	XopW	BAE66792.1	Xoo MAFF311018	AP008229.1
	hrcN	AKU51301.1	Xaj 417	CP012251.1
	HrpG	ENZ95267.1	X. fragariae LMG 25863	AJRZ: Xanthomonas fragariae LMG 25863 (wgs)
	HrpX	ENZ95268.1	X. fragariae LMG 25863	AJRZ: Xanthomonas fragariae LMG 25863 (wgs)
	HrcJ	AKU51304.1/WP_006451031.1	Xaj 417	CP012251.1
	HrcQ	SOT93783.1	Xaf CFBP 6762	OEQE: Xanthomonas arboricola pv. fragariae strain CFBP 6762 (wgs)
	HrcR	AAM40524.1/WP_006451038.1	Xcc str. ATCC 33913	AE008922.1
	HrcS	AKU51312.1/WP_016901415.1	Xaj 417	CP012251.1
	HrcT	AKU51299.1	Xaj 417	CP012251.1
	HrcU	WP_053045999.1	Xaj CPBF 427	LR861807.1
	HrcV	AKU51308.1	Xaj 417	CP012251.1
	HrpB1	WP_026064309.1*	X. arboricola	not applicable
	HrpF	AKU51319.1	Xaj 417	CP012251.1
	HrpW	SOT93768.1	Xaf CFBP 6762	OEQE: Xanthomonas arboricola pv. fragariae strain CFBP 6762 (wgs)
	HrcC	AKU51298.1/WP_026064306.1	Xaj 417	CP012251.1
	HpaA	SOU04637.1/AFV80108.1	Xaf LMG 19146	OEQG: Xanthomonas arboricola pv. fragariae strain LMG 19146 (wgs)
	HpaB	WP_005912720.1	Xanthomonas	not applicable
	Hpa2	WP_003486465.1	Xanthomonas	not applicable
	HrcL	WP_017118385.1*	Xanthomonas vasicola	not applicable
	HrpB2	WP_014505147.1*	Xanthomonas oryzae	not applicable
	HrpB4	WP_017171652.1*	Xanthomonas phaseoli	not applicable
	HrpB7	WP_019301113.1*	Xanthomonas oryzae	not applicable
	HpaC/sctP	WP_008575445.1*	Xanthomonas	not applicable
	HrcD	WP_005994475.1*	Xanthomonas vesicatoria	not applicable
	HrpD6	WP_006451041.1*	Xanthomonas	not applicable
	HrpE	WP_011036417.1*	Xanthomonas campestris	not applicable

*REFSEQ: This record represents a single, non-redundant, protein sequence which may be annotated on many different RefSeq genomes from the same, or different, specie

CLUSTAL O(1.2.4) multiple sequence alignment

CLUSTAL O(1.2.4) multiple sequence
alignments(424_XTG000468^ax766_XCY000470x765_XCY000604x427_XSP003
898)

CLUSTAL O(1.2.4) multiple sequence alignment

CLUSTAL O(1.2.4) multiple sequence
alignment(424_XTG000469x766_XCY000471x765_XCY000605x427_XSP00389
6)

blastp results:

-XTG000468⁽¹⁾ (CPBF 424): hits with hypothetical proteins of several species including *X. euroxanthea*, *X. arboricola*, *X. hortorum* and *Burkholderia* spp.

-XTG000469⁽³⁾ (CPBF 424): hits with hypothetical proteins of several *Xanthomonas* species and 1 hit for hpa3 of *Xanthomonas campestris* pv. *raphani* 756C

-XSP003897⁽⁴⁾ (CPBF 427): best blast hit-IS5 family transposase



Chapter V

DNA markers for Detection and
Genotyping of *Xanthomonas*
euroxanthea

Chapter V include the following publication:

Silva KG*, Martins L*, Teixeira M, Pothier JF, Tavares F (2022) DNA Markers for Detection and Genotyping of *Xanthomonas euroxanthea*. *Microorganisms* 2022, 10: 1078

* These authors contributed equally to this work.

Part of the results from this chapter were presented at:

Martins, L. Silva, K. Teixeira, M. Pothier, J. F. Tavares, F. *Xanthomonas euroxanthea*-specific DNA Markers for Genotyping and Multiplex PCR-based Detection. 14th International Congress of Plant Pathogenic Bacteria (ICPPB) “The impact of Plant Pathogenic Bacteria on Global Plant Health” - Assisi, Italy, 3-8 July 2022 (Oral presentation)

Silva, K.G. Martins, L. Teixeira, M. Pothier, J.F. Tavares, F. Highlighting the genome of *Xanthomonas euroxanthea*: DNA Markers for its Detection and Genotyping. 15th Encontro de Investigação Jovem da U.Porto (IJUP 2022), Porto, Portugal, May 4-6, 2022.

DNA Markers for Detection and Genotyping of

Xanthomonas euroxanthea

Abstract

Xanthomonas euroxanthea is a bacterial species encompassing both pathogenic and non-pathogenic strains, and is frequently found colonizing the same host plants as *X. arboricola*. This calls for the need to develop a detection and genotyping assay able to track these bacteria in microbial consortia with other xanthomonads. Eight *X. euroxanthea*-specific DNA markers (XEA1 to XEA8) were selected by comparative genomics; and validated *in silico* regarding their specificity and consistency by BLASTn, synteny analysis, GC content, codon usage (CAI/eCAI values), and genomic proximity to plasticity determinants. *In silico*, the selected eight DNA markers were found to be specific and conserved across the genomes of 11 *X. euroxanthea* strains, and, particularly, five DNA markers (XEA4, XEA5, XEA6, XEA7 and XEA8) were unfailingly found in these genomes. A multiplex PCR targeting markers of different amplicon size (XEA1, XEA5, and XEA8) was shown to successfully detect *X. euroxanthea* down to 1 ng of DNA (per PCR reaction). The topology of trees generated with the concatenated sequences of three markers (XEA5, XEA6, and XEA8) and four housekeeping genes (*gyrB*, *rpoD*, *fyuA*, and *acnB*) underlined the equal discriminatory power of these features, and thus the suitability of the DNA markers to discriminate *X. euroxanthea* lineages. Overall, this study displays a method conciliating the detection and genotyping of *X. euroxanthea* strains, contributing to monitoring for its presence in *X. arboricola*-colonizing habitats.

1. Introduction

Following extensive genotyping and comparative genomics studies carried out on walnut-associated bacterial isolates, it has been shown that not all the isolates could be identified as *Xanthomonas arboricola* pv. *juglandis*, the phytopathogen commonly acknowledged as causing walnut bacterial blight (WBB) (Moragrega & Özaktan, 2010; Hajri *et al.*, 2010; Moragrega *et al.*, 2011). Further studies showed that some walnut-associated *Xanthomonas* isolates were taxonomically distinct from any of the other described *Xanthomonas* species and were proposed as members of the new species *Xanthomonas euroxanthea* (Martins *et al.*, 2020). Pathogenicity assays indicated that *X. euroxanthea* encompasses non-pathogenic, and pathogenic strains that can cause

WBB-like symptoms, being in a privileged position to investigate genetic determinants of pathogenesis and its evolution (Fernandes *et al.*, 2018; Martins *et al.*, 2020; Teixeira *et al.*, 2020). Interestingly, besides its occurrence in walnuts (*Juglans regia*), recent evidence was gathered reporting the isolation of *X. euroxantha* from very distinct plant host species, such as *Carya illinoensis* (pecan; strains CPBF 761 and CPBF 766), that together with walnut (*Juglans regia*; strains CPBF 367, CPBF 424^T, CPBF 426 and CFBP 7653) belong to the Juglandaceae family; *Solanum lycopersicum* (tomato plants; strains BRIP 62409, BRIP 62411, BRIP 62415 and BRIP 62418) a member of the Solanaceae family (Roach *et al.*, 2018; Fernandes *et al.*, 2021b; Martins *et al.*, 2022); and *Phaseolus vulgaris* (common bean; strain CFBP 7622, previously misclassified as *X. arboricola* (Kałużna *et al.*, 2021; Zarei *et al.*, 2022)), a Fabaceae plant species. More strikingly is that *X. arboricola* strains were also isolated from all the mentioned plant species, suggesting that both *X. euroxantha* and *X. arboricola* share the same host plants, including the same plant specimen which raises questions regarding co-colonization and niche-specific adaptations (Fernandes *et al.*, 2021a,b; Martins *et al.*, 2022). Still, taking into consideration the recent taxonomic refinement of *X. arboricola* species and the unearthing of six new *X. euroxantha* strains (2949, 2955, 2957, 2974, 3640 and F2) (Zarei *et al.*, 2022) it is foreseeable the identification of additional *X. euroxantha* isolates from plant species that have passed unnoticed so far.

Altogether, the common ecological niche of these two closely related species, plus their cosmopolitan distribution and co-occurrence in different host plant species resulted in misclassifying some *X. euroxantha* isolates as *X. arboricola* (Fernandes *et al.*, 2021b; Kałużna *et al.*, 2021; Zarei *et al.*, 2022), which calls for the need to develop methods for the detection and genotyping of this bacterial species. Over the years, different diagnostic and molecular typing tools have been developed for a variety of xanthomonads (Catara *et al.*, 2021), namely for *X. arboricola* including their most studied pathovars *juglandis* (Fernandes *et al.*, 2017; Martins *et al.*, 2019) and *pruni* (Palacio-Bielsa *et al.*, 2011, 2015; Pothier *et al.*, 2011a; Loreti *et al.*, 2015), aiming to address their diversity within a geographic or epidemiological context. Generally, these approaches consisted of culture-based detection of the phytopathogen by PCR, followed by multilocus sequence analysis (MLSA) of several housekeeping genes to define haplotypes. In fact, multiplex-PCR and dot-blot hybridization assays showed that three *X. arboricola* pv. *juglandis* specific DNA markers (XAJ1, XAJ6 and XAJ8) were absent from *X. euroxantha* strains CPBF 367, CPBF 424^T and CPBF 426, which formed a distinct MLSA cluster (Fernandes *et al.*, 2017, 2021a). A robust method to detect *X. euroxantha* while undoubtedly distinguishing it from *X. arboricola* has not been described so far. Furthermore, while several genotyping techniques have been used to

assess the diversity of phytosanitary regulated *Xanthomonas*, as recently reviewed (Catara *et al.*, 2021), the data currently available regarding genotyping of *X. euroxanthae* is scarce and limited to MLSA studies carried out on *Xanthomonas* isolated from walnut trees (Fernandes *et al.*, 2021a).

In this study, comparative genomics, and *in silico* validation tools were combined as previously described (Albuquerque *et al.*, 2012) to select eight *X. euroxanthae*-specific DNA markers located in conserved genomic regions. While XEA1, XEA5 and XEA8 DNA markers were chosen to optimize a multiplex PCR detection method for the reliable detection of the target bacteria, altogether the number of SNPs recorded for DNA markers XEA5, XEA6 and XEA8 within the studied *X. euroxanthae* strains, revealed an allelic variation capable to discriminate *X. euroxanthae* strains as efficiently as the housekeeping genes used in MLSA. Ultimately, this work may unlock the possibility to conciliate bacterial detection and diversity assessment using the same DNA markers, which may be particularly useful to survey *X. euroxanthae* populations in environmental samples.

2. Materials and Methods

2.1. *In silico* selection and validation of *X. euroxanthae*-specific DNA markers

A synteny analysis of 11 *X. euroxanthae* strains and other 24 representative xanthomonads, downloaded from the NCBI database (Table V.1), was performed using MaGe v3.15.3 (Vallenet *et al.*, 2020). This led to the identification of *X. euroxanthae*-specific coding DNA sequences (CDSs) that are concomitantly present in *X. euroxanthae* genomes and absent from non-*X. euroxanthae* strains.

Then, a BLAST was performed in NCBI (accessed in November 2021 at <https://www.ncbi.nlm.nih.gov/>) using as query the putative *X. euroxanthae*-specific CDSs, and the database nr/nt to further confirm their specificity. Sequences with hits pertaining solely to *X. euroxanthae* were considered putative *X. euroxanthae*-specific genomic regions (Supplementary Table V.S1) and were used for primer and specific DNA marker design using Geneious® 9.1.8 (Kearse *et al.*, 2012). The affinity and complementarity of these primers to target *X. euroxanthae* CDSs was checked using NCBI's Primer-BLAST tool (Stothard, 2000; Kearse *et al.*, 2012). The obtained eight DNA markers were then evaluated for specificity by a BLASTn analysis in Geneious® 9.1.8 and NCBI (using the database nr/nt). The genomic context of the markers was assessed to ensure its consistency across the diversity of *X. euroxanthae* strains. Particularly, a

comparative genomics analysis was performed by local alignment of CDSs in MaGe (Supplementary Figure V.S1 (a-d), to evaluate the syntenic context.

Table V.1. Bacterial strains used for MaGe synteny analysis to retrieve *Xanthomonas euroxantha* specific coding sequences (CDSs).

<i>Xanthomonas</i> pathovars	species	and Strains	GenBank, NCBI Accession/ WGS prefix
<i>X. euroxantha</i>		CPBF 367	LR861803.1
<i>X. euroxantha</i>		CPBF 424 ^T	LR994544.1
<i>X. euroxantha</i>		CPBF 426	LR861805.1
<i>X. euroxantha</i>		CPBF 761	HG999363.1
<i>X. euroxantha</i>		CPBF 766	HG999364.1
<i>X. euroxantha</i>		CFBP 7622	MIGF01.1
<i>X. euroxantha</i>		CFBP 7653	MIGK01.1
<i>X. euroxantha</i>		BRIP 62409	QEJZ01.1
<i>X. euroxantha</i>		BRIP 624011	QEZI01.1
<i>X. euroxantha</i>		BRIP 62415	QEZH01.1
<i>X. euroxantha</i>		BRIP 62418	QEZG01.1
<i>X. arboricola</i>		CPBF 1494	HG999362.1
<i>X. arboricola</i>		CPBF 765	HG999365.1
<i>X. arboricola</i> pv. <i>juglandis</i>		CPBF 427	LR861807.1
<i>X. campestris</i> pv. <i>campestris</i>		LMG 568 ^{PT}	NC_003902
<i>X. campestris</i> pv. <i>campestris</i>		8004	NC_007086
<i>X. campestris</i> pv. <i>campestris</i>		B100	NC_010688
<i>X. citri</i> pv. <i>citri</i>		306	NC_003919
<i>X. citri</i> pv. <i>bilvae</i>		NCPPB 3213 ^{PT}	CDHI01
<i>X. phaseoli</i> pv. <i>phaseoli</i>		CFBP 412	NZ_CP020964.2
<i>X. citri</i> subsp. <i>aurantifolli</i>		ICPB 10535	ACPY01
<i>X. vasicola</i> pv. <i>musacearum</i>		BCC282	RRCQ01
<i>X. vasicola</i> pv. <i>musacearum</i>		NCPPB 4381	ACHT01
<i>X. oryzae</i> pv. <i>oryzae</i>		PX099 ^A	NC_010717.1
<i>X. oryzae</i> pv. <i>oryzae</i>		MAFF 311018	NC_007705.1
<i>X. oryzae</i> pv. <i>oryzae</i>		KACC 10331	NC_006834.1
<i>X. oryzae</i> pv. <i>oryzicola</i>		BLS256	NZ_AAQN
<i>X. translucens</i> pv. <i>translucens</i>		569	VIWM01.1
<i>X. translucens</i> pv. <i>translucens</i>		LMG 876 ^{PT}	NZ_CAPJ.1
<i>X. sacchari</i>		NCPPB 4393	AGDB01.1
<i>X. hortorum</i> pv. <i>gardneri</i>		ICMP 7383	CP018731.1
<i>X. hortorum</i> pv. <i>gardneri</i>		LMG 962 ^T	NZ_AEQX.1
<i>X. vesicatoria</i>		LMG 911 ^T	NZ_AEQV
<i>X. euvesicatoria</i> pv. <i>perforans</i>		91-118	NZ_AEQW
<i>X. albilineans</i>		GPE PC73	NC_013722

Additionally, features such as CAI/eCAI values (Puigbò *et al.*, 2008a); GC content (deducted from MaGe); chromosomal location and proximity to determinants of genomic plasticity, namely transposons, integrases, recombinases and phage related ORFs (Geneious® 9.1.8) were annotated. The number of Single Nucleotide Polymorphisms

(SNPs) included in the eight DNA markers and housekeeping genes of the 11 *X. euroxantha* genomes considered in this study were summed up recurring to Geneious® 9.1.8, and normalized using the formula $\left(\frac{\text{SNPs number}}{\text{total nucleotide sequence length}} \times 100\right)$. The described procedure followed for specific *X. euroxantha* DNA markers design is systematized in Figure V.1.

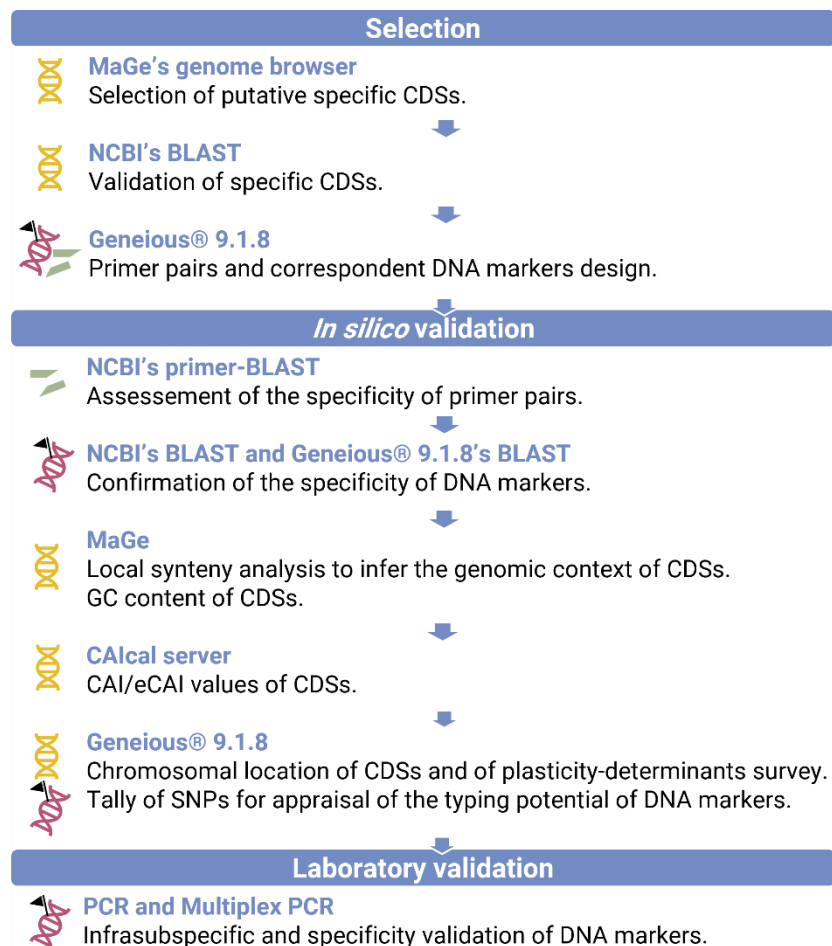


Figure V.1. Flowchart for selection and validation of *X. euroxantha*-specific DNA markers

2.2. Bacterial strains, culture conditions and DNA extraction

The bacterial strains used for validation of the eight *X. euroxantha*-specific markers are listed in Table V.2, and include seven *X. euroxantha* strains; and other closely related and niche-sharing strains, namely 10 strains of *X. arboricola*, seven strains representing six pathovars of *X. arboricola*, and 11 strains belonging to non-*arboricola* *Xanthomonas* species. Bacterial strains were cultured as previously described (Fernandes *et al.*, 2017) or in peptone-sucrose-agar (PSA) medium (10 g peptone; 10 g sucrose; 1 g glutamic acid; 15 g agar and distilled water up to 1.0 L) at 28°C. DNA was extracted from pure

cultures using the EZNA Bacterial DNA Purification kit (Omega Bio-Tek, Norcross, GA), according to the manufacturer's instructions, and quantified using a DS-11 microvolume spectrophotometer (DeNovix, Wilmington, USA).

Table V.2. List of bacterial strains used for validation of the *Xanthomonas euroxantha*-specific DNA markers.

<i>Xanthomonas</i> species and pathovars	Strains ¹	Geographic origin	Year of isolation
<i>X. euroxantha</i>	CPBF 367	Portugal (Loures)	2016
<i>X. euroxantha</i>	CPBF 424 ^T	Portugal (Loures)	2016
<i>X. euroxantha</i>	CPBF 426	Portugal (Loures)	2016
<i>X. euroxantha</i>	CPBF 761	Portugal (Alcobaça)	2016
<i>X. euroxantha</i>	CPBF 766	Portugal (Alcobaça)	2016
<i>X. euroxantha</i>	CFBP 7622	USA	1985
<i>X. euroxantha</i>	CFBP 7653	France	2008
<i>X. arboricola</i>	CPBF 122	Portugal (Ponte da Barca)	2015
<i>X. arboricola</i>	CPBF 237	Portugal (Ponte de Lima)	2015
<i>X. arboricola</i>	CPBF 554	Portugal (Carraceda de Ansiães)	2016
<i>X. arboricola</i>	CPBF 765	Portugal (Alcobaça)	2016
<i>X. arboricola</i>	CPBF 796	Portugal (Alcobaça)	2016
<i>X. arboricola</i>	CPBF 1494	Portugal (Alcobaça)	2014
<i>X. arboricola</i>	CPBF 1483	Portugal (Alcobaça)	2014
<i>X. arboricola</i>	CPBF 1514	Portugal (Estremoz)	2014
<i>X. arboricola</i>	CPBF 1567	Portugal (Bombarral)	2015
<i>X. arboricola</i>	CPBF 1586	Portugal (Loures)	2015
<i>X. arboricola</i> pv. <i>juglandis</i>	CPBF 427	Portugal (Loures)	2016
<i>X. arboricola</i> pv. <i>juglandis</i>	CPBF 1521	Portugal (Loures)	2014
<i>X. arboricola</i> pv. <i>celebensis</i>	LMG 677 ^{PT}	New Zealand	1960
<i>X. arboricola</i> pv. <i>corylina</i>	LMG 689 ^{PT}	USA	1939
<i>X. arboricola</i> pv. <i>fragariae</i>	LMG 19145 ^{PT}	Italy	1993
<i>X. arboricola</i> pv. <i>populi</i>	CFBP 3123 ^{PT}	Netherlands	1979
<i>X. arboricola</i> pv. <i>pruni</i>	LMG 852 ^{PT}	New Zealand	1953
<i>X. citri</i> pv. <i>citri</i>	LMG 9322 ^T	USA	1989
<i>X. campestris</i> pv. <i>campestris</i>	LMG 568 ^{PT}	United Kingdom	1957
<i>X. axonopodis</i> pv. <i>dieffenbachiae</i>	LMG 695 ^{PT}	Brazil	1965
<i>X. fragariae</i>	LMG 708 ^T	USA	1960
<i>X. oryzae</i> pv. <i>oryzicola</i>	LMG 797 ^{PT}	Malaysia	1964
<i>X. translucens</i> pv. <i>translucens</i>	LMG 876 ^{PT}	USA	1933
<i>X. vesicatoria</i>	LMG 911 ^T	New Zealand	1955
<i>X. euvesicatoria</i> pv. <i>euvesicatoria</i>	LMG 922	USA	1939
<i>X. hortorum</i> pv. <i>gardneri</i>	LMG 962 ^T	Yugoslavia	1953
<i>X. euvesicatoria</i> pv. <i>perforans</i>	NCPBP 4321 ^T	USA	1933
<i>X. oryzae</i> pv. <i>oryzae</i>	LMG 5047 ^{PT}	India	1965

¹ CPBF: Portuguese Collection of Phytopathogenic Bacteria, Instituto Nacional de Investigação Agrária e Veterinária, I.P. Oeiras, Portugal. CFBP: French Collection for Plant-associated Bacteria, Institut National de la Recherche Agronomique, Angers, France. LMG: Belgian Coordinated Collections of Microorganisms/ LMG Bacteria Collection, Universiteit Gent –

Laboratorium voor Microbiologie, Gent, Belgium. NCPPB: National Collection of Plant Pathogenic Bacteria, Fera Science Ltd., York, United Kingdom.
Superscript following strain names indicate ^T the type strain of a species and ^{PT} the pathotype strain for a pathovar.

2.3. Experimental validation of putative *X. euroxantha*-specific DNA markers by multiplex PCR

A multiplex PCR targeting the most promising DNA markers was optimized to validate a method to rapidly identify *X. euroxantha* isolates. XEA1, XEA5 and XEA8 were the chosen markers attending their distinct amplicon lengths of 819 bp, 295 bp and 648 bp, respectively; and their broad occurrence in the tested *X. euroxantha* strains, apart from XEA1 (absent in CFBP 7622). A 20 µL PCR reaction mix consisted of 1× DreamTaq Buffer (ThermoFisher Scientific, Massachusetts, USA), 0.2 mM of each deoxynucleotide triphosphate (dNTP) (Grisp, Porto, Portugal), 0.2 mM of each forward and reverse primers (Supplementary Table V.S3), 1.5 U of DreamTaq DNA Polymerase (ThermoFisher Scientific, Massachusetts, USA), and 25 ng of DNA template. Sterile distilled water was used as the negative control. PCR cycling parameters consisted of a first amplification cycle of 5 min at 95°C, followed by 35 cycles of 95°C for 30 s, 61°C for 15 s, and 72°C for 30 s, and a final DNA extension at 72°C for 10 min. The same DNA samples were used as template in PCR reactions using each of the markers individually (XEA1, XEA5 and XEA8), and 1.0 U of DreamTaq DNA polymerase per PCR reaction. PCR products were separated by electrophoresis on a 0.8% agarose gel (1 × TAE buffer) and visualized using Xpert Green DNA stain (Grisp, Porto, Portugal) with a Molecular Imager Gel Doc XR+ System (Bio-Rad, California, USA). The obtained PCR products for each marker and each strain were purified using the Illustra GFX GEL Band Purification kit (GE Healthcare, Buckingham-shire, United Kingdom), following the reference protocol available, and sequenced on both strands (STAB Vida, Caparica, Portugal) to confirm their identity and determine the number of SNPs.

Table V.3. Selected *Xanthomonas euroxanthae*-specific DNA markers (XEA1 to XEA8), corresponding primer pair sequences, expected amplicon sizes, and best BLASTn hits of amplicons with non-*X. euroxanthae* genomes.

DNA markers	CDS (MaGe) ¹	Locus tag (NCBI)	Gene annotation (MaGe)	Primers	Sequences (5'→3')	Length (bp)	LASTn hit with non- <i>X. euroxanthae</i> (E value / Query coverage)
XEA1	XE424_v1_a0582XTG_000508		Conserved protein of unknown function	XEA1F XEA1R	CTGCCGAGCGTGAAATCCAG CCTTCAGTTGCACCGAACGC	819	-
XEA2	XE424_v1_a2605XTG_002379		Conserved protein of unknown function	XEA2F XEA2R	AGTCCACCAATGCCATCGCC AGTCCACCAATGCCATCGCC	425	-
XEA3	XE424_v1_a2606XTG_002380		Conserved protein of unknown function	XEA3F XEA3R	CGGATCGGACAATGACGCTG GCTCTACATCGCCGCTGGAG	612	-
XEA4	XE424_v1_a1415XTG_001287		TetR/AcrR family transcriptional regulator	XEA4F XEA4R	GACGCATCCGCCACGACC TAGGCGGCAGACCCCTTCC	341	<i>Dickeya zeae</i> A586-S18-A17 (2e-50 / 95%)
XEA5	XE424_v1_a0462XTG_000401		MarR family transcriptional regulator	XEA5F XEA5R	AACGACGCTGACCTGGACC CGACACCGCACGACCCCG	295	<i>Sphingomonas</i> sp. AP4-R1 (5e-13 / 73%)
XEA6	XE424_v1_a0617XTG_000542		Conserved protein of unknown function	XEA6F XEA6R	GCGGCTGCAGCGTCGTTG TCACCTGATGATCGAAGCCTGG	237	<i>Xanthomonas</i> sp. GW (2e-04 / 19%)
XEA7	XE424_v1_a1414n/a		Protein of unknown function	XEA7F XEA7R	GGACGCGCCATGATCTGCC GGTGTCCGAGGTCAGGTGC	212	-
XEA8	²	²	²	XEA8F XEA8R	ATCGCCTCTGGATGACGGC GGTGATGTCTGGCAAGCTCG	648	<i>Dickeya zeae</i> A586-S18-A17 (1e-95 / 73%)

Selected *Xanthomonas euroxanthae*-specific DNA markers (XEA1 to XEA8), corresponding primer pair sequences, expected amplicon sizes, and best BLASTn hits of amplicons with non-*X. euroxanthae* genomes.

n/a: not available.

¹ - no significant hit has been found.

² - XEA8 DNA marker was designed within the genomic subsequent and partially overlapping CDSs for markers XEA7 and XEA4.

2.4. PCR detection limit

The detection limit of the multiplex PCR was determined using 10 µL from each of the 10-fold dilutions of *X. euroxanthea* CPBF 424^T chromosomal DNA prepared in distilled sterile water, ranging from 100 ng to 1 pg per PCR reaction. Multiplex PCR conditions were kept as described above.

2.5. Typing potential of *X. euroxanthea*-specific DNA markers

Unrooted trees using XEA5, XEA6 and XEA8 markers (transversal to all strains of *X. euroxanthea*) (Supplementary Table V.S2) and partial sequence analysis of the housekeeping genes *acnB*, *fyuA*, *gyrB* and *rpoD* (Supplementary Table V.S3) of 11 *X. euroxanthea* strains were built to infer the typing potential of these DNA markers. Markers XEA5 (295 bp), XEA6 (237 bp) and XEA8 (648 bp) sequences were concatenated using the Geneious® v. 9.1.7 and used to build a maximum-likelihood tree based on Tamura-Nei model on MEGA X (Kumar *et al.*, 2018), as previously described (Fernandes *et al.*, 2017). The nucleotide sequences of housekeeping genes *acnB*, *fyuA*, *gyrB* and *rpoD* were retrieved from the 11 *X. euroxanthea* genomes, aligned and trimmed to 513 bp, 640 bp, 828 bp and 793 bp, respectively, and subsequently concatenated using the Geneious® v. 9.1.7 to build a maximum-likelihood tree as described for the markers. Since these three markers are *X. euroxanthea* specific, and no homologous could be found to define a coherent outgroup, the relatedness of the *X. euroxanthea* strains was inferred by an unrooted tree.

3. Results

3.1. *In silico* selection of DNA markers for *X. euroxanthea*

The selection of *X. euroxanthea*-specific DNA markers was carried out according to the workflow detailed in Figure V.1, which consisted on the use of a platform of comparative genomics (Mage) to screen for unique *X. euroxanthea* CDSs, further validated by BLASTn.

A list of CDSs exclusively present in the genomes of 11 *X. euroxanthea* strains, and absent from other 12 *Xanthomonas* species, including the closely related *X. arboricola*, was deduced from synteny analysis between the 24 genomes (Table V.1) along the full length of *X. euroxanthea* CPBF 424^T genome used as reference (Supplementary Figure V.S1). The candidate CDSs (from the reference genome of *X. euroxanthea* CPBF 424^T) were submitted to a BLASTn to confirm their exclusivity in *X. euroxanthea* considering high stringency values of E-value, percentage of identity and query coverage, and their

suitability as *X. euroxanthea*-specific DNA markers appraised by the design of robust primers. The data showed that, four CDSs had no significant BLAST hits; three CDSs showed no hits with other *Xanthomonas* sp., and solely one CDS coding for a putative conserved protein of unknown function matched with *Xanthomonas* sp. GW (E-value of 2e-04, and query coverage of 19%) (Supplementary Table V.S3). These seven CDSs were selected for the design of eight DNA markers, designated XEA1 to XEA8; being that two consecutive and partly overlapping CDSs (XE424_v1_a1415 and XE424_v1_a1414) were used to design marker XEA8 (Supplementary Table V.S3). Five of these CDSs are predicted to be proteins of unknown function, and two to be transcriptional regulators (Supplementary Table V.S3). Ultimately, five DNA markers, XEA4 to XEA8, are unfailingly present in the genome of the 11 *X. euroxanthea* strains analyzed; XEA1 is present in eight genomes; and XEA2 and XEA3 are present in five strains (Figure V.2).

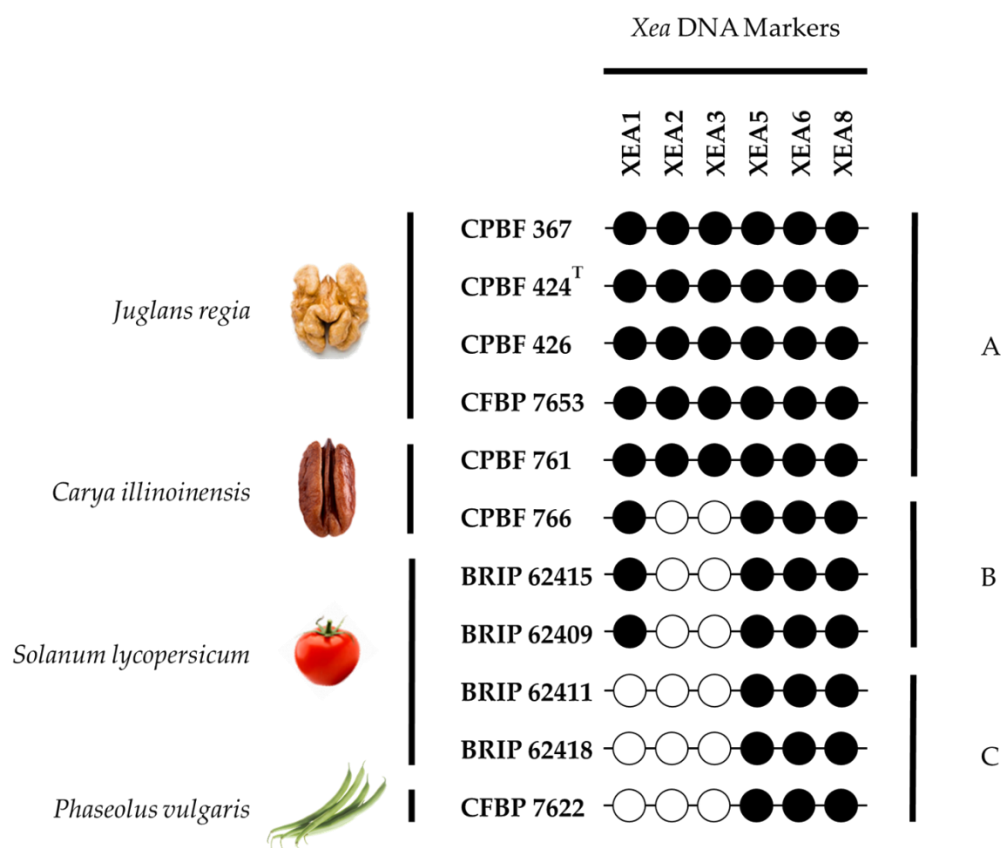


Figure V.2. Distribution of six *Xanthomonas euroxanthea* (*Xea*)-specific DNA markers (XEA1, XEA2, XEA3, XEA5, XEA6 and XEA8) in 11 *X. euroxanthea* genomes. The presence/absence of six XEA DNA markers was assessed by BLASTn analysis in Geneious, allowing to disclose three patterns, A to C, that do not translate strain-host affinities. (An updated version of figure is included in Chapter VI – Supplementary Figure VI.S1)

3.2. Genomic analysis unearths the stability of XEA DNA markers

To further assess the uniqueness of these *X. euroxantha*-specific CDSs used for DNA marker design, several features, including SNPs number, CAI/eCAI values, GC content, chromosomal location and chromosomal proximity to genomic plasticity-determinants were surveyed (Figure V.3).

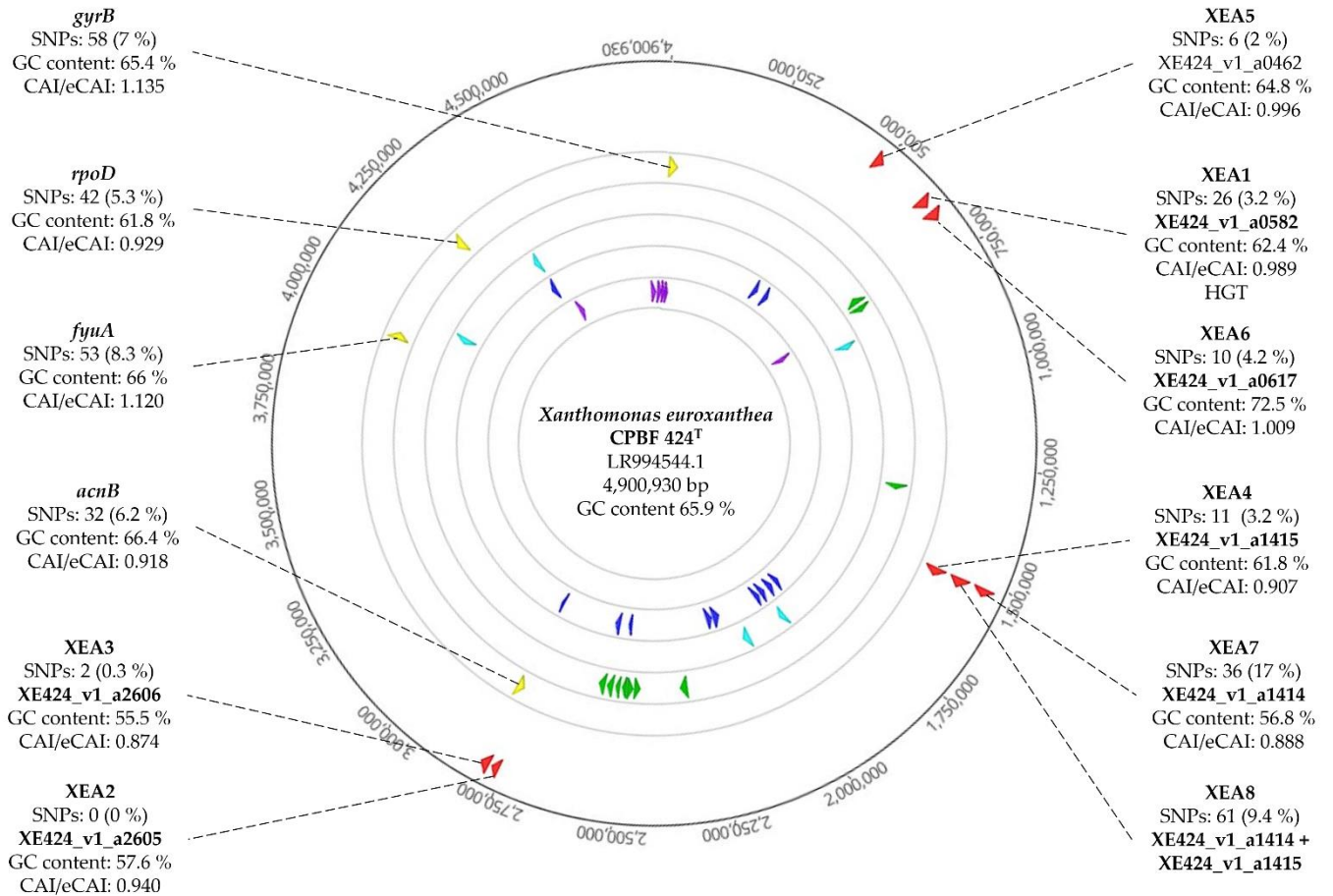


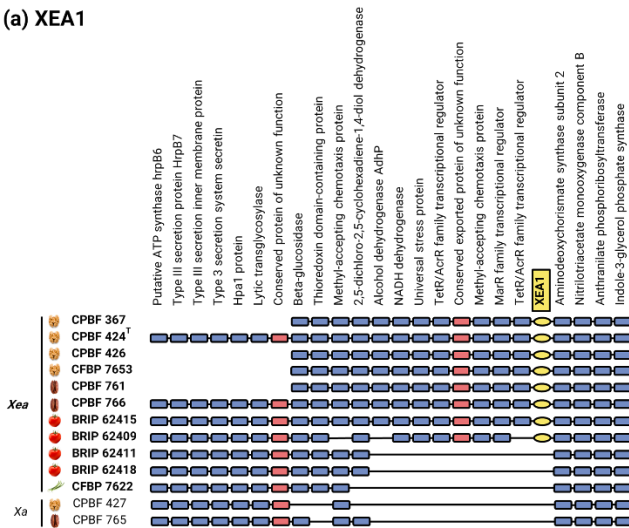
Figure V.3. Circular map of *Xanthomonas euroxantha* strain CPBF 424^T chromosome. Outside to inner circles are showing genome coordinates (bp); *X. euroxantha*-specific DNA markers XEA1 to XEA8 (red); housekeeping genes *gyrB*, *rpoD*, *fyuA* and *acnB* (yellow); transposases (green); recombinases (light blue); integrases (dark blue) and phage-related ORFs (purple). For each XEA DNA marker and housekeeping gene the number of SNPs (calculated based on 11 genomes of *X. euroxantha*), GC content and CAI/eCAI values are shown.

Within the 11 *X. euroxantha* genomes considered in this study, the number of SNPs for each marker and housekeeping gene was determined and normalized to the full length of the sequence (i.e. DNA marker or housekeeping gene) as a percentual value indicative of the allelic diversity of the sequences (Figure V.3). CAI/eCAI values of CDSs range between 0.874 (for XE424_v1_a2606 corresponding to DNA marker XEA3) to 1.009 (for XE424_v1_a0617 for DNA marker XEA6); which are values similar to the range attained with housekeeping genes, namely, 0.918 (*acnB*) to 1.135 (*gyrB*) (Figure

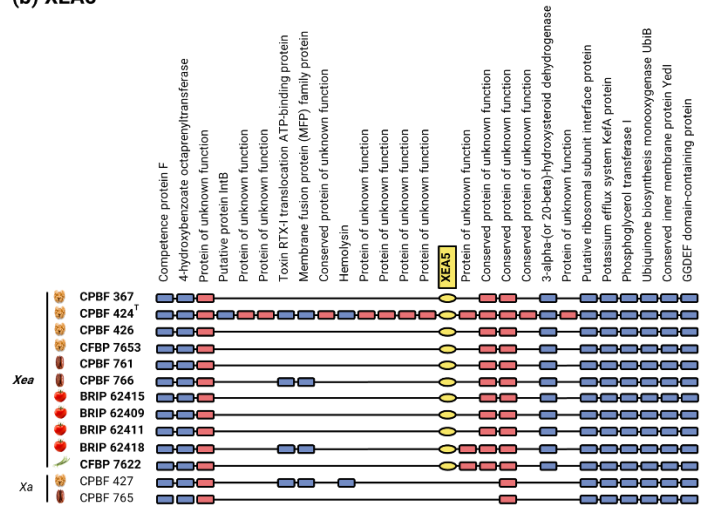
V.3). Also, GC content values for *X. euroxantha*-specific CDSs varied between 55.5 % (XE424_v1_a2606 XEA3) to 72.5 % (XE424_v1_a0617_XEA6), which parallels with *X. euroxantha* CPBF 424^T genome GC content value of 65.9 % (Figure V.3). Moreover, chromosomal location studies revealed that the XEA markers are scattered throughout the first half of the chromosome, and not in the vicinity of genomic plasticity-determinants, including, transposases, recombinases, integrases and phage related ORFs (Figure V.3).

The synteny analysis performed with MaGe allowed to investigate the genomic context of the most representative DNA markers, i.e. XEA1, XEA5, XEA6 and XEA8, across 11 *X. euroxantha* genomes and comparatively to *Xanthomonas arboricola* strains CPBF 427 (*X. arboricola* pv. *juglandis*) and CPBF 765 (Figure V.4), and other *Xanthomonas* spp. representative strains (Supplementary Figure V.S1).

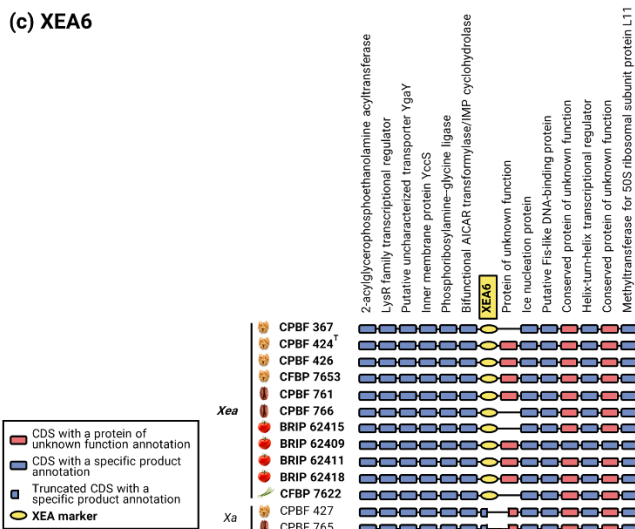
(a) XEA1



(b) XEA5



(c) XEA6



(d) XEA8

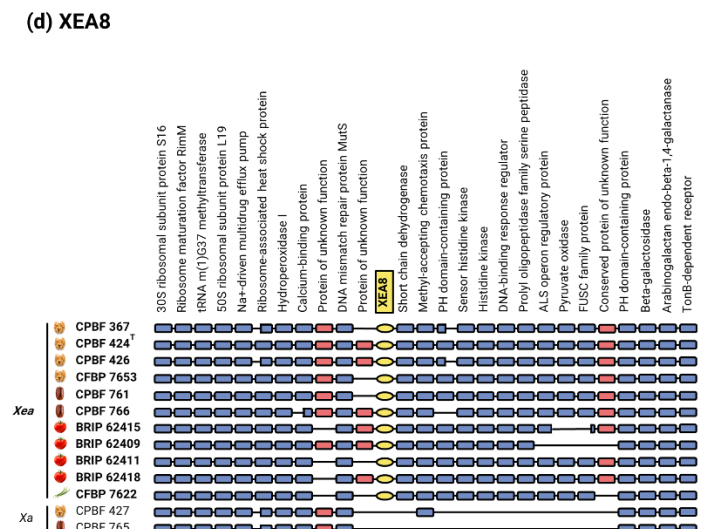


Figure V.4. Comparative syntenic maps of four *Xanthomonas euroxantha*-specific DNA marker-harboring-regions (a) XEA1 (designed from a conserved protein of unknown function sequence), (b) XEA5 (design from a MarR family transcriptional regulator), (c) XEA6 (designed from a conserved protein of unknown function sequence) and (d) XEA8 (designed from a protein of unknown function, and a TetR/AcrR family transcriptional regulator sequences) DNA markers across 11 *X. euroxantha* (Xea) and two *Xanthomonas arboricola* (Xa) strains.

The data showed that XEA1, XEA5, XEA6 and XEA8 are located in highly syntenic regions across all the *X. euroxantha* genomes studied, underlining the absence of genomic rearrangements (Figure V.4). Furthermore, CDSs used for the design of markers XEA5, XEA6 and XEA8 are exclusive of *X. euroxantha*, being either absent (CDSs encompassing XEA5 and XEA8) or truncated (CDS encompassing XEA6) in *X. arboricola* strains (Figure V.4). Particularly, the region upstream and including marker XEA1 appears to have suffered erosion in the *X. arboricola* strains analyzed (CPBF 427 and CPBF 765) and in three of the *X. euroxantha* strains studied (BRIP 62411, BRIP 62418 and CFBP 7622). In parallel, this scenario is observed for the region upstream of XEA5 for all strains, except for CPBF 424^T; and downstream of XEA8 for the non-*X. euroxantha* strains. It is further noticeable that within the flanking regions of XEA markers, the CDSs annotated as unknown proteins are the ones which presence/absence across strains is inconsistent, suggesting that they have been decaying. While for XEA6 and XEA8 markers these events are limited to a single CDS immediately flanking the markers, for XEA5 this paradigm is particularly clear as several CDSs annotated as unknown proteins are present in CPBF 424^T, have been lost by the other strains.

3.3. Multiplex PCR allows for the confident identification of *X. euroxantha* strains

Reliable identification of *X. euroxantha* isolates by multiplex PCR was optimized for DNA markers XEA1, XEA5 and XEA8, as these originate amplicons of distinguishable size (819 bp, 295 bp and 648 bp, respectively) and are located in three independent genomic regions of *X. euroxantha*. The three markers were successfully amplified in all *X. euroxantha* strains analyzed, with exception for XEA1 in *X. euroxantha* CFBP 7622 strain (Figure V.5) as expected by *in silico* studies. In addition, no amplification was observed for any of the other 28 xanthomonads, namely 17 strains of *X. arboricola*, including different pathovars, and 9 non-*arboricola* *Xanthomonas* species.

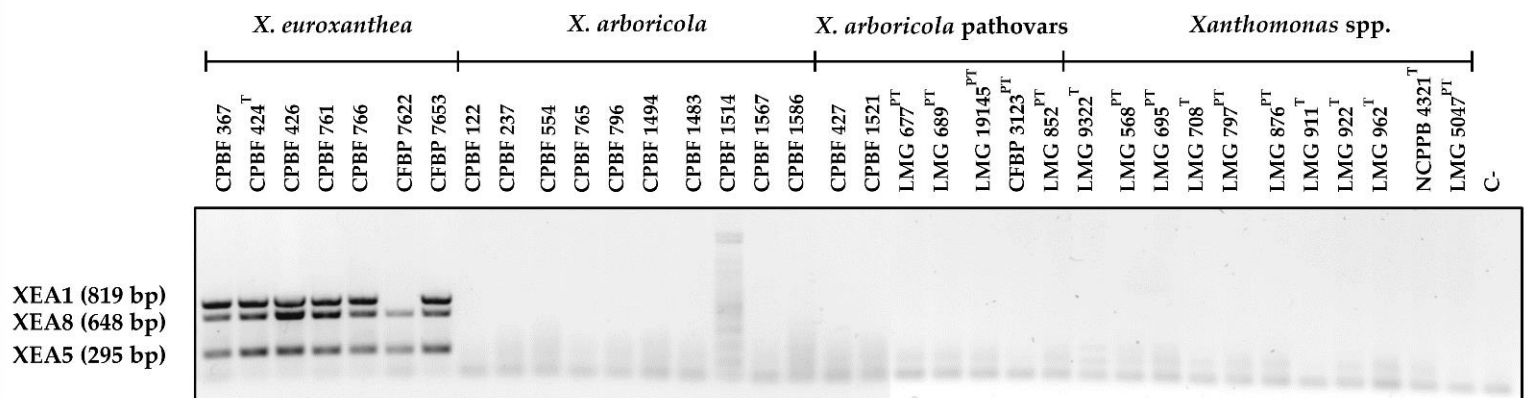


Figure V.5. Multiplex PCR using *Xanthomonas euroxantha*-specific DNA markers XEA1 (819 bp), XEA8 (648 bp) and XEA5 (295 bp) on 7 *X. euroxantha* strains, 10 *Xanthomonas arboricola* strains, 6 pathovars of *Xanthomonas arboricola*, and 9 non-*arboricola* *Xanthomonas* species. Markers XEA5 and XEA8 were successful in detecting *X. euroxantha* strains, while XEA1 identified all *X. euroxantha* strains, except for CFBP 7622. No amplification was observed for any of the other xanthomonads tested, namely *X. arboricola* and other *Xanthomonas* species. C-: negative control.

3.4. Detection limit of multiplex PCR with XEA DNA markers

The detection limit of the multiplex PCR targeting the *X. euroxantha*-specific markers XEA1, XEA5, and XEA8, determined through serial dilution of chromosomal DNA, was 1 ng for PCR reaction (Supplementary Figure V.S2). When assessing the PCR detection limit of each marker individually, while for XEA1 and XEA8 the detection limit was identical to the multiplex PCR, i.e. 1 ng/PCR reaction, it was ascertained that the limit of detection lowered to 100 pg/PCR reaction for XEA5.

3.5. Typing potential of informative XEA DNA markers

The concatenated sequences of XEA5, XEA6 and XEA8 (1180 bp), for each *X. euroxantha* strain studied, were aligned and used to generate a maximum-likelihood tree to investigate the discriminatory potential of these *X. euroxantha*-specific markers comparatively to four housekeeping genes, corresponding to a concatenated sequence length of 2774 bp, (*acnB*, *fyuA*, *gyrB* and *rpoD* genes) commonly used for MLSA (Figure V.6).

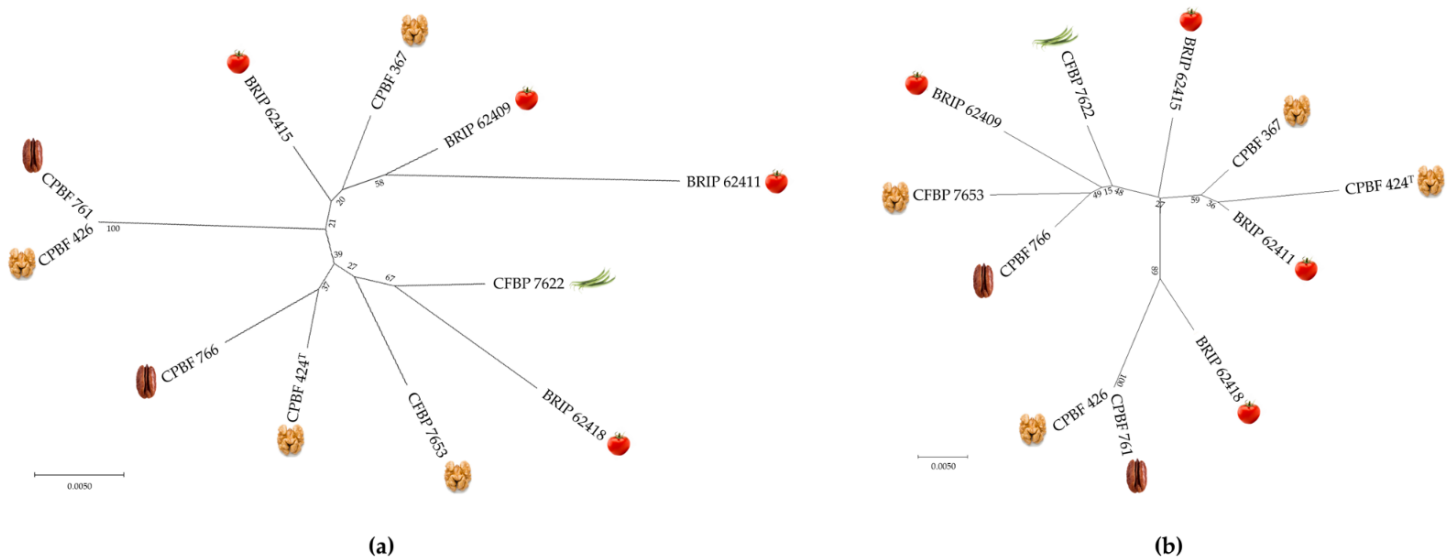


Figure V.6. Maximum-likelihood phylogenetic tree based on concatenated sequences of (a) DNA markers XEA5, XEA6 and XEA8 (1180 bp); and (b) partial housekeeping gene sequences for *acnB*, *fyuA*, *gyrB* and *rpoD* (2774 bp) extracted from 11 *X. euroxantha* genomes. The tree was constructed using the Tamura-Nei model using MEGA X (Kumar et al., 2018). Supporting values from 1000 bootstrap replicates are indicated near nodes.

Attending that these XEA detection markers are highly specific to *X. euroxantha*, no homologous sequences have been found in other bacterial taxa, and therefore the allelic diversity determined by the number of SNP (6 SNP/295nt for XEA5; 10 SNP/237nt for XEA6; and 61 SNP/648 nt for XEA8, Figure V.3) were represented by an unrooted Maximum-likelihood tree. For both trees, each *X. euroxantha* strain is represented by an independent tree branch, with exception for strains CPBF 426 and CPBF 761 that clustered together in a single branch. Furthermore, the topology of both trees does not reflect any clustering according to plant-host of isolation, and pathogenicity or non-pathogenicity phenotypes (Figure V.6).

4. Discussion

The development of a reliable and efficient method, for the accurate detection and identification of *X. euroxantha* strains and concomitant differentiation from the closely related *X. arboricola* strains that frequently colonize the same host plants, is essential to investigate the distribution and role played by each of these *Xanthomonas* species in the common hosts (Fernandes *et al.*, 2021a). Additionally, early detection is a critical first step towards timely sanitary intervention aimed at eradicating the pathogen and at

reducing the inoculum spread to other plants; thus lessening the disease-induced damage in crops from growth to postharvest processing of products, and ensuring agricultural sustainability (Fang & Ramasamy, 2015). Likewise, discriminating one bacterial species from the other would be rendered useful in the application of suitable management procedures.

Recommended standard diagnostic protocols from OEPP/EPPO for other plant-diseases caused by *Xanthomonas* spp. (namely, *X. arboricola* pv. *corylina*, *X. axonopodis* pv. *dieffenbachiae*, *X. arboricola* pv. *pruni*, *X. axonopodis* pv. *citri*, *X. fragariae*, *X. oryzae*, *X. axonopodis* pv. *allii*, *X. euvesicatoria*, *X. hortorum* pv. *gardneri*, *X. perforans* and *X. vesicatoria*) still require several steps - observation of disease symptoms, microscopic examination, pathogen isolation, pathogenicity tests and molecular tests ("PM 7/022 (1) *Xanthomonas arboricola* pv. *corylina*," 2004; "PM 7/044 (1) *Xanthomonas axonopodis* pv. *citri*," 2005; "PM 7/065 (1) *Xanthomonas fragariae*," 2006; "PM 7/080 (1) *Xanthomonas oryzae*: Diagnostics," 2007; "PM 7/023 (2) *Xanthomonas axonopodis* pv. *dieffenbachiae*," 2009; "PM 7/110 (1) *Xanthomonas* spp. (*Xanthomonas euvesicatoria*, *Xanthomonas gardneri*, *Xanthomonas perforans*, *Xanthomonas vesicatoria*) causing bacterial spot of tomato and sweet pepper," 2013; "PM 7/128 (1) *Xanthomonas axonopodis* pv. *allii*," 2016; "PM 7/064 (2) *Xanthomonas arboricola* pv. *pruni*," 2021). Having in mind that only recently *X. euroxanthea* was proposed as a new species (Martins *et al.*, 2020), isolated from different plant hosts (Roach *et al.*, 2018; Fernandes *et al.*, 2021b; Kałużna *et al.*, 2021; Martins *et al.*, 2022), for which the nature of the bacteria-plant interaction is still unknown and no distinct symptoms have been described (Martins *et al.*, 2020), the herein proposed multiplex PCR is the tool available for accurate detection and identification of *X. euroxanthea*.

While multiplex PCRs have been proposed to detect *X. arboricola* pv. *juglandis* (Fernandes *et al.*, 2017) or *X. arboricola* pv. *pruni* (Pothier *et al.*, 2011b), the present work describes eight *X. euroxanthea*-specific DNA markers (XEA1 to XEA8) and proposes a methodology to detect and genotype *X. euroxanthea* strains. In view of that, using a comparative genomics strategy and by assessing a pool of numerous xanthomonads genomes, including several *X. euroxanthea* and *X. arboricola*, it was possible to narrow down *X. euroxanthea*-specific genomic regions to only seven CDSs. Genomes alignment is emerging as fast and convenient solution to rapidly identify species-specific DNA markers to implement PCR-based detection methods, as recently evidenced for *Xanthomonas campestris* pv. *raphani* (Inoue *et al.*, 2021).

The specificity of these putative *X. euroxanthea*-specific DNA markers were further validated by BLASTn analysis using *X. euroxanthea* type strain CPBF 424^T as query sequence, followed by an *in silico* workflow essentially as described in previous studies

(Albuquerque *et al.*, 2012; Almeida *et al.*, 2013), to determine the genomic context of each *X. euroxanthea*-specific loci, and particularly to ensure that these putative DNA markers are within conserved and stable genomic regions. Specifically, four CDSs showed no significant BLASTn hits outside of *X. euroxanthea*, three CDSs showed some similarity with non *Xanthomonas* species, and solely one CDS matched with a *Xanthomonas* spp. but with a poor similarity as shown by an E-value of 2e-04 and query coverage of 19%. Based on these data, eight primer pairs corresponding to eight markers (XEA1 to XEA8), were designed to nest in these seven CDSs. Marker XEA8 was designed from two consecutive and partly overlapping CDSs. Furthermore, the chromosomal distance between each of the *X. euroxanthea* DNA markers (XEA1 to XEA8) and genomic plasticity-determinants such as transposases, recombinases, integrases and phage-related ORFs attests to the low genomic plasticity and high stability of the DNA marker-harboring-regions as previously suggested (Albuquerque *et al.*, 2012). All seven specific CDSs on which XEA markers are founded, seem to be well adapted to the codon usage and GC content, suggesting that these CDSs were not recently acquired by Horizontal Gene Transfer – HGT (Puigbò *et al.*, 2008b; Albuquerque *et al.*, 2012). Such findings are consistent with the hypothesis that these CDSs are conserved in the *X. euroxanthea* genomes, and likely present across all the *X. euroxanthea* strains regardless of infrasubspecific variability. A BLAST analysis of each DNA-marker (XEA1 to XEA8) against the genomes of 11 *X. euroxanthea* showed that five DNA markers (XEA4-XEA8) are unfailingly present in the genome of the 11 *X. euroxanthea* strains analyzed. One marker (XEA1) is absent in three out of the 11 *X. euroxanthea* genomes analyzed, and markers XEA2 and XEA3 are present in the four *X. euroxanthea* strains isolated from walnut (CPBF 367, CPBF 424^T, CPBF 426, and CFBP 7653), and in one strain isolated from pecan (CPBF 766). These results suggest that the CDSs used for the design of markers XEA4-XEA8 are within the core genome of the 11 *X. euroxanthea* studied, whereas the CDSs corresponding to XEA1-XEA3 markers, although *X. euroxanthea*-specific, are part of the accessory genome (Segerman, 2012).

An interesting attribute of *X. euroxanthea* is the encompassment of pathogenic (CPBF 424^T), and non-pathogenic strains (CPBF 367) to the same host (Martins *et al.*, 2020). Considering the pathogenicity and non-pathogenicity phenotypes; different plant hosts of isolation; and the occurrence of markers XEA1 to XEA8 across the 11 studied *X. euroxanthea*, one may conclude that these *X. euroxanthea*-specific markers are not biased by pathogenicity phenotype or plant host species. To determine the allelic variation of the markers and infer their potential to discriminate *X. euroxanthea* strains, the number of SNPs of markers shown to be present in the 11 *X. euroxanthea* strains

(i.e. XEA5, XEA6 and XEA8) was determined. The data indicate that XEA8 with a 9.4% of SNPs (61 SNP/648 bp) stands as the most informative *X. euroxanthea*-specific DNA marker for typing purposes, as it surpasses the 2% and 4.2% of SNPs recorded for XEA5 and XEA6 markers respectively (6 SNP/295 bp for XEA5; and 10 SNP/237 bp for XEA6). Although the allelic variation of XEA5 and XEA6 may be inferior to the allelic variation observed for the housekeeping genes commonly used for MLSA (*acnB* (6.2%), *fyuA* (8.3%), *gyrB* (7%) and *rpoD* (5.3%)), when combined with XEA8 it is nevertheless enough to separate the 11 *X. euroxanthea* strains as efficiently as the housekeeping genes. These results suggest that these three *X. euroxanthea*-specific DNA markers (XEA5, XEA6 and XEA8) may be used both for the detection and genotyping of *X. euroxanthea*. Since the allelic variation was carried out considering 11 genomes of *X. euroxanthea*, the number of SNPs may be underestimated, i.e. novel *X. euroxanthea* strains may likely reveal novel single nucleotide substitutions in the sequence of the mentioned markers.

The synteny analyses conducted to investigate the genomic context of the XEA markers across 11 *X. euroxanthea* and two strains of the closely related species *X. arboricola*, revealed that XEA markers and their flanking CDSs are syntenic for *X. euroxanthea*. It is worth mentioning that within the flanking regions of XEA markers, CDSs annotated as unknown proteins are intermittently absent when compared to CDSs of proteins of known function, suggesting genomic erosion. This paradigm is further supported by the genomic context of XEA5 marker where two CDSs, namely, the toxin RTX-I translocation ATP-binding protein and the membrane fusion protein (MFP) family protein remain present in CPBF 424^T, CPBF 766, BRIP 62418 and CPBF 427 despite the overall loss of the flanking unknown proteins. The fact that the CDS of XEA6 marker is present in *X. euroxanthea* strains and truncated in *X. arboricola* strains, suggests a common and recent ancestry of these two species.

Distinctively, markers designed from CDSs annotated as a family transcriptional regulator, namely XEA5 and XEA8, are likely to portray an essential role in cell life, and consequently remain conserved in *X. euroxanthea*, which is further supported by their inclusion as part of the core genome.

Overall, the data gathered regarding synteny, CAI/eCAI values, and the GC content of *X. euroxanthea*-specific CDSs concerning markers XEA1, XEA5, XEA6 and XEA8, suggest that these markers and their flanking regions are not the result of recent HGT acquisitions events by *X. euroxanthea*, which increases its reliability for *X. euroxanthea* detection and identification, as has been hypothesized for other bacteria (Albuquerque *et al.*, 2012; Almeida *et al.*, 2013).

A multiplex PCR detection assay targeting XEA1, XEA5 and XEA8 markers was successfully developed for the identification of *X. euroxanthea*. The annealing temperature and extension time were optimized to steer clear of non-specific amplifications of DNA, which was also reportedly been done in previous studies [40]. This multiplex PCR was shown to be both specific and efficient given that neither non-specific amplifications (bands of unexpected sizes) in *X. euroxanthea* nor amplification in non-target *Xanthomonas* spp. tested were observed. As predicted by the *in silico* assessments, the amplicons corresponding to XEA1, XEA5 and XEA8 markers were obtained for all *X. euroxanthea* strains, with exception of XEA1 for *X. euroxanthea* CFBP 7622. These results emphasize the robustness of the multiplex PCR for the detection and identification of *X. euroxanthea*.

When assessing the PCR detection limit of each marker individually, it was observed that XEA5 lowered the limit of detection value to 10 pg/ μ L, indicating that the sensitivity of the simplex PCR targeting XEA5 marker is ten fold higher than the multiplex PCR. The detection limit of 0.1 ng/ μ L attained for the multiplex PCR is similar to what has been reported in other studies (0.02 ng/ μ L and 0.5 ng/ μ L) (Fonseca *et al.*, 2019; Bangratz *et al.*, 2020).

To investigate the genotyping potential of the chosen markers, the allelic variation of the three markers present in all of the studied *X. euroxanthea* (XEA5, XEA6 and XEA8), was compared with four housekeeping genes commonly used for MLSA (*acnB*, *fyuA*, *gyrB* and *rpoD*) and represented as unrooted trees. The data showed that, for both trees, each strain is allocated to an independent tree branch; except for strains CPBF 426 and CPBF 761 that clustered together in a single branch attending that no SNPs were observed neither within the concatenated sequences of the XEA markers (1180 bp), nor the partial housekeeping genes (2774 bp).

The trees topology reveals that markers and housekeeping genes are similarly informative in discriminating *X. euroxanthea* strains. Furthermore, the scattered tree branches indicate that the studied *X. euroxanthea* strains are not clustered according to plant-host of isolation, pathogenicity or non-pathogenicity phenotypes.

By comparing marker's presence profiles to plant host species from which the *X. euroxanthea* strains were isolated, three patterns identified as A, B, and C (Figure V.2) are observed. Attending that all XEA markers are in syntenic genomic regions that overlap with homologs regions in the closely related *X. arboricola* strains, we may hypothesize that the *X. euroxanthea* strains isolated from two Juglandaceae species (*Juglans regia* and *Carya illinoensis*) with pattern A, i.e. possessing all XEA markers (XEA1 to XEA8), are ancestors of strains included in pattern B (i.e. which lost XEA2 and XEA3) and isolated from *Carya illinoensis* and Solanaceaceae (*Solanum*

lycopersicum), and pattern C (i.e. which lost XEA1, XEA2 and XEA3), isolated from *Solanum lycopersicum* and a Fabaceae (*Phaseolus vulgaris*). Thus, rather than an acquisition of XEA2, and XEA3 markers in five of the eleven studied *X. euroxanthea* strains (pattern A), these markers were likely lost in the other *X. euroxanthea* strains, followed by XEA1 marker shown to be absent in 3 out the 11 *X. euroxanthea* strains, leading to the emergence of recent *X. euroxanthea* lineage characterized by marker-profile C (BRIP 62411, BRIP 62418) and *Phaseolus vulgaris* (CFBP 7622). These data suggest that the lost of markers XEA2, XEA3, and XEA1 i.e. patterns A-C, occurred progressively as *X. euroxanthea* lineages extended to a new host, specifically, from walnut, to pecan, to tomato, to the common bean. These results are aligned with studies describing events of genome erosion as a consequence of bacterial adaptation to a new plant-host (Brown & Wernegreen, 2019; Armbruster *et al.*, 2021). To sum up we may infer that the evolutionary trend of XEA marker's loss follows progressive host-jump events from walnut, to pecan, to tomato, and finally to common bean.

5. Conclusions

The present study proposes specific, efficient and reliable eight DNA markers for detection and identification of *X. euroxanthea* isolates. The allelic variation of some of these markers allow to conciliate detection and genotyping of *X. euroxanthea* strains, contributing to survey these bacteria in ecological niches colonized by the closely related *X. arboricola*. The multiplex PCR was shown to be highly specific, as solely target DNA (i.e. *X. euroxanthea*) was amplified; and efficient, as an amplicon was observed with all tested *X. euroxanthea* strains. The present study also provided a successful workflow for the selection of molecular markers, able to be implemented in the selection of species-specific-genomic regions for any other taxa. By analyzing marker's presence across strains colonizing different plant host species we may infer that *X. euroxanthea* colonization of different plant host species occurred at different moments.

Author Contributions

Conceptualization, F.T., J.F.P., and L.M.; validation, writing—original draft preparation, formal analysis, K.G.S., F.T., and L.M.; investigation, K.G.S., L.M. and M.T.; writing—review and editing, K.G.S., F.T., J.F.P., L.M. and M.T.; supervision, F.T. and L.M.; project administration, funding acquisition, resources F.T.; All authors have read and agreed to the published version of the manuscript.

Funding

Funding for this work was provided by the EU Horizon 2020 Research and Innovation program (grant agreement 857251), European Structural & Investment Funds (ESIFs) through COMPETE 2020, National Funds through FCT (in the scope of the projects EVOXANT-PTDC/BIA-EVF/3635/2014-POCI-01-0145-FEDER-016600; and UIDB/50027/2020), and the Operational Thematic Program for Competitiveness and Internationalization (POCI) under the PORTUGAL 2020 Partnership Agreement through the European Regional Development Fund (FEDER). L.M. was supported by fellowship from FCT (SFRH/BD/137079/2018). J.F.P. acknowledges support from the Department of Life Sciences and Facility Management of the Zurich University of Applied Sciences (ZHAW) in Wädenswil. This article is based upon work from COST Action CA16107 EuroXanth, supported by COST (European Cooperation in Science and Technology). The APC was funded by EU's Horizon 2020 Research and Innovation program through BIOPOLIS (grant agreement 857251).

References

- Albuquerque P, Caridade CMR, Rodrigues AS *et al.*, 2012. Evolutionary and experimental assessment of novel markers for detection of *Xanthomonas euvesicatoria* in plant samples. *PLoS ONE* **7**.
- Almeida A, Albuquerque P, Araujo R, Ribeiro N, Tavares F, 2013. Detection and discrimination of common bovine mastitis-causing streptococci. *Veterinary Microbiology* **164**, 370–377.
- Armbruster CR, Marshall CW, Garber AI *et al.*, 2021. Adaptation and genomic erosion in fragmented *Pseudomonas aeruginosa* populations in the sinuses of people with cystic fibrosis. *Cell Reports* **37**.
- Bangratz M, Wonni I, Kini K *et al.*, 2020. Design of a new multiplex PCR assay for rice pathogenic bacteria detection and its application to infer disease incidence and detect co-infection in rice fields in Burkina Faso. *PLoS ONE* **15**.
- Brown BP, Wernegreen JJ, 2019. Genomic erosion and extensive horizontal gene transfer in gut-associated Acetobacteraceae. *BMC Genomics* **20**, 1–15.
- Catara V, Cubero J, Pothier JF *et al.*, 2021. Trends in Molecular Diagnosis and Diversity Studies for Phytosanitary Regulated *Xanthomonas*. *Microorganisms* 2021, Vol. 9, Page 862 **9**, 862.
- Fang Y, Ramasamy RP, 2015. Current and prospective methods for plant disease detection. *Biosensors* **5**, 537–561.
- Fernandes C, Albuquerque P, Mariz-Ponte N, Cruz L, Tavares F, 2021a. Comprehensive diversity assessment of walnut-associated xanthomonads reveal the occurrence of distinct *Xanthomonas arboricola* lineages and of a new species (*Xanthomonas euroxanthea*) within the same tree. *Plant Pathology* **70**, 5.
- Fernandes C, Albuquerque P, Sousa R, Cruz L, Tavares F, 2017. Multiple DNA Markers for Identification of *Xanthomonas arboricola* pv. *juglandis* Isolates and its Direct Detection in Plant Samples. *Plant Disease* **101**, 858–865.
- Fernandes C, Blom J, Pothier JF, Tavares F, 2018. High-Quality Draft Genome Sequence of *Xanthomonas* sp. Strain CPBF 424, a Walnut-Pathogenic Strain with Atypical Features. *Microbiology Resource Announcements* **7**.

- Fernandes C, Martins L, Teixeira M *et al.*, 2021b. Comparative Genomics of *Xanthomonas euroxanthea* and *Xanthomonas arboricola* pv. *juglandis* Strains Isolated from a Single Walnut Host Tree. *Microorganisms* 2021, Vol. 9, Page 624 **9**, 624.
- Fonseca NP, Felestrino ÉB, Caneschi WL *et al.*, 2019. Detection and identification of *Xanthomonas* pathotypes associated with citrus diseases using comparative genomics and multiplex PCR. *PeerJ* **2019**, 1–22.
- Hajri A, Meyer D, Delort F, Guillaumès J, Brin C, Manceau C, 2010. Identification of a genetic lineage within *Xanthomonas arboricola* pv. *juglandis* as the causal agent of vertical oozing canker of Persian (English) walnut in France. *Plant Pathology* **59**, 1014–1022.
- Inoue Y, Fujikawa T, Takikawa Y, 2021. Detection and identification of *Xanthomonas campestris* pv. *campestris* and pv. *raphani* by multiplex polymerase chain reaction using specific primers. *Applied Microbiology and Biotechnology* **105**, 1991–2002.
- Kaluźna M, Fischer-Le Saux M, Pothier JF *et al.*, 2021. *Xanthomonas arboricola* pv. *juglandis* and pv. *corylina*: Brothers or distant relatives? Genetic clues, epidemiology, and insights for disease management. *Molecular Plant Pathology* **22**, 1481–1499.
- Kearse M, Moir R, Wilson A *et al.*, 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K, 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**, 1547–1549.
- Loreti S, Pucci N, Perez G *et al.*, 2015. Detection and identification of *Xanthomonas arboricola* pv. *pruni* from symptomless plant material: Results of an Italian test performance study. *EPPO Bulletin* **45**, 41–51.
- Martins L, Fernandes C, Albuquerque P, Tavares F, 2019. Assessment of *Xanthomonas arboricola* pv. *juglandis* Bacterial Load in Infected Walnut Fruits by Quantitative PCR. *Plant disease* **103**, 2577–2586.
- Martins L, Fernandes C, Blom J, Dia NC, Pothier JF, Tavares F, 2020. *Xanthomonas*

euroxanthea sp. nov., a new xanthomonad species including pathogenic and non-pathogenic strains of walnut. *International Journal of Systematic and Evolutionary Microbiology* **70**, 9.

Martins L, Teixeira M, Fernandes C *et al.*, 2022. Genomic features of *Xanthomonas arboricola* and *X. euroxanthea* sharing the same plant host species preferences (walnut, pecan and tomato) seems unbiased by the host species.

Moragrega C, Matias J, Aletà N, Montesinos E, Rovira M, 2011. Apical necrosis and premature drop of Persian (English) walnut fruit caused by *Xanthomonas arboricola* pv. *juglandis*. *Plant Disease* **95**, 1565–1570.

Moragrega C, Özaktan H, 2010. Apical necrosis of persian (english) walnut *Juglans regia*): An update. *Journal of Plant Pathology* **92**.

Palacio-Bielsa A, Cubero J, Cambra MA, Collados R, Berruete IM, López MM, 2011. Development of an efficient real-time quantitative PCR protocol for detection of *Xanthomonas arboricola* pv. *pruni* in *Prunus* species. *Applied and Environmental Microbiology* **77**, 89–97.

Palacio-Bielsa A, López-Soriano P, Bühlmann A *et al.*, 2015. Evaluation of a real-time PCR and a loop-mediated isothermal amplification for detection of *Xanthomonas arboricola* pv. *pruni* in plant tissue samples. *Journal of Microbiological Methods* **112**, 36–39.

PM 7/022 (1) *Xanthomonas arboricola* pv. *corylina*, 2004. *EPPO Bulletin* **34**, 179–181.

PM 7/023 (2) *Xanthomonas axonopodis* pv. *dieffenbachiae*, 2009. *EPPO Bulletin* **39**, 393–402.

PM 7/044 (1) *Xanthomonas axonopodis* pv. *citri*, 2005. *EPPO Bulletin* **35**, 289–294.

PM 7/064 (2) *Xanthomonas arboricola* pv. *pruni*, 2021. *EPPO Bulletin* **51**, 240–266.

PM 7/065 (1) *Xanthomonas fragariae*, 2006. *EPPO Bulletin* **36**, 135–144.

PM 7/080 (1) *Xanthomonas oryzae*: Diagnostics, 2007. *EPPO Bulletin* **37**, 543–553.

PM 7/110 (1) *Xanthomonas* spp. (*Xanthomonas euvesicatoria*, *Xanthomonas gardneri*, *Xanthomonas perforans*, *Xanthomonas vesicatoria*) causing bacterial spot of tomato and sweet pepper, 2013. *EPPO Bulletin* **43**, 7–20.

- PM 7/128 (1) *Xanthomonas axonopodis* pv. *allii*, 2016. *EPPO Bulletin* **46**, 429–443.
- Pothier JF, Pagani MC, Pelludat C, Ritchie DF, Duffy B, 2011a. A duplex-PCR method for species- and pathovar-level identification and detection of the quarantine plant pathogen *Xanthomonas arboricola* pv. *pruni*. *Journal of Microbiological Methods* **86**, 16–24.
- Pothier JF, Vorhölter FJ, Blom J *et al.*, 2011b. The ubiquitous plasmid pXap41 in the invasive phytopathogen *Xanthomonas arboricola* pv. *pruni*: Complete sequence and comparative genomic analysis. *FEMS Microbiology Letters* **323**, 52–60.
- Puigbò P, Bravo IG, Garcia-Vallve S, 2008a. CAIcal: A combined set of tools to assess codon usage adaptation. *Biology Direct* **3**, 1–8.
- Puigbò P, Bravo IG, Garcia-Vallvé S, 2008b. E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinformatics* 2008 9:1 **9**, 1–7.
- Roach R, Mann R, Gambley CG, Shivas RG, Rodoni B, 2018. Identification of *Xanthomonas* species associated with bacterial leaf spot of tomato, capsicum and chilli crops in eastern Australia. *European Journal of Plant Pathology* **150**, 595–608.
- Segerman B, 2012. The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Frontiers in cellular and infection microbiology* **2**, 116.
- Stothard P, 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* **28**.
- Teixeira M, Martins L, Fernandes C *et al.*, 2020. Complete Genome Sequences of Walnut-Associated *Xanthomonas euroxanthea* Strains CPBF 367 and CPBF 426 Obtained by Illumina/Nanopore Hybrid Assembly (DA Baltrus, Ed.). *Microbiology Resource Announcements* **9**.
- Vallenet D, Calteau A, Dubois M *et al.*, 2020. MicroScope: An integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Research* **48**, D579–D589.
- Zarei S, Taghavi SM, Rahimi T *et al.*, 2022. Taxonomic Refinement of *Xanthomonas*

arboricola. <https://doi.org/10.1094/PHYTO-12-21-0519-R>.

Supplementary Material

The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/microorganisms10061078/s1>, Table V.S1. MaGe labels of the seven selected *X. euroxanthea*-specific CDSs of 11 *X. euroxanthea* genomes used to design XEA1 to XEA8 markers. Table V.S2. Genomic coordinates of the eight *X. euroxanthea*-specific DNA markers of 11 *X. euroxanthea* genomes. Table V.S3. MaGe labels of four housekeeping genes from 11 *X. euroxanthea* genomes used in the construction of an unrooted tree. Figure V.S1. Synteny map of DNA markers (a) XEA1, (b) XEA5, (c) XEA6 and (d) XEA8 across the genomes of 11 *X. euroxanthea* and 24 other *Xanthomonas* spp. strains. Figure V.S2. PCR detection limits assessed using purified DNA from CPBF 424T. C-: negative control (sterile distilled water).



Chapter VI

General Discussion

General Discussion

The evolutionary background of *Xanthomonas euroxanthea* is still a mystery and one that would be precious to unravel. Since its discovery, the study of this species has been an interesting journey. The cornerstone of this PhD project was the discovery of *Xanthomonas arboricola*-like isolates during an extensive diversity study of *Xanthomonas arboricola* pv. *juglandis* (Xaj) across walnut orchards throughout Portugal (2014-2017) (Fernandes *et al.*, 2021a). Further characterization of these strains through typing using Multilocus Sequence Analysis with four housekeeping genes *acnB*, *gyrB*, *fyuA* and *rpoD* divided all the isolates into ten clusters (I-X), from I to VI containing typical *Xanthomonas arboricola* pv. *juglandis* strains, VII to VIII encompassing *X. arboricola* and IX to X grouping *Xanthomonas* sp. (Fernandes *et al.*, 2021a). The genomes of three atypical *Xanthomonas* isolates from cluster X, CPBF 367, CPBF 424^T (Fernandes *et al.*, 2018a), and CPBF 426 were sequenced along with putative typical *X. arboricola* pv. *juglandis* CPBF 427 and CPBF 1521 (cluster VI) (Fernandes *et al.*, 2018b) using the short-read sequencing technology Illumina HiSeq platform.

Daring to be different - *X. euroxanthea* discovery story

The availability of whole genome sequence data of a multitude of *Xanthomonas* isolates has allowed a deeper insight into the phylogenetic relationship within the genus, including taxonomic revision and reclassifications (An *et al.*, 2019), owing particular importance to Average Nucleotide Identity (ANI) (Konstantinidis & Tiedje, 2005) to perform higher resolution taxogenomics of bacteria. This was of particular importance for our study because it provided us the data to perform whole genomic ANI comparisons over a representative sample of the *Xanthomonas* genus. This way, we disclosed that CPBF 367, CPBF 424^T and CPBF 426 could not be attributed to any known species, subsequently proposing *X. euroxanthea* as a new species (Chapter II - Martins *et al.*, 2020). The characterization of *X. euroxanthea* as a new species laid the groundwork for the identification of more strains belonging to this species, including the reclassification of previously wrongly-identified *Xanthomonas arboricola*-like strains and other *Xanthomonas* that were considered as *Xanthomonas* sp. or mistaken as *X. arboricola* (Chapter III - Fernandes *et al.*, 2021b; Kałużna *et al.*, 2021; Chapter IV - Martins *et al.*, 2022; Zarei *et al.*, 2022) (More on this subject further along in this chapter). Particularly, during an extensive isolation effort from symptomatic and asymptomatic walnut material to study walnut blight in Romagna region in Italy, Giovanardi *et al.*, (2016) identified morphological differences of colonies from strains identified as Xaj based on pathogenic ability in walnut

fruitlets. A yet considerable number of non-pathogenic strains, denominated as *Xaj*-like strains (59 from 136 total isolates) were isolated both from diseased and healthy walnut tissues and could not be easily distinguished from *Xaj* strains. It would be interesting to check if these *Xaj*-like strains and morphologically distinct *Xaj* could in fact belong to *X. euroxanthea* species, as strains from this species were also found together with *Xaj* and *X. arboricola* strains in walnut and other hosts, respectively. In light of the discovery of *X. euroxanthea*, its genetic closeness to *X. arboricola* pv. *juglandis* and likewise inhabitant of walnut trees (Chapter II - Martins et al., 2020; Chapter III - Fernandes et al., 2021b), it's possible that the reported morphological and genetic diversity of the *juglandis* pathovar (Scortichini et al., 2001; Marcelletti et al., 2010; Hajri et al., 2010; Giovanardi et al., 2016; Merda et al., 2016) could have been overestimated as some of the strains might not belong to the same pathovar but otherwise be a different species. Particularly, carried out an essential study to unravel the population structure of *X. arboricola* depicted three groups (A-C) (Merda et al., 2016). Group A include epidemic clones represented by pathovars with the most important pathogens (that is, pathogens that are responsible for worldwide spread diseases i.e. pv. *juglans* (WBB), *corylina* (bacterial blight on hazelnut) and *pruni* (bacterial spot on stone fruit tree)); Groups B and C included highly recombinant non-pathogenic or low virulence strains; Group C was the most diverse, showing four lineages, with one of them being constituted by strains CFBP 7622 and CPBF7653 (now acknowledged as *X. euroxanthea*). Kałużna et al. (2021) updated this segmentation, with all known strains of *X. euroxanthea* belonging to group C.

Until recently, in order to confirm the identity of these *X. arboricola* or *X. euroxanthea* strains, housekeeping genes would have to be sequenced and concatenated for multilocus sequence analysis or whole genomic sequences obtained and made publicly available. Now, with the establishment of the optimized multiplex PCR with *X. euroxanthea*-specific markers XEA1, XEA5 and XEA8 (XEA) developed in this study (Chapter V - Silva et al., 2022) it is possible to profile these dubious strains in a single PCR run.

X. euroxanthea and *X. arboricola* partner up on multiple host plants - a sympatric powerhouse duo?

Another interesting finding of this study was the realization that *X. arboricola* and *X. euroxanthea* strains coexistence in an exact same individual plant host is not exclusive to walnut (Chapter III - Fernandes et al., 2021a), but also occurs in pecan (Chapter IV - Martins et al., 2022). These two bacterial species have also been reported in the same

plant species, bean plants (Merda *et al.*, 2017) and tomato plants from the same region (Stanthorpe) (Roach *et al.*, 2018), but it is not detailed whether these isolates were from the same host specimen. Nonetheless, the frequent host sharing for these two bacteria in multiple hosts may not be a circumstantial but evidence of a sympatric lifestyle. In ecology, the concept of “sympatric species” describes species that occupy the same geographical range, consistently or only in particular regions (Marko, 2008). Despite this concept not being frequently applied on bacterial plant pathogens, (Moliner *et al.*, 2010) referred to the intracellular pathogens of an amoeba as a sympatric population. Given these definitions, it seems appropriate to consider these two bacterial species, *X. euroxanthea* and *X. arboricola*, recurrently found together in different hosts upon multiple geographic regions, as sympatric species. Is still to be unraveled if the co-colonization in a host contributes to infection as a synergistic interaction, or if low-virulence/non-pathogenic *X. euroxanthea* can “piggyback” their way into the host, by opportunistically making use of the mechanisms employed by pathogenic *X. arboricola* pv. *juglandis*, as also hypothesized by Merda *et al.*, (2017) regarding the simultaneous finding of non-pathogenic and pathogenic *X. arboricola* strains on the same host. Moreover, the presence of commensal strains could prime host defenses by inducing Systemic acquired resistance (SAR) which activates a network of durable, wide-ranging resistance (Gao *et al.*, 2015; An *et al.*, 2019).

T3SS and T3E evolution in *Xanthomonas* - more is not necessarily more

The T3SS and respective effectors are a major virulence asset in Gram-negative pathogens and particularly within the *Xanthomonas* genus. In which all pathogenic species encompass and rely on this machinery's function, with exception of two pathogenic species, *X. sacchari* and *X. albilineans* (An *et al.*, 2019; Meline *et al.*, 2019). Other *Xanthomonas* species than lack T3SS/T3E display a non-pathogenic phenotype, such as *X. cannabis* and *X. maliensis* or were considered as commensals, as reported for *X. arboricola* (Triplett *et al.*, 2015; An *et al.*, 2019).

A preliminary comparative investigation of T3SS and T3E present in *X. arboricola* pv. *juglandis* strain CPBF 427 and *X. euroxanthea* CPBF 367, CPBF 424^T and CPBF 426 pointed us to the disparity of T3E profiles between walnut pathogenic strains *Xaj* CPBF 427 and *X. euroxanthea* CPBF 424^T (Martins *et al.*, 2020) and was the starting point for a deeper analysis. Pathogenic *Xanthomonas* species and pathovars usually display around 25 to 35 effectors (Teper *et al.*, 2021), however it was previously noticed that in *X. arboricola* the effectome can be variable (Merda *et al.*, 2017). The reporting of a reduced

ancestral core repertoire of seven effectors for pathogen *X. arboricola* also supported the finding of a reduced effectome for pathogenic *X. euroxanthea* that was enough to cause disease, even if showing a weaker phenotype than the specialized walnut pathovar *Xaj* strains (Merda et al., 2017; Chapter II-Martins et al., 2020).

Tackling *X. euroxanthea* and *X. arboricola* sympatric lifestyle with comparative genomics

On the quest to understand the evolution of these two bacteria and their lifestyle it is crucial to ascertain their genomic characteristics and pathogenicity traits. While taxonomical genome similarity calculations do not demand very high-quality genomes, in order to obtain reliable comparative genomics assessments, the genomes should be closed or have low contig number, this way guaranteeing none or reduced number of gaps that could mean gene disruption and skewing of analysis (Smits, 2019). Sequencing of CPBF 367, CPBF 424^T, CPBF 426, CPBF 761, CPBF 766 (*X. euroxanthea*), CPBF 765, CPBF 1494 (*X. arboricola*), CPBF 427 and CPBF 1521 (*X. arboricola* pv. *juglandis*) strains through long-read technologies (Teixeira et al., 2020, 2021; Fernandes et al., 2021b) coupled to short-read sequencing genomes obtained previously (Fernandes et al., 2018a,b) or *de novo* within the group, (Martins et al., 2022) provided the opportunity to perform hybrid assemblies of these genomes. Thus achieving a closed genome assembly of the chromosomal sequence i.e. encompassed in a single contig (Teixeira et al., 2020, 2021; Fernandes et al., 2021b; Martins et al., 2022).

As an effort to tackle the sympatric lifestyle of *X. euroxanthea* and *X. arboricola* pv. *juglandis* and discern the genomic differences of these two species colonizing the same tree, we applied a comparative genomics approach focusing on general and virulence genomic profiles of CPBF 367, CPBF 424^T, CPBF 426 (*X. euroxanthea*) and CPBF 427 (*Xaj*) of differing pathogenicity phenotypes (Chapter III-Fernandes et al., 2021b). This global analysis considered several virulence factors, however, there were not major differences between the two species or within the *X. euroxanthea* species apart from a few genes coding for chemotaxis-related proteins, non-fimbrial adhesins, T2SS, extracellular enzymes and T4P. This study also revealed that non-pathogenic *X. euroxanthea* strains display the similarly reduced T3SS and T3E profile as commensal *X. arboricola* strains. While pathogenic *X. euroxanthea* show a reduced effectome comparing with members of the *juglandis* pathovar (Fernandes et al., 2021b).

Nonetheless, pathogenicity in bacteria is not only a matter of a major ensemble of virulence traits but also the lack of it, as host specialization follows major genome size

constrictions (Georgiades & Raoult, 2011). So, in our view, to fully understand if *X. euroxanthea* carries out a generalist lifestyle (i.e. generalist pathogens may have the ability to colonizes multiple hosts; vs. specialist pathogens that can only infect one host species (Baumler & Fang, 2013)) or is pathogenic solely on walnut trees and uses the other known isolation hosts as reservoirs. Genomic prediction will not suffice and this knowledge must be complemented with pathogenicity tests for hitherto all known *X. euroxanthea* strains to a wide range of plant hosts, including the variety of hosts of isolation known so far: walnut, pecan, tomato, bean and ground cherry.

Providing the tools for *X. euroxanthea* identification and typing

Xanthomonas euroxanthea includes both non-pathogenic and pathogenic strains, with the latter being reported to cause bacterial blight-like symptoms in walnut (Martins *et al.*, 2020). Considering the recurrent encounter of *X. euroxanthea* and *X. arboricola* on the same plant hosts and the similar symptomatology caused by both pathogens on walnut leaves (Martins *et al.*, 2020, 2022; Fernandes *et al.*, 2021b), it is of utmost importance to rapidly detect and discern these two bacteria in order to apply custom-made phytosanitary measures on concerned diseased plants. Towards this, *X. euroxanthea* specific DNA markers were developed following an *in silico* genomics investigation of gene determinants that could serve as ID (identification card) to unequivocally track this species in multiple hosts. Of a genomics comparison encompassing publicly available genomes of *X. euroxanthea* strains hitherto known, and genomes representative of other *Xanthomonas* species, eight markers were established (XEA1-XEA8), characterized, and tested for specificity through both *in silico* and wet lab approaches (Silva *et al.*, 2022). Furthermore, following the publication of this work, genomes of additional *X. euroxanthea* strains (more on this on the next topic) were divulged, allowing to extend the *in silico* XEA marker prediction for these stains (Supplementary Figure VI.S1). The broadening of the search made possible detecting an additional XEA marker pattern, totaling four patterns (A-D). Interestingly, it not clear the source of this division, as neither can be attributed to host, pathogenicity, year or location of isolation.

Evolution of *X. euroxanthea*

Xanthomonas genus can be divided in two clades according to nucleotide sequence comparison: Clade I, is pointed as early branching group, which include monocot isolated species; and Clade II, the group that includes the great portion of species and the most

economically important (Parkinson *et al.*, 2007; Jacques *et al.*, 2016). Core genome phylogeny has attributed *X. euroxanthea* to Clade II of *Xanthomonas* spp. (Mafakheri *et al.*, 2022). Unsurprisingly, phylogenetic analyses also indicate that *X. arboricola* is a sister species to *X. euroxanthea* (Martins *et al.*, 2020; Fernandes *et al.*, 2021b; Mafakheri *et al.*, 2022). Although this species does not appear to be particularly virulent and destructive so far, it is paramount to keep monitoring for its pathogenicity potential, as it was hypothesised before (Merda *et al.*, 2016) that these non-pathogenic, highly recombinant strains constitute a source pool to become pathogenic/and aggressive. In fact it has been observed an arise of more aggressive strains within *Xanthomonas* genus (Coutinho *et al.*, 2022). Moreover, the inclusion of both pathogenic and non pathogenic strains constitutes a great model to understand evolution of pathogenicity in *Xanthomonas*.

Current situation of *X. euroxanthea* – What we know so far

At the time this thesis was written, genomic sequences of 19 *X. euroxanthea* strains were available in the genome taxonomy database (<https://gtdb.ecogenomic.org/species?id=Xanthomonas%20euroxanthea>) and NCBI database (Table VI.1). As mentioned above, it is expected that nineteen is an underestimation of the number of strains already isolate. When this PhD project started, we were only aware of the existence of three *X. euroxanthea* strains: CPBF 367, CPBF 424^T and CPBF 426 that were included in the first publication that constitutes this PhD thesis (Chapter II - Martins *et al.*, 2020). Later on, with genomic analyses of these three strains of *X. euroxanthea* and *X. arboricola* pv. *juglandis* CPBF 427 strain isolated from a single walnut tree, two other *X. euroxanthea* strains, CFBP 7653, CFBP 7635, previously isolated in France, also from the same host, were clearly assigned to *X. euroxanthea* species, based on a phylo-taxogenomics approach, more specifically, ANI and dDDH calculations and core genome phylogeny (Chapter III - Fernandes *et al.*, 2021b). Moreover, during an extensive genomic analysis on searching physiological and pathogenicity traits that may be responsible for *X. arboricola* and *X. euroxanthea* strains colonization detected on multiple hosts, Portuguese strains isolated from pecan tree CPBF 761 and CPBF 766 and Australian strains BRIP 62409, BRIP 62411, BRIP 62415, BRIP 62418 isolated from tomato were also confirmed as *X. euroxanthea*, using the same approach as mentioned above (Chapter IV).

Table VI.I. Database of *X. euroxanthea* strains described so far

Strain	Host of Isolation	Sample	Geographic origin	Yr of isolation	Reference publication
CPBF 367	<i>Juglans regia</i>	buds	Loures, Portugal	2016	Martins et al. 2020
CPBF 424 ^T	<i>Juglans regia</i>	buds	Loures, Portugal	2016	Martins et al. 2020
CPBF 426	<i>Juglans regia</i>	buds	Loures, Portugal	2016	Martins et al. 2020
CPBF 761	<i>Carya illinoensis</i>	leaf	Alcobaça (Ganilhos), Portugal	2016	unpublished
CPBF 766	<i>Carya illinoensis</i>	leaf	Alcobaça (Ganilhos), Portugal	2016	unpublished
CFBP 7653	<i>Juglans regia</i>	buds	Périgord, France	2008	Essakhi et al., 2015
CFBP 7635a	<i>Juglans regia</i> cv. <i>Franquette</i>	healthy bud	Isère, France	2009	Essakhi et al., 2015
CFBP 7622	<i>Phaseolus vulgaris</i>	Leaf washings	USA	1985	Merda et al. 2017
BRIP 62415	<i>Solanum lycopersicum</i>	Leaf?	Stanthorpe, Australia	2015	Roach et al. 2019
BRIP 62409	<i>Solanum lycopersicum</i>	Leaf?	Stanthorpe, Australia	2015	Roach et al. 2019
BRIP 62418	<i>Solanum lycopersicum</i>	Leaf?	Stanthorpe, Australia	2015	Roach et al. 2019
BRIP 62411	<i>Solanum lycopersicum</i>	Leaf?	Stanthorpe, Australia	2015	Roach et al. 2019
2955	Environment (rain)	n/d	n/d	n/d	Potnis et al. unpublished
2957	Environment (rain)	n/d	n/d	n/d	Potnis et al. unpublished
3640	Environment (rain)	n/d	n/d	n/d	Potnis et al. unpublished
F2	<i>Physalis peruviana</i>	n/d	n/d	n/d	Potnis et al. unpublished
2974	Environment (rain)	n/d	n/d	n/d	Potnis et al. unpublished
3338a	Environment (rain)	n/d	n/d	n/d	Potnis et al. unpublished
2949	Environment (rain)	n/d	n/d	n/d	Potnis et al. unpublished

Meanwhile, CFBP 7622 isolated from bean plant in France, F2 isolated from physalis plant and 2949, 2955, 2957, 2974, 3640 isolated from rainwater were also included in *X. euroxanthea* group (Kałużna *et al.*, 2021; Zarei *et al.*, 2022; Wagner *et al.*, 2022). Additionally, strain 3338 also isolated from rainwater seems to belong to *X. euroxanthea* as well (data not shown). Curiously, CFBP 7622 was obtained from washes of bean plant (Wagner *et al.*, 2022), evidencing the epiphytic phase in the semi epiphytic/endophytic lifestyle currently accepted for *Xanthomonas* sp. (An *et al.*, 2019) or it can suggest that it maintains in the epiphytic form, using bean plant as a reservoir. In fact, (Zarei *et al.*, 2018) evaluated the risk of a multitude of non-host plants as possible reservoirs of *X. arboricola* and pointed out a considerable group of plants (bean plant included) that demonstrated high reservoir potential. The hypothesis of *X. euroxanthea* including epiphytic strains is not totally surprising, as most of the strains discovered so far appear to be non-pathogenic or lacking the T3SS (thus pointing to their non-pathogenic status). Additionally, the fact that some strains were recovered from rainwater may point to it as a means of dispersal for these bacteria to move from host to host like illustrated in the herein proposed *X. euroxanthea* lifestyle model (Figure VI.1).

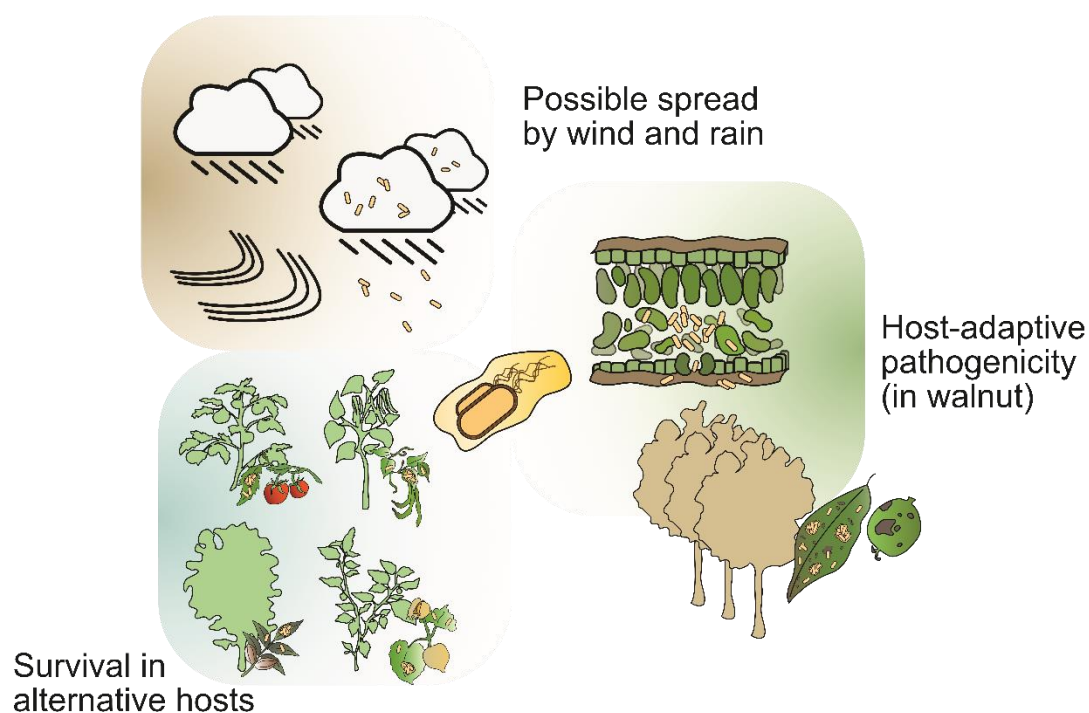


Figure VI.1. Working model depicting *X. euroxanthea* lifestyle

Final remarks and Future perspectives

This thesis attempted to disclose the intricate and curious association between *X. arboricola* and *X. euroxantha*. However, a deeper understanding of the interaction and behavior of *X. arboricola* – *X. euroxantha* upon co-colonization, namely on plant infection should be addressed by co-infection assays on host plants, and also by following the traffic of fluorescently-marked representative of *X. arboricola* and *X. euroxantha* within the plant.

The XEA markers herein proposed were successful in specifically amplifying *X. euroxantha* DNA, obtaining no off-target amplification of close-related and/or niche sharing species. Although it constitutes an invaluable tool for rapid identification of *X. euroxantha* in the lab, at this moment is still requires bacteria culture and isolation steps. In order to upgrade this method for a culture-independent and field-use tool further optimization is needed to achieve direct detection from plant samples extracts. It could also be adapted for qPCR scheme, as this technique is routinely used for phytopathogen's surveillance and detection programs for by reference labs and governmental entities.

References

- An SQ, Potnis N, Dow M *et al.*, 2019. Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen *Xanthomonas*. *FEMS Microbiology Reviews* **44**, 1–32.
- Baumler A, Fang FC, 2013. Host Specificity of Bacterial Pathogens. *Cold Spring Harbor Perspectives in Medicine* **3**, a010041.
- Coutinho TA, Jacques M-A, Jones J, 2022. Editorial: Emergence and re-emergence of plant diseases caused by *Xanthomonas* species. *Frontiers in Microbiology* **13**, 4851.
- Fernandes C, Albuquerque P, Mariz-Ponte N, Cruz L, Tavares F, 2021a. Comprehensive diversity assessment of walnut-associated xanthomonads reveal the occurrence of distinct *Xanthomonas arboricola* lineages and of a new species (*Xanthomonas euroxanthea*) within the same tree. *Plant Pathology* **70**, 5.
- Fernandes C, Blom J, Pothier JF, Tavares F, 2018a. High-Quality Draft Genome Sequence of *Xanthomonas* sp. Strain CPBF 424, a Walnut-Pathogenic Strain with Atypical Features. *Microbiology Resource Announcements* **7**.
- Fernandes C, Blom J, Pothier JF, Tavares F, 2018b. High-Quality Draft Genome Sequence of *Xanthomonas arboricola* pv. *juglandis* CPBF 1521, Isolated from Leaves of a Symptomatic Walnut Tree in Portugal without a Past of Phytosanitary Treatment. *Microbiology Resource Announcements* **7**.
- Fernandes C, Martins L, Teixeira M *et al.*, 2021b. Comparative Genomics of *Xanthomonas euroxanthea* and *Xanthomonas arboricola* pv. *juglandis* Strains Isolated from a Single Walnut Host Tree. *Microorganisms* 2021, Vol. 9, Page 624 **9**, 624.
- Gao QM, Zhu S, Kachroo P, Kachroo A, 2015. Signal regulators of systemic acquired resistance. *Frontiers in Plant Science* **6**.
- Georgiades K, Raoult D, 2011. Defining pathogenic bacterial species in the genomic era. *Frontiers in Microbiology* **1**, 151.
- Giovanardi D, Bonneau S, Gironde S, Saux MF le, Manceau C, Stefani E, 2016. Morphological and genotypic features of *Xanthomonas arboricola* pv. *juglandis* populations from walnut groves in Romagna region, Italy. *European Journal of Plant Pathology* **145**.

- Hajri A, Meyer D, Delort F, Guillaumès J, Brin C, Manceau C, 2010. Identification of a genetic lineage within *Xanthomonas arboricola* pv. *juglandis* as the causal agent of vertical oozing canker of Persian (English) walnut in France. *Plant Pathology* **59**, 1014–1022.
- Jacques M-A, Arlat M, Boulanger A *et al.*, 2016. Using Ecology, Physiology, and Genomics to Understand Host Specificity in *Xanthomonas*. *Annual Review of Phytopathology* **54**, 163–187.
- Kałużna M, Fischer-Le Saux M, Pothier JF *et al.*, 2021. *Xanthomonas arboricola* pv. *juglandis* and pv. *corylina*: Brothers or distant relatives? Genetic clues, epidemiology, and insights for disease management. *Molecular Plant Pathology* **22**, 1481–1499.
- Konstantinidis KT, Tiedje JM, 2005. Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 2567–2572.
- Mafakheri H, Taghavi SM, Zarei S *et al.*, 2022. *Xanthomonas bonasiae* sp. nov. and *Xanthomonas youngii* sp. nov., isolated from crown gall tissues. *International Journal of Systematic and Evolutionary Microbiology* **72**, 005418.
- Marcelletti S, Ferrante P, Scortichini M, 2010. Multilocus Sequence Typing Reveals Relevant Genetic Variation and Different Evolutionary Dynamics among Strains of *Xanthomonas arboricola* pv. *juglandis*. *Diversity* **2**, 1205–1222.
- Marko PB, 2008. Sympatry. *Encyclopedia of Ecology, Five-Volume Set*, 3450–3458.
- Martins L, Fernandes C, Blom J, Dia NC, Pothier JF, Tavares F, 2020. *Xanthomonas euroxanthea* sp. nov., a new xanthomonad species including pathogenic and non-pathogenic strains of walnut. *International Journal of Systematic and Evolutionary Microbiology* **70**, 9.
- Martins L, Teixeira M, Fernandes C *et al.*, 2022. Genomic features of *Xanthomonas arboricola* and *X. euroxanthea* sharing the same plant host species preferences (walnut, pecan and tomato) seems unbiased by the host species.
- Meline V, Delage W, Brin C *et al.*, 2019. Role of the acquisition of a type 3 secretion system in the emergence of novel pathogenic strains of *Xanthomonas*. *Molecular Plant Pathology* **20**, 33–50.

- Merda D, Bonneau S, Guimbaud JF *et al.*, 2016. Recombination-prone bacterial strains form a reservoir from which epidemic clones emerge in agroecosystems. *Environmental Microbiology Reports* **8**, 572–581.
- Merda D, Briand M, Bosis E *et al.*, 2017. Ancestral acquisitions, gene flow and multiple evolutionary trajectories of the type three secretion system and effectors in *Xanthomonas* plant pathogens. *Molecular Ecology* **26**, 5939–5952.
- Moliner C, Fournier PE, Raoult D, 2010. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiology Reviews* **34**, 281–294.
- Parkinson N, Aritua V, Heeney J, Cowie C, Bew J, Stead D, 2007. Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *International Journal of Systematic and Evolutionary Microbiology* **57**, 2881–2887.
- Roach R, Mann R, Gambley CG, Shivas RG, Rodoni B, 2018. Identification of *Xanthomonas* species associated with bacterial leaf spot of tomato, capsicum and chilli crops in eastern Australia. *European Journal of Plant Pathology* **150**, 595–608.
- Scortichini M, Marchesi U, di Prospero P, 2001. Genetic diversity of *Xanthomonas arboricola* pv. *juglandis* (synonyms: *X. campestris* pv. *juglandis*; *X. juglandis* pv. *juglandis*) Strains from different geographical areas shown by repetitive polymerase chain reaction genomic fingerprinting. *Journal of Phytopathology* **149**, 325–332.
- Silva KG, Martins L, Teixeira M, Pothier JF, Tavares F, 2022. DNA Markers for Detection and Genotyping of *Xanthomonas euroxanthea*. *Microorganisms* 2022, Vol. 10, Page 1078 **10**, 1078.
- Smits THM, 2019. The importance of genome sequence quality to microbial comparative genomics. *BMC Genomics* **20**, 1–4.
- Teixeira M, Fernandes C, Chaves C, Pinto J, Tavares F, Fonseca NA, 2021. Complete Genome Sequence Obtained by Nanopore and Illumina Hybrid Assembly of *Xanthomonas arboricola* pv. *juglandis* CPBF 427, Isolated from Buds of a Walnut Tree (SR Gill, Ed.). *Microbiology Resource Announcements* **10**.
- Teixeira M, Martins L, Fernandes C *et al.*, 2020. Complete Genome Sequences of Walnut-Associated *Xanthomonas euroxanthea* Strains CPBF 367 and CPBF 426 Obtained by Illumina/Nanopore Hybrid Assembly (DA Baltrus, Ed.). *Microbiology Resource Announcements* **9**.

- Teper D, Pandey SS, Wang N, 2021. The HrpG/HrpX Regulon of Xanthomonads—An Insight to the Complexity of Regulation of Virulence Traits in Phytopathogenic Bacteria. *Microorganisms* **9**, 187.
- Triplett LR, Verdier V, Campillo T *et al.*, 2015. Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali and proposal of *Xanthomonas maliensis* sp. nov. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* **107**, 869–881.
- Wagner N, Albuquerque M, Ecker N *et al.*, 2022. Natural language processing approach to model the secretion signal of type III effectors. *Frontiers in Plant Science* **13**, 4339.
- Zarei S, Taghavi SM, Hamzehzarghani H, Osdaghi E, Lamichhane JR, 2018. Epiphytic growth of *Xanthomonas arboricola* and *Xanthomonas citri* on non-host plants. *Plant Pathology* **67**, 660–670.
- Zarei S, Taghavi SM, Rahimi T *et al.*, 2022. Taxonomic Refinement of *Xanthomonas arboricola*. <https://doi.org/10.1094/PHYTO-12-21-0519-R>.

Supplementary material

<i>Xanthomonas euroxanthea</i>									
Hosts	Strains		DNA Markers						
			XEA1	XEA2	XEA3	XEA4	XEA5	XEA6	XEA7
			XEA8						
<i>Juglans regia</i>		CPBF 367	●	●	●	●	●	●	●
<i>Juglans regia</i>		CPBF 424 ^T	●	●	●	●	●	●	●
<i>Juglans regia</i>		CPBF 426	●	●	●	●	●	●	●
<i>Juglans regia</i>		CPBF 7653	●	●	●	●	●	●	●
<i>Carya illinoensis</i>		CPBF 761	●	●	●	●	●	●	●
Rain		2974	●	●	●	●	●	●	●
Rain		3640	●	●	●	●	●	●	●
<i>Physalis peruviana</i>		F2	●	●	●	●	●	●	●
Rain		2955	○	●	●	●	●	●	●
Rain		2957	○	●	●	●	●	●	●
Rain		3338	○	●	●	●	●	●	●
<i>Carya illinoensis</i>		CPBF 766	●	○	○	●	●	●	●
<i>Solanum lycopersicum</i>		BRIP 62415	●	○	○	●	●	●	●
<i>Solanum lycopersicum</i>		BRIP 62409	●	○	○	●	●	●	●
<i>Solanum lycopersicum</i>		BRIP 62411	○	○	○	●	●	●	●
<i>Solanum lycopersicum</i>		BRIP 62418	○	○	○	●	●	●	●
<i>Phaseolus vulgaris</i>		CFBP 7622	○	○	○	●	●	●	●
Rain		2949	○	○	○	●	●	●	●

Supplementary Figure VI.S1. Distribution of six *Xanthomonas euroxanthea* (Xea)-specific DNA markers (XEA1 - XEA8) in hitherto known and available *X. euroxanthea* genomes assessed by BLASTn analysis, disclosing patterns A-D