

## Article

# Orthology-Based Estimate of the Contribution of Horizontal Gene Transfer from Distantly Related Bacteria to the Intraspecific Diversity and Differentiation of *Xylella fastidiosa*

Giuseppe Firrao<sup>1,2,\*</sup>, Marco Scortichini<sup>3</sup> and Laura Pagliari<sup>1</sup><sup>1</sup> Dipartimento di Scienze Agroalimentari, Ambientali e Animali, University of Udine, Udine, Italy<sup>2</sup> Istituto Nazionale Biostrutture e Biosistemi, Rome, Italy<sup>3</sup> Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria-Centro di ricerca per l'Olivicoltura, Frutticoltura e Agrumicoltura, Via di Fioranello, 52, 00134 Rome, Italy; marco.scortichini@crea.gov.it

\* Correspondence: firrao@uniud.it

**Abstract:** *Xylella fastidiosa* is a xylem-limited bacterium phylogenetically related to the xanthomonads, with an unusually large and diversified range of plant hosts. To ascertain the origin of its peculiarities, its pan-genome was scanned to identify the genes that are not coherent with its phylogenetic position within the order *Xanthomonadales*. The results of the analysis revealed that a large fraction of the genes of the *Xylella* pan-genome have no ortholog or close paralog in the order *Xanthomonadales*. For a significant part of the genes, the closest homologue was found in bacteria belonging to distantly related taxonomic groups, most frequently in the *Betaproteobacteria*. Other species, such as *Xanthomonas vasicola* and *Xanthomonas albilineans* which were investigated for comparison, did not show a similar genetic contribution from distant branches of the prokaryotic tree of life. This finding indicates that the process of acquisition of DNA from the environment is still a relevant component of *Xylella fastidiosa* evolution. Although the ability of *Xylella fastidiosa* strains to recombine among themselves is well known, the results of the pan-genome analyses stressed the additional relevance of environmental DNA in shaping their genomes, with potential consequences on their phytopathological features.

**Keywords:** hierarchical orthologous group; pangenome; HGT; *Xanthomonadales*; *Xanthomonas vasicola*; *albilineans*



**Citation:** Firrao, G.; Scortichini, M.; Pagliari, L. Orthology-Based Estimate of the Contribution of Horizontal Gene Transfer from Distantly Related Bacteria to the Intraspecific Diversity and Differentiation of *Xylella fastidiosa*. *Pathogens* **2021**, *10*, 46. <https://doi.org/10.3390/pathogens10010046>

Received: 24 November 2020

Accepted: 5 January 2021

Published: 7 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

*Xylella fastidiosa* is a Gram-stain-negative, xylem-limited, vector-borne, nutritionally fastidious phytopathogen that can colonize many wild and cultivated plant species. The host range of *Xylella fastidiosa* is unusually large among plant pathogenic bacteria, and comprises 563 plant species grouped into 82 botanical families, including many of socio-economic interest (i.e., grapevine, citrus, coffee, and olive trees) [1]. Nevertheless, only a limited number of infected plant species shows the typical wilting of leaves, twigs and branches and the subsequent plant death, caused by the pathogen-induced blockage of xylem vessels [2]. The currently available data suggest that *Xylella fastidiosa* is associated with a large number of plant species as a commensalist, but only particular clades -or even genotypes- are responsible for a small number of plant diseases [3–5].

According to the taxonomy based on ribosomal RNA sequence analysis, *Xylella fastidiosa* is a member of the *Xanthomonadaceae* family [6]. Genome analyses have confirmed that the status of *Xylella fastidiosa* as a member of the family is phylogenetically well supported [7], even if its physiological, cultural, morphological and phytopathological features greatly differ from *Xanthomonas* spp. Three subspecies are recognized within the *Xylella fastidiosa* species, namely *fastidiosa*, *multiplex* and *pauca*, while a closely related species is *Xylella taiwanensis* which causes pear leaf scorch in Taiwan [8–11].

The study of the phylogenetic relationships among *Xylella fastidiosa* isolates and the clade organization in relation to both host plant and geographical location has been promoted by MultiLocus Sequence Type analysis (MLST) [12]. The same method also revealed recombination between *Xylella fastidiosa* strains [13] and genetic introgression of different subspecies within the native *Xylella fastidiosa* strains of Central America; substantial recombination was observed between some *Xylella fastidiosa* subsp. *multiplex* and subsp. *fastidiosa* strains of the United States, likely associated with host shifts of the strains [14]. Relevant gene flow between *Xylella fastidiosa* subspecies has also been detected for strains of subsp. *pauca* and *multiplex* in Brazil [15]. Despite the occurrence of recombination events between strains, *Xylella fastidiosa* shows a low ratio of recombination to point mutation on a per housekeeping gene basis when compared to other bacterial species [12]. Genome-wide approaches also revealed recombination events among strains, so that ancestral and more recent recombination was detected for all subspecies, with some strains showing a high recombination rate [16,17].

Most studies published so far have focused on intraspecific recombination events, therefore little is known about the proportion of genes possibly acquired by *Xylella fastidiosa* from related or taxonomically distant species. Denancé et al. [11], by assessing the *Xylella fastidiosa* strain k-mers of the genomes, confirmed the variability of the species and found (in one strain isolated from *Hibiscus fragilis*) unique regions both in the chromosome and in the plasmid with high homology to genomic regions of other bacterial genera, such as *Comamonas*, *Cupriavidus*, *Pseudomonas*, *Klebsiella*, *Bordetella*, *Burkholderia* as well as *Xanthomonas euvesicatoria*.

To investigate whether parts of the *Xylella fastidiosa* genome could have originated by horizontal transfers, we applied a genomic approach that allows the comparison of either related or distantly related bacterial strains. To accomplish the task, we built up the pan-genome of *Xylella fastidiosa* from a large sample of genome drafts, and scanned it asserting orthology for each gene. As comparison references, we applied the same approach to other two species of the *Xanthomonadaceae*, i.e., *Xanthomonas vasicola* and *Xanthomonas albilineans*, which are a typical and an atypical member species of the genus *Xanthomonas*, respectively [18]. Similarly to *Xylella fastidiosa*, *Xanthomonas vasicola* colonizes many different host plants such as corn, sugarcane, sorghum, banana, ensete, areca nut and *Tripsacum laxum* [19], thus showing strong adaptive potential. On the other hand, *Xylella fastidiosa* shares with *Xanthomonas albilineans* some relevant features: both actively colonize the plant xylem, they do not have *hrp* genes encoding for the type III secretion system, and both experienced a reductive genome evolution during their descent from the common ancestor, presumably due to adaptation to the nutritionally poor xylem niche [20]. We show here that, from a genomic point of view, *Xylella fastidiosa*, differently from the two other *Xanthomonadales*, has a significant portion of its genome that was acquired by horizontal genetic transfer (HGT) from distantly related bacteria.

## 2. Results

### 2.1. Orthology Assertion

Each putative protein-encoding gene of the pan-genome of *Xylella fastidiosa* and, for comparison, of the pan-genomes of *Xanthomonas vasicola* and *Xanthomonas albilineans*, was investigated to ascertain whether or not it shares a common ancestry with genes of other members of the order *Xanthomonadales*, a taxon that includes all three named species.

To gain an insight into the ancestry of the genes, we exploited the OMA (Orthologous Matrix), which distinguishes itself through high-quality orthology inferences, the broad coverage of bacteria, and the availability of a programmatic access [21,22]. We systematically searched the pan-genomes for genes missing orthologs of close paralogs in the class *Xanthomonadales*.

The procedure outlined in Figure S1 integrates the results of sequence similarity searches, orthology inferences, and global alignments, to select, from each pan-genome, the genes whose origin cannot be traced back to the *Xanthomonadales*. Genes that were most

similar to genes carried by phylogenetically distant bacteria were not included in the selection, when a putative hierarchical ortholog at the *Xanthomonadales* level was detected. This conservative approach implies discharging a priori hypotheses of independent acquisition or loss and gain events, but provided a non-subjective, well-defined frame for comparison.

The results of the assignments are hereafter presented in detail, first restricted to phylogenetically conserved subsets, and then extended to the entire pan-genome.

## 2.2. Analysis of the Core Genome

The first set investigated was restricted to the so-called BUSCO (Benchmarking Universal Single-Copy Orthologs) genes [23]. BUSCO sets for the major bacterial lineages, available through OrthoDB ([www.orthodb.org](http://www.orthodb.org)), are selected from orthologous groups with genes present as single-copy orthologs in at least 90% of the species, and are hence regarded as having strict evolutionary coherence. Within the BUSCO dataset compiled for the *Gammaproteobacteria* there are 452 genes. In the genomes of *Xanthomonas vasicola* and *Xanthomonas albilineans*, seven and eight of these genes, respectively, were missing from all genomes. Considering the genomes of *Xylella fastidiosa*, as many as 15 BUSCO genes were missing, including RecX, an inhibitor of RecA (Table 1). All BUSCO genes of all three species belong to a hierarchical orthologous group (HOG) at the level of *Xanthomonadales*.

**Table 1.** Benchmarking Universal Single-Copy Orthologs (BUSCO) genes missing in the genomes of *Xanthomonas vasicola*, *Xanthomonas albilineans*, and *Xylella fastidiosa*.

<i>Xylella fastidiosa</i>	<i>Xanthomonas vasicola</i>	<i>Xanthomonas albilineans</i>	Description
POG0909014H	POG0909014H	POG0909014H	UDP- <i>N</i> -acetylmuramoylalanine-D-glutamate ligase
POG090901AA	POG090901AA	POG090901AA	Small ribosomal subunit biogenesis GTPase RsgA
POG090901YD	POG090901YD	POG090901YD	membrane protein
POG09090233	POG09090233	POG09090233	apolipoprotein <i>N</i> -acyltransferase
POG090902S9	POG090902S9	POG090902S9	threonylcarbamoyl-AMP synthase
POG090903DC	POG090903DC	POG090903DC	argininosuccinate lyase
POG0909029O	POG0909029O		Thymidylate kinase
		POG090901QT	peptide chain release factor 1
		POG090900DY	tRNA N6-adenosine threonylcarbamoyltransferase
POG090900A4			tRNA modification GTPase MnmE
POG090900LM			Bifunctional uridylyltransferase/uridylyl-removing enzyme
POG0909004H			Bifunctional glutamine synthetase adenylyltransferase/adenylyl-removing enzyme
POG090901JJ			D-aminoacyl-tRNA deacylase
POG090901Q3			RNA-binding protein
POG09090284			tRNA-dihydrouridine synthase B
POG090902SD			membrane protein
POG090903LW			Regulatory protein RecX

A second set, named Core, is a subset of the pan-genome that includes the genes shared by all genomes of each species examined. The number of genes included reflects the average size and diversity of the genomes used for the construction of the pan-genome, and also the completeness and quality of the drafts used [24].

For *Xanthomonas vasicola*, which has an average genome size of about 4.9 Mbp containing 3888 genes (max: 4068, min: 3174), the core genome constructed using 65 genome

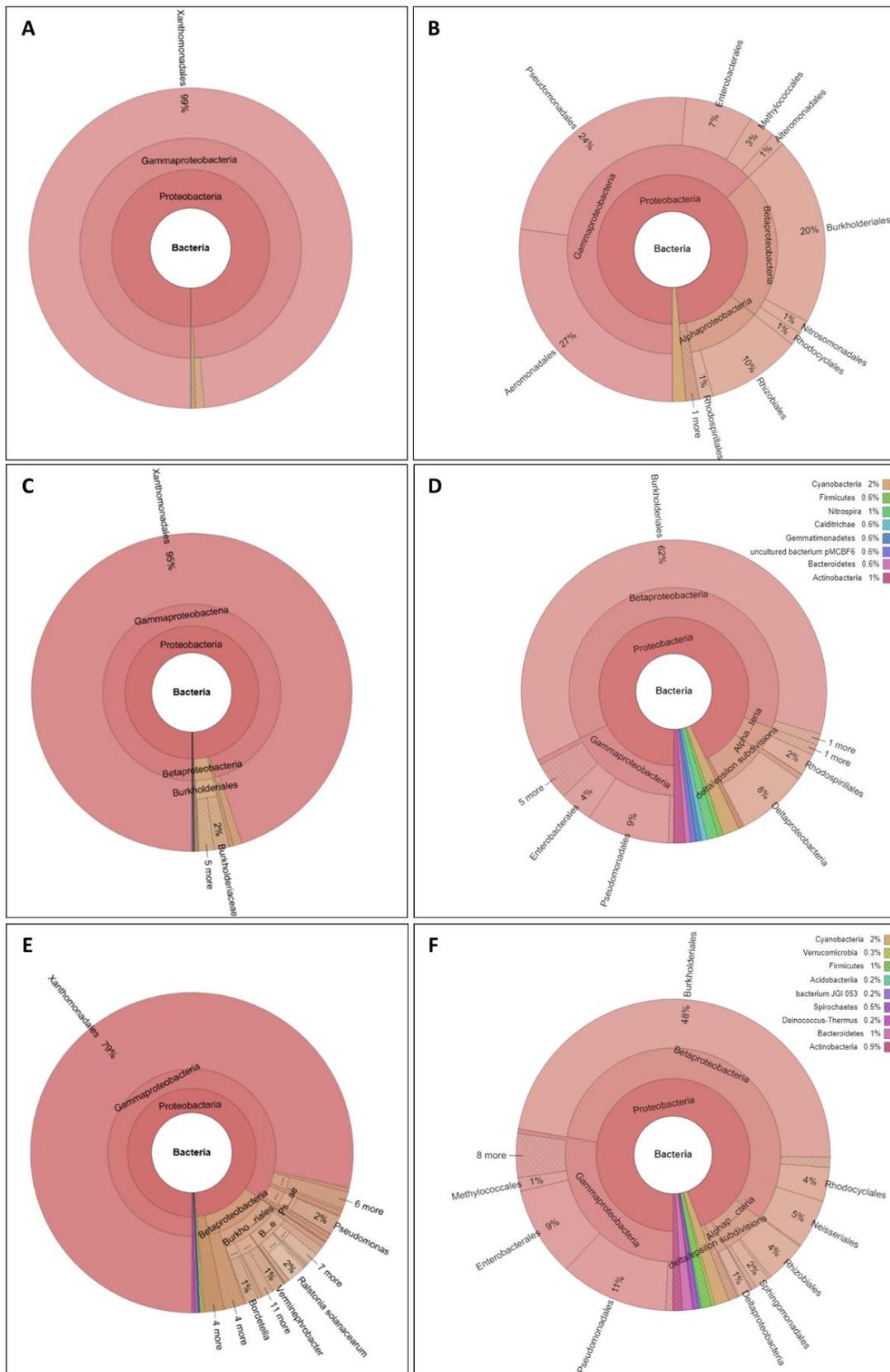
drafts comprises 1828 orthologous groups (OGs) of protein encoding genes, (47% of the average genome), all having HOGs at the *Xanthomonadales* level and all having their best BLAST hit within the *Xanthomonadales* (Table 2, Figure S2).

**Table 2.** Comparison of the relevant data in the pan-genome analysis of *Xanthomonas vasicola*, *Xanthomonas albilineans* and *Xylella fastidiosa*.

<i>Xanthomonas vasicola</i>	<i>Xanthomonas albilineans</i>	<i>Xylella fastidiosa</i>	
4.9	3.7	2.5	Average genome size (Mbp)
3888	2919	2176	Average number of predicted proteins
<b>Core genome:</b>			
1828	2386	1045	Genes searched for orthology and homology
1828 (100%)	2371 (99.4%)	1038 (99.3%)	Genes that have orthologs or close paralogs in the <i>Xanthomonadales</i>
0 (0%)	15 (0.6%)	7 (0.7%)	Genes that have orthologs in taxa not belonging to <i>Xanthomonadales</i>
<b>Pan-genome:</b>			
5296	3536	3673	Protein encoding genes in pan-genome
5155	3369	3049	Genes searched for orthology and homology
5085 (98.6%)	3198 (94.9%)	2395 (78.6%)	Genes that have orthologs or close paralogs in the <i>Xanthomonadales</i>
70 (1.4%)	171 (5.1%)	654 (21.4%)	Genes that have orthologs in taxa not belonging to <i>Xanthomonadales</i> (Figure 1)

The genome of *Xanthomonas albilineans* has an average size of 3.7 Mbp and contains 2919 genes (max: 3083, min: 2764); its core genome constructed from 16 genome drafts comprises 2386 OGs of protein-encoding genes (81.7% of the average genome). Among them, 2371 (99.4%) have HOGs at the *Xanthomonadales* level; the remaining 15 genes (corresponding to 0.6% of the total genes of the core genome) included several degrading enzymes, enzymes connected with chorysmate metabolism, a non-ribosomal peptidase synthetase, and a methyl transferase (Table 2). According to our analysis, most of those genes have homologs in the genomes of organisms in the beta- and delta-subdivision of the Proteobacteria (Figure S3). Moreover, for 14 of the 2371 genes, the detected HOG is not the best BLAST hit in the database and may therefore be the results of gene loss and gain from taxonomically unrelated bacteria.

In *Xylella fastidiosa*, the average genome is 2.54 Mbp in size and contains, on average, 2176 genes (min: 1930, max: 2413), and the core genome comprises 1045 genes (55.6% of the average genome) (Table 2). Among them, there were seven genes (corresponding to 0.7% of the total genes of the core genome) for which no ortholog or close paralog could be found in the *Xanthomonadales*, including a membrane protein and a quinolinate synthase NadA. In addition, there were 15 detected HOGs for which the best BLAST hit in the database was not in the *Xanthomonadales* and may therefore be the results of gene loss and gain from taxonomically unrelated bacteria (Table 2, Figure S4).



**Figure 1.** Taxonomic subdivision of the nearest ortholog assignment of all genes in the pan-genome (A,C,E) and of the subset of genes that do not have an ortholog in the *Xanthomonadales* (B,D,F) as detected in the pan-genomes of *Xanthomonas vasicola* (A,B), *Xanthomonas albilineans* (C,D) and *Xylella fastidiosa* (E,F). For better reading, the same graphics are available in interactive form as Supplementary Figures S2–S4.

### 2.3. Analysis of the Pan-Genome

Although the analysis of the core genome may provide clues of ancient HGTs that have been fixed within the genome, thus contributing to the definition of the ecological features of the species, the pan-genome analysis includes genes that were recorded in some (at least two) but not all genomes. Thus, it defines features that may affect the host range or the differences among subspecies or strains [24].

The pan-genome of *Xanthomonas vasicola* comprises 5296 genes, 141 (2.6%) of which had no significant homology with other proteins in the database and no associated putative function. Among the remaining 5155 genes, there were 70 genes (1.4%) for which no HOG at the *Xanthomonadales* level was found (Figure 1); with the sole exception of a membrane protein, they encode proteins presumably associated with phages or plasmids (Figure S2).

The pan-genome of *Xanthomonas albilineans* comprises 3536 genes, 167 of which (4.7%) had no significant homology with other proteins in the database and no associated putative function. Among the remaining 3369 genes, 95% have HOGs *Xanthomonadales*, while the remaining 171 OGs included coding sequences that could be traced primarily to the *Betaproteobacteria*, and also to other classes (Figure 1 and Figure S3). Most of these proteins are annotated as hypothetical proteins, phage and plasmid related proteins, or DNA processing enzymes, reducing the number of genes with putative associated function and possible impact on adaptation to those already found as part of the core genome set and eight additional proteins, which included two components of membrane transport system, one adhesin, two enzymes with putative biosynthetic activities, and another non ribosomal peptide synthetase. Interestingly, the *Xanthomonas albilineans* genome includes genes encoding non-ribosomal peptide synthetases that are absent in other *Xanthomonas* spp. but present in a wide range of bacteria.

The pan-genome of *Xylella fastidiosa* comprises 3673 protein genes, among which as many as 616 (nearly one-fifth of the putative coding sequences of the pan-genome) had no significant homology with other proteins in the database, and no associated predicted function. Among the others (3049, i.e., 83% of the pan-genome), 75% have homologs in the order *Xanthomonadales*, while as many as 654 (17.8% of the pan-genome) could not be tracked to the class (Figure 1; Figure S4). Among them, 473 (12.8%) encode functions that could not be tracked to the *Gammaproteobacteria*. Although a significant fraction of the genes with no HOG in common to *Xanthomonadales* encodes phage/transposon-related proteins or putative proteins with no function determined, we identified a set of 313 genes with predicted function.

### 2.4. Genomic Profiling of Pathway Activity for *Xylella fastidiosa*

To obtain clues about the function of this gene set, we enriched the annotation with ko-terms using the KEGG facilities (KAAS) and classified the CDSs according to their function and their taxonomically closest relative. A summary of the KEGG annotation is shown in Figure 2, where the CDSs that were not assigned by the KAAS annotator were manually assigned to additional categories according to the results of BLAST searches.



According to the analysis, the largest contribution to the diversification of *Xylella fastidiosa* strains was given by the *Betaproteobacteria*, particularly from the orders *Burkholderiales* (47.3% of the gene set), *Neisseriales* (4.8%) and *Rhodocyclales* (3.9%) (Figure 2). Additionally, members of *Gammaproteobacteria* and the *Alphaproteobacteria* played a role in shaping the *Xylella fastidiosa* pan-genome (28 and 6.8%, respectively), in particular the orders *Enterobacteriales* (10.1%), *Pseudomonadales* (11.0%) and *Rhizobiales* (4.5%) (Figure 2). The predominant input of *Burkholderiales* genes is recognizable in the five most populated functional categories, i.e., “Prokaryotic defense system”, “Replication and repair”, “Enzymes”, “Secretion system” and “Glycosaminoglycan binding proteins” (Figure 2). “Prokaryotic defense system” includes proteins of the type II toxin–antitoxin system (roughly 40% of the total proteins annotated within this category), among which 15 out of 25 (i.e., 60%) are from the *Burkholderiales*. Additionally, ATP-dependent helicases, a major group within the “Replication and repair” category (27%), constitutes proteins probably derived from HGT with members of *Burkholderiales* (Table S2). “Enzymes” collects the long list of various genes encoding biosynthetic or degrading enzymes that for their limited functional characterization were reported in this general group. They include hydroxylases, acyltransferases, deaminases, aminomutase, decarboxylase, superoxide dismutase, lipase, tautomerase, quinolinate synthase, anthranilate synthase, alkene reductase, ketoreductases, and peptidase of various families, as summarized in Table S3.

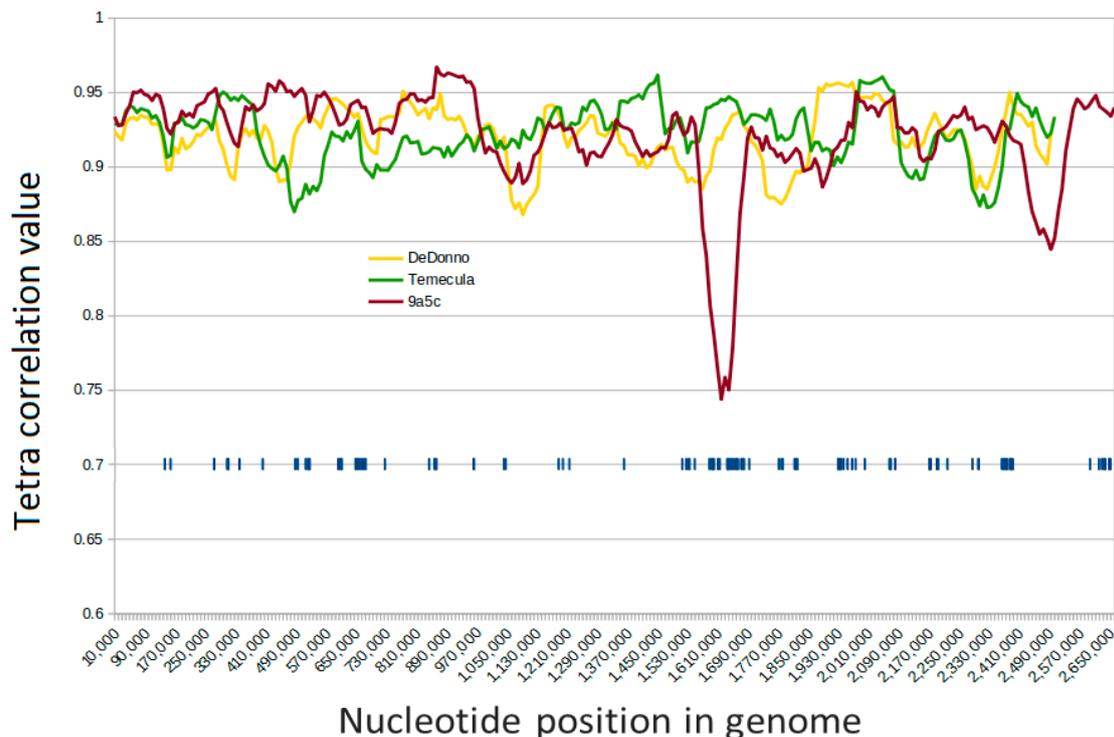
Roughly one-fifth of them had their closest orthologous groups within the *Burkholderiales* order (Table S2). Additionally, the genes for hemagglutinins, which cover more than the 50% of the “Glycosaminoglycan binding proteins” category, are similar to the homologous genes of members of the *Burkholderiales* (Table S2). Interestingly, all of the approximately 20 genes in the “Secretion systems” had their closest orthologs within the *Burkholderiales* and are associated with conjugal transfer (Table S2). Additionally, “Prokaryotic defense systems” and “Replication and repair” include many genes associated with plasmids and transposons (Table S2). Furthermore, an analysis of the sequences deposited as plasmids in the databases showed that 431 out of the 654 (66%) OGs that we identified as missing an ortholog in the *Xanthomonadales* included at least one member that has been found on a plasmid (Table S2). Beside *Burkholderiales*, other bacterial orders contributed to *Xylella fastidiosa* diversity. For example, several genes encoding proteins of the toxin–antitoxin (TA) system, which are listed under the descriptor heading “Prokaryotic defense systems”, had their closest relative distributed over a wide taxonomic range (i.e., *Xanthomonadales*, *Rhizobiales*, *Burkholderiales*, *Enterobacteriales*, *Pseudomonadales*, *Nitrosomonadales*, *Methylococcales*, and *Desulfobacteriales*).

### 2.5. Evidence of Recent HGT in *Xylella fastidiosa* Genomes

The 6697 nts region of the genome of *Xylella fastidiosa* subsp. *fastidiosa* 9a5c, from positions 1,644,305 to 1,651,001 (NCBI reference sequence: NC\_002488.3), is almost identical (90% identity at the DNA level) to the 6758 nts region, from position 6,187,137 to 6,192,834, of the distantly related *Pseudomonas* sp. CCA1 (NCBI reference sequence: NZ\_BDGS01000001.1). The region includes three genes encoding aldo/keto reductases, and genes encoding a 4-oxalocrotonate tautomerase, a glucose 1-dehydrogenase, an NADP-alcohol dehydrogenase, and an MFS transporter, in the same strictly conserved order. The region is common to the *Xylella fastidiosa* subsp. *pauca* strains isolated from plum, citrus and coffee in South America, and not to *Xylella fastidiosa* subsp. *fastidiosa* from North America nor to other subspecies.

In addition to the high similarity at the DNA level and the conservation of the gene order, further evidence that this region represents a recent introduction into the *Xylella fastidiosa* chromosome was gained by comparing the tetranucleotide patterns in the region and in the entire chromosome. The analysis was performed on 15 genomes (Figure S5 and Figure 3); Figure 3 shows a sliding window graph of the tetra correlation values between the 100,000 nts windows and the chromosome of three representative strains. Drops of the correlation values occurred between positions 1,610,000 and 1,670,000 and between

positions 2,480,000 and 2,520,000 in the chromosome of *Xylella fastidiosa* subsp. *pauca* 9a5c. Divergence in the tetranucleotide patterns is regarded as a hallmark of recent horizontal gene transfer events [25]: DNA sequences that are introduced through recent horizontal transfer often bear distinct sequence characteristics and can be distinguished from ancestral DNA but, over time, these sequences ameliorate to reflect the DNA composition of the new genome [26].



**Figure 3.** Sliding window graph of the tetra correlation values between a 100 k window and the entire chromosome. The DNA sequence of the chromosomes of three *X. fastidiosa* strains (*Xylella fastidiosa* subsp. *fastidiosa* strain Temecula, green line; *Xylella fastidiosa* subsp. *pauca* strain DeDonno, yellow line; *Xylella fastidiosa* subsp. *pauca* strain 9a5c, red line) were scanned and for each position ( $x$ -axis) the tetra correlation value calculated between the 100,000 nucleotides following that position and the entire chromosome sequence was reported in the  $y$ -axis. The blue ticks mark the position on the chromosome of strain 9a5c of the ORFs for the proteins that were found to be missing an ortholog in the *Xanthomonadales* in this work.

### 3. Discussion

In this work, by applying an orthology grouping approach on different gene sets and different genomes, we showed that at least one-fifth of the genes of the pan-genome of *Xylella fastidiosa* originated in distant branches of the evolutionary tree and was presumably acquired by HGT. The pan-genome defines characteristics that may affect the host range or the differences among subspecies or strains [24], therefore several features that delineate the ecological and physiological peculiarity of *Xylella fastidiosa* may indeed rely on genetic material common to organisms that share its niche rather than its phylogeny. HGT has played an important role in bacterial evolution and in facilitating the origins of bacterial diversity, on which the concept of adaptive HGT has been built [27]. Further investigation of the time when the introduction of the genetic determinants of diversity impacted the evolutionary history of a pathogen may shed light on the key phases of its adaptation to the environment.

#### 3.1. BUSCO Analysis Confirmed the Reduction of *Xylella fastidiosa* Genome

The BUSCO gene dataset compiled for the *Gammaproteobacteria* was compared with the genes borne by *Xanthomonas vasicola*, *Xanthomonas albilineans* and *Xylella fastidiosa* genomes,

showing that these three species are missing allegedly ubiquitous genes, accordingly with the genome reduction that characterized their evolution [7,20]. The extensive genome reduction observed in some xanthomonads has been connected with their endophytic lifestyle and their typical commitment to a reduced host range [28]. It is therefore noticeable that *Xylella fastidiosa* was the species with the highest number of missing genes, despite being characterized by a broad host range. *Xylella fastidiosa* misses the regulatory protein RecX, a small protein inhibiting RecA [29]. RecA is involved in the induction of the SOS response [30], essential to increase bacterial tolerance to DNA damage associated with several factors including HGT [31]. Thus, for *Xylella fastidiosa* the absence of RecX may imply enhanced exchange of genetic material even in the absence of strong selective pressure.

### 3.2. Distantly Related Bacteria Contributed to the Shape of the *Xylella fastidiosa* Accessory Genome

For the genome comparison of the three species considered, we preliminary focused the analysis on the core genome, which is the pool of genes shared by all the strains of the same taxon and hence differs from the BUSCO set for the inclusion of genes that are not ubiquitous, but typical of that species. When compared to the BUSCO set, the core genome set highlights the genes that were exploited by the species to gain access to the ecological niche that allowed speciation. Instead, the third set, the pan-genome, highlights the strain-specific features that were exploited for strain differentiation, likely associated with adaptation to a new host or environment.

*Xanthomonas vasicola* was found to have a compact core genome (representing roughly half of the whole genome), as already pointed out by Studholme and coworkers [32], who associated the reductive evolution tendency of *Xanthomonas vasicola* core genome with the genomic gains arisen by recent genetic exchange with other bacteria. Rodriguez-R and coworkers [7] reported a level of genomic loss and gain that was almost twice as large as that exhibited by *Xanthomonas albilineans*, suggesting that the *Xanthomonas vasicola* genomes are very dynamic. Our study detected no gene of the core genome without an ortholog or close paralog within the *Xanthomonadales*, while about 100 genes (2%) in the pan-genome were the result of HGT from distant taxa. In summary, the dynamic of the large accessory genomes of *Xanthomonas vasicola* is more relevant in the diversification of strains and their virulence features rather than in the context of the definition of the primary ecological niche that leads to speciation.

Differently from *Xanthomonas vasicola*, both the core genome and pan-genome of *Xanthomonas albilineans* are characterized by the presence of genes with origin by HGT from distant taxa. Nevertheless, in the *Xanthomonas albilineans* pan-genome only a few of these genes may have an impact on adaptation and most of them were actually found within the core genome set.

Consistently with the narrow host range of *Xanthomonas albilineans*, limited to sugarcane and few other monocots in the *Poaceae* family [33], here the contribution of HGT from distant taxa has been more relevant in defining the characteristics of the species rather than in strain-to-strain diversification. Some of the genes inherited from distant taxa, such as serine hydrolase, different degrading enzymes, non-ribosomal peptide synthetases and the two methyl transferases, are regarded as virulence factors [34–38], suggesting the possibility that *Xanthomonas albilineans* exploited external genomic resources to cope with host defense in a phytopathological context that is not typical of the xanthomonads.

*Xylella fastidiosa* is yet another case. Similarly to *Xanthomonas albilineans*, a few genes of the core genome were identified as not originating within the *Xanthomonadales*. On the other hand, in the pan-genome the number of the genes presumably transferred horizontally from distant taxa is extremely large, suggesting the intense trading with environmental bacteria as a major determinant of strain diversity. Further evidence of recent introduction in the genomes comes from the detection of DNA stretches with very high sequence similarity to bacteria widely spread in the environment, although not typically associated to the xylem: only a very recent gene exchange may explain the extremely high similarity

at the DNA level of the 6697 nts region that we detected in the genome of *Xylella fastidiosa* subsp. *pauca* 9a5c and related strains with the corresponding region in the genome of the distantly related *Pseudomonas* sp. CCA1, a lignin-degrading bacterium isolated from soil in Japan for which the name *Pseudomonas humi* has been proposed [39]. Connections between rhizosphere and phyllosphere bacterial populations have already been reported [40].

The genome plasticity of *Xylella fastidiosa* is well known, as it was for *Xanthomonas vasicola*; it was shown in ref 7 that a large gene flux characterizes strain diversity. We confirmed the large size of the accessory genomes, but we also showed that the gene reservoir accessible to *Xanthomonas vasicola* is primarily contained within the order *Xanthomonadales*. Conversely, *Xylella fastidiosa* showed a similar largeness of the accessory genome, despite its reduced genome size, but the reference reservoir for this species does not appear to be confined within the order boundaries, because it extensively draws genes from the bacterial divisions that are most abundant in the plant endosphere.

The pool of genes that has been gained from distantly related bacteria is part primarily of the accessory and only to a lesser extent of the core genome, therefore we deduce that those genes play a role in strain-to-strain differentiation even more markedly than in shaping the features needed for the access to the habitat that is typical of the whole species. In other words, the results presented here highlight the relevance of the genes gained from distantly related environmental bacteria in subspecies and strain differentiation, and in determining host specificity.

### 3.3. *Xylella fastidiosa* Gained Different Functions from Various Clades

*Xylella fastidiosa* genes with homology only to genes found in distantly related bacteria were classified, enriching the annotation with ko-terms using the KEGG facilities. The analysis showed that bacteria belonging to the order *Burkholderiales* contributed to almost one-tenth of the entire *Xylella fastidiosa* pan-genome, covering different functional categories. These data match with the fact that the genes of *Xylella fastidiosa* associated with the type IV secretion system were particularly abundant and were similar to their homologs in the *Burkholderiales*. Type IV secretion system genes mediate the transfer of DNA and proteins from cell to cell in various biological contexts, including conjugation [40]. Conceivably, the availability of a conjugation system in common with the *Betaproteobacteria* results from, but also promotes, the HGT with bacteria belonging to this large and diverse bacterial division.

Concerning the contributions to *Xylella fastidiosa* diversity given by other bacterial orders, beside *Burkholderiales*, it is worth naming the genes of the TA system, which encode a toxin protein capable of inhibiting cell growth and an antitoxin that counteracts the toxin. In *Xylella fastidiosa*, the TA system genes can be encoded by both plasmids and chromosomes, but all genes coding for TA systems derived by HGT with phylogenetically distant bacteria are located on plasmids. Plasmid-encoded TA systems confer stable plasmid inheritance by post-segregational killing of plasmid-free daughter cells [41]. Their diverse origin is in line with the great variability characterizing TA genes in the *Xylella fastidiosa* genome, which can contain multiple copies of the TA prophage region, with each copy bearing a different TA system, depending on the strain or the subspecies [42]. A further relevant finding concerns the relatively large number of genes encoding putative degrading and biosynthetic activities that have no ortholog or close paralog within the *Xanthomonadales*, and appear to originate in widely diverse regions of the prokaryotic tree of life. These genes add to the significant fraction of the genome of *Xylella fastidiosa* that is not annotated due to the lack of similarity with proteins with known function.

### 3.4. *Xylella fastidiosa* and the Xylem Sap Microbiome

Sequence similarity between *Xylella fastidiosa* and bacteria belonging to the orders *Burkholderiales*, *Rhizobiales*, and *Rhodocyclales* has already been reported [43–47]. Nevertheless, the impact on the *Xylella fastidiosa* pan-genome given by genes coming from evolutionary distant bacteria, systemically analyzed in this study, may appear in contrast with the pathogen lifestyle, confined into xylem vessels.

Xylem is populated by many other xylem-limited bacteria [48,49], and also by endophytes that use it as the main transport route for systemic colonization of the whole plant [50–52]. Studies on the composition of the xylem sap microbiome in citrus plants and grapevines showed not only that *Xylella fastidiosa* shares its ecological niche with many other bacteria endophytes, but also that it interacts with them modulating its growth and its gene expression [53,54]. The interconnections between *Xylella fastidiosa* and other bacterial components of the phyllosphere and the natural inclination of *Xylella fastidiosa* to take up DNA from its environment [55] allow us to hypothesize, in the evolutionary history of *Xylella fastidiosa*, frequent genomic exchanges with a variety of phyllosphere inhabiting bacteria, possibly favored by the recruitment of a plasmid conjugation machinery from the *Betaproteobacteria* (that prompted genetic exchange with phyllosphere bacteria) and by the loss of RecX (that enhanced the plasticity of the genome).

The access to this large gene reservoir may have had (and still may have) a role in *Xylella fastidiosa* adaptation capability. It is well known that *Xylella fastidiosa* strains can recombine with each other [56] and recombination among strains is regarded as a pivotal condition for the emergence of new variants with an enlarged or modified host range. For this reason, after the introduction of *Xylella fastidiosa* subsp. *pauca* in Apulia, a major concern in Europe has been the prevention of the introduction of new strains with which the current epidemic strain may recombine and enlarge the host range [46]. Our work, however, stresses the relevance of genes that originated outside the genus *Xylella*. Without questioning the relevance of the risk of creating new genotypes by intraspecific recombination, we evidenced the potential role of HGT from the environmental microbiota for the phytopathogen adaptation and the possibility that preventing intraspecific recombination may not guarantee the avoidance of shifts to new hosts.

## 4. Materials and Methods

### 4.1. Obtaining the Pan-Genomes

Datasets were constructed for three groups of bacteria representing taxonomically and genomically well defined clusters: (i) 65 genomes of *Xanthomonas vasicola*, including *Xanthomonas vasicola* pv. *holcicola* (22 genomes), *Xanthomonas vasicola* pv. *vasculorum* (43 genomes), and 8 genomes of the maize isolates previously named *Xanthomonas vasicola* pv. *zeae*; (ii) 16 genomes of *Xanthomonas albilineans*; (iii) 45 representative genomes of *Xylella fastidiosa* (this figure is lower than the number of all assemblies presently available from public database, which would exceed the computational capacity). Table S1 lists the assemblies used and the total sizes of the predicted proteomes, which range from about 730,000 to about 880,000 amino acids.

The downloaded NCBI-annotated sequences of the genomes of each species (*Xanthomonas vasicola*, *Xanthomonas albilineans*, *Xylella fastidiosa*) were submitted to the standalone version of the Orthologous Matrix tool (OMA-standalone; <http://omabrowser.org/standalone/>) to identify orthologous groups (OGs) of genes. The output was further analyzed with several custom Perl/Bash scripts for the construction of three gene sets, named BUSCO, Core, and Pan-genome. Gene sequences found in only one strain (singletons) were not included in the above-mentioned gene sets. Hence, the pan-genomes identified in this work differ from those publicly available from UNIPROT ([ftp://ftp.uniprot.org/pub/databases/uniprot/current\\_release/knowledgebase/pan\\_proteomes/](ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/pan_proteomes/)) and roughly correspond to the concept of the shell genome used by Vanhove and coworkers [56].

### 4.2. Classification of Gene According to Orthology and Homology Searches

As a reference representative gene product of each OGs, the centroid amino acid sequence resulting from the pairwise distance calculation carried out with ClustalW2 [57] was selected (OGRP: orthologous group reference protein). Each OGRP was submitted to the analysis flow outlined in Figure S1 and described later on in this section, for the identification of genes not shared with other *Xanthomonadales*. The procedure in Figure S1 is

based on the integration of the results of BLASTP searches and interrogations of the OMA database, obtained as follows.

Each OG was submitted to BLASTP analysis using a custom database from which all the entries from the genus/species investigated were excluded, to identify the bacteria carrying the most similar genes. The preliminary filtering of the database was carried out with an ad hoc BASH script on the NCBI reference sequence (RefSeq) database. Three BLASTP search sets were carried out on customized databases made with all prokaryotic sequences excluding: (i) for the *Xanthomonas vasicola* pan-genome, all sequences deposited as belonging to *Xanthomonas vasicola* and its pathovars or to *Xanthomonas campestris* pv. *musacearum*, because comparative phylogenomic studies supported its phylogenetic relatedness to *Xanthomonas vasicola* [32,58]; (ii) for the *Xylella fastidiosa* pan-genome, all sequences deposited as belonging to the genus *Xylella*, including *Xylella taiwanensis*; (iii) for the *Xanthomonas albilineans* pan-genome, all sequences deposited as belonging to *Xanthomonas albilineans*, *Xanthomonas pseudoalbilineans*, and the related strains *Xanthomonas* sp. MUS 60 and GPE39, for which the name “*Xanthomonas pseudoalbilineans*” has been proposed. The results of the BLASTP search were then filtered by keeping only hits with e-values lower than  $1 \times 10^{-6}$ . The BLASTP results were reformatted and each homologous group was thus provided with a taxonomic identifier using the Entrez e-service [59]. When a BLASTP search indicated a best match in a taxon not comprised in the *Xanthomonadales*, the result was further validated with a tBLASTn search in custom databases built from all *Xanthomonadales* genomes included in the JGI database (downloaded in December 2019).

All identifiers assigned by the OMA database (OMAdb; [21]) to the species under investigation were downloaded and assigned to the matching OGRP using the cross-reference service provided by the OMAdb. Due to the limitation of the number of genomes included in the OMAdb, several OGRPs had no identifier with which they could be associated. In such cases, the OMAdb was searched by BLASTP and the best fitting identifier was provisionally associated with the querying OGRP.

As illustrated in Figure S1, the first step of the analysis of each OGRP was to check whether or not its best BLASTP hit was within the class *Xanthomonadales*; in this case the taxID of the class was assigned and the OGRP was not further investigated. The second step was to check whether or not the OGRP under examination had an existing entry in the OMAdb. If the corresponding identifier belonged to a hierarchical orthologous group (HOG; [60]) at the level of the *Xanthomonadales*, the taxID of the class was assigned and the OGRP was not further investigated. If an existing entry could not be found, the OMAdb was homology searched for the most similar identifier: if the best fitting identifier itself or one of its orthologs belonged to an HOG at the level of the *Xanthomonadales*, again the taxID of the class was assigned and the OGRP was not further investigated (step 3).

The OGRPs that have not been excluded to this point were further investigated with the aim of determining the taxon of their putative origin. Using the Python interface to the REST-API of the OMAdb [22] for each OG, all protein sequences recorded as orthologs of the OG-associated identifier were then downloaded and aligned with the OGRP using FASMA [61]. In each global alignment, the four sequences with highest identity scores were selected and their lower common ancestor rank determined using the ETE3 toolkit [62].

Finally, the information obtained from the OMAdb was merged with the results of the BlastP searches of the RefSeq database and checked for taxonomic congruence. In the case of conflict, the taxID assigned to the OG was determined by choosing the higher sequence identity score in global alignment to the OG centroid after alignment of the best BlastP hit sequence and that of the closest protein among the putative ortholog list provided by OMAdb. The results were visualized using the Krona tools [63].

#### 4.3. Functional Assignments and Other Analyses

The program SplitByTaxa of the suite BBMap, available at: <https://sourceforge.net/projects/bbmap/>, was used for discarding the representative gene sequence of each OG in agreement to its taxonomic position according to the previously assigned taxID. An ad hoc

Perl script was then used to interrogate the KAAS service [64] for functional classification and KEGG pathway assignment of the representative gene sequences of each OG.

The sliding window plot of tetra correlation values was built using an ad hoc script that calculate tetra z-scores according to [65], and used a window size of 100,000 nts, increasing the positions for calculation by 10,000 nts for each plotted point.

## 5. Conclusions

*Xylella fastidiosa* is phylogenetically strictly related to other xanthomonads, but with a rather atypical, unique phytopathogenic behavior and very wide host range and combination of host–pathogen relationship. In this paper, we show that the *Xylella fastidiosa* pan-genome is not peculiar for its size (it is less than 1.5 times the average genome) nor for its ratio accessory versus core genome. The pan-genome of *Xylella fastidiosa* is instead peculiar for the origin of the genes that it includes, which in a large fraction have not evolved within the *Xanthomonadales* but have rather been acquired by horizontal transfer from distantly related bacteria. The access to the environmental gene pool may have been the key for the species success and, as such, actively pursued.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/2076-0817/10/1/46/s1>. Figure S1: Schematic representation of the procedure used for classification of gene according to orthology and homology searches. Figure S2: Taxonomic subdivision of their nearest ortholog assignment of genes in the pan-genome of *Xanthomonas vasicola*. Figure S3: Taxonomic subdivision of their nearest ortholog assignment of genes in the pan-genome of *Xanthomonas albilineans*. Figure S4: Taxonomic subdivision of their nearest ortholog assignment of genes in the pan-genome of *Xylella fastidiosa*. Figure S5: Sliding window graphs of the tetra correlation values between a 100 k window and the entire chromosome. The DNA sequence of the chromosomes of 12 *X. fastidiosa* strains were scanned and for each position (*x*-axis, nucleotide position in genome) the tetra correlation value calculated between the 100,000 nucleotides following that position and the entire chromosome sequence was reported in the *y*-axis. Table S1: List of assemblies used for the construction of the pan-genomes of *Xanthomonas vasicola*, *Xanthomonas albilineans* and *Xylella fastidiosa* and total sizes of their predicted proteomes. Table S2: Genes of the pan-genome of *Xylella fastidiosa* without HOG in the *Xanthomonadales*, listed according to their KEGG pathways or GO terms. Table S3: List of the annotation of degrading and synthetic enzyme encoding genes in the pan-genome of *Xylella fastidiosa* that have no homologues in the *Xanthomonadales*. The column on the right reports the detected number of groups of ortholog genes found analyzing 45 genomes of *Xylella fastidiosa*.

**Author Contributions:** All three authors conceived the idea; G.F. and L.P. planned the work; G.F. carried out the bioinformatic analyses; L.P. prepared the figures; all three authors wrote the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The study analyzes publicly available data.

**Acknowledgments:** We acknowledge the CyVerse cyberinfrastructure (supported by the National Science Foundation under Award Numbers DBI-0735191, DBI-1265383, and DBI-1743442. URL: [www.cyverse.org](http://www.cyverse.org)) for making available the computational resources used in this work.

**Conflicts of Interest:** The authors declare that they have no conflict of interest.

## References

1. Jeger, M.; Caffier, D.; Candresse, T.; Chatzivassiliou, E.; Dehnen-Schmutz, K.; Gilioli, G.; Grégoire, J.C.; Jaques Miret, J.A.; MacLeod, A.; Navajas Navarro, M.; et al. Updated pest categorisation of *Xylella fastidiosa*. *EFSA J.* **2018**, *16*. [[CrossRef](#)]
2. Janse, J.D.; Obradovic, A. *Xylella fastidiosa*: Its biology, diagnosis, control and risks. *J. Plant Pathol.* **2010**, *92*, S35–S48.
3. Chatterjee, S.; Almeida, R.P.P.; Lindow, S. Living in two worlds: The plant and insect lifestyles of *Xylella fastidiosa*. *Annu. Rev. Phytopathol.* **2008**, *46*, 243–271. [[CrossRef](#)] [[PubMed](#)]
4. Hopkins, D.L.; Purcell, A.H. *Xylella fastidiosa*: Cause of Pierce’s disease of grapevine and other emergent diseases. *Plant Dis.* **2002**, *86*, 1056–1066. [[CrossRef](#)] [[PubMed](#)]

5. Nunney, L.; Vickerman, D.B.; Bromley, R.E.; Russell, S.A.; Hartman, J.R.; Morano, L.D.; Stouthamer, R. Recent evolutionary radiation and host plant specialization in the *Xylella fastidiosa* subspecies native to the United States. *Appl. Environ. Microbiol.* **2013**, *79*, 2189–2200. [[CrossRef](#)]
6. Chen, J.; Jarret, R.L.; Qin, X.; Hartung, J.S.; Banks, D.; Chang, C.J.; Hopkins, D.L. 16S rDNA sequence analysis of *Xylella fastidiosa* strains. *Syst. Appl. Microbiol.* **2000**, *23*, 349–354. [[CrossRef](#)]
7. Rodriguez-R, L.M.; Grajales, A.; Arrieta-Ortiz, M.L.; Salazar, C.; Restrepo, S.; Bernal, A. Genomes-based phylogeny of the genus *Xanthomonas*. *BMC Microbiol.* **2012**, *12*, 43. [[CrossRef](#)]
8. Su, C.-C.; Deng, W.-L.; Jan, F.-J.; Chang, C.-J.; Huang, H.; Shih, H.-T.; Chen, J. *Xylella taiwanensis* sp. nov., causing pear leaf scorch disease. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 4766–4771. [[CrossRef](#)]
9. Schaad, N.W.; Postnikova, E.; Lacy, G.; Fatmi, M.; Chang, C.-J. *Xylella fastidiosa* subspecies: *X. fastidiosa* subsp. [correction] *fastidiosa* [correction] subsp. nov., *X. fastidiosa* subsp. *multiplex* subsp. nov., and *X. fastidiosa* subsp. *pauca* subsp. nov. *Syst. Appl. Microbiol.* **2004**, *27*, 290–300. [[CrossRef](#)]
10. Marcelletti, S.; Scortichini, M. Genome-wide comparison and taxonomic relatedness of multiple *Xylella fastidiosa* strains reveal the occurrence of three subspecies and a new *Xylella* species. *Arch. Microbiol.* **2016**, *198*, 803–812. [[CrossRef](#)]
11. Denancé, N.; Briand, M.; Gaborieau, R.; Gaillard, S.; Jacques, M.-A. Identification of genetic relationships and subspecies signatures in *Xylella fastidiosa*. *BMC Genom.* **2019**, *20*, 239. [[CrossRef](#)] [[PubMed](#)]
12. Scally, M.; Schuenzel, E.L.; Stouthamer, R.; Nunney, L. Multilocus sequence type system for the plant pathogen *Xylella fastidiosa* and relative contributions of recombination and point mutation to clonal diversity. *Appl. Environ. Microbiol.* **2005**, *71*, 8491–8499. [[CrossRef](#)] [[PubMed](#)]
13. Nunney, L.; Hopkins, D.L.; Morano, L.D.; Russell, S.E.; Stouthamer, R. Intersubspecific recombination in *Xylella fastidiosa* strains native to the United States: Infection of novel hosts associated with an unsuccessful invasion. *Appl. Environ. Microbiol.* **2014**, *80*, 1159–1169. [[CrossRef](#)] [[PubMed](#)]
14. Nunney, L.; Ortiz, B.; Russell, S.A.; Ruiz Sánchez, R.; Stouthamer, R. The complex biogeography of the plant pathogen *Xylella fastidiosa*: Genetic evidence of introductions and subspecific introgression in Central America. *PLoS ONE* **2014**, *9*, e112463. [[CrossRef](#)] [[PubMed](#)]
15. Coletta-Filho, H.D.; Francisco, C.S.; Lopes, J.R.S.; Muller, C.; Almeida, R.P.P. Homologous recombination and *Xylella fastidiosa* host-pathogen associations in South America. *Phytopathology* **2017**, *107*, 305–312. [[CrossRef](#)]
16. Potnis, N.; Kandel, P.P.; Merfa, M.V.; Retchless, A.C.; Parker, J.K.; Stenger, D.C.; Almeida, R.P.P.; Bergsma-Vlami, M.; Westenberg, M.; Cobine, P.A.; et al. Patterns of inter-and intrasubspecific homologous recombination inform eco-evolutionary dynamics of *Xylella fastidiosa*. *ISME J.* **2019**, *13*, 2319–2333. [[CrossRef](#)]
17. Castillo, A.I.; Chacón-Díaz, C.; Rodríguez-Murillo, N.; Coletta-Filho, H.D.; Almeida, R.P.P. Impacts of local population history and ecology on the evolution of a globally dispersed pathogen. *BMC Genom.* **2020**, *21*. [[CrossRef](#)]
18. Pieretti, I.; Pesic, A.; Petras, D.; Royer, M.; Suessmuth, R.D.; Cociancich, S. What makes *Xanthomonas albilineans* unique amongst xanthomonads? *Front. Plant Sci.* **2015**, *6*, 289. [[CrossRef](#)]
19. Perez-Quintero, A.L.; Ortiz-Castro, M.; Lang, J.M.; Rieux, A.; Wu, G.; Liu, S.; Chapman, T.A.; Chang, C.; Ziegler, J.; Peng, Z.; et al. Genomic acquisitions in emerging populations of *Xanthomonas vasicola* pv. *vasculorum* infecting corn in the United States and Argentina. *Phytopathology* **2020**, *110*, 1161–1173. [[CrossRef](#)]
20. Pieretti, I.; Royer, M.; Barbe, V.; Carrere, S.; Koebnik, R.; Cociancich, S.; Couloux, A.; Darrasse, A.; Gouzy, J.; Jacques, M.-A.; et al. The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited *Xanthomonadaceae*. *BMC Genom.* **2009**, *10*, 616. [[CrossRef](#)]
21. Altenhoff, A.M.; Glover, N.M.; Train, C.-M.; Kaleb, K.; Warwick Vesztrocy, A.; Dylus, D.; de Farias, T.M.; Zile, K.; Stevenson, C.; Long, J.; et al. The OMA orthology database in 2018: Retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. *Nucleic Acids Res.* **2018**, *46*, D477–D485. [[CrossRef](#)] [[PubMed](#)]
22. Kaleb, K.; Vesztrocy, A.W.; Altenhoff, A.; Dessimoz, C. Expanding the Orthologous Matrix (OMA) programmatic interfaces: REST API and the OmaDB packages for R and Python. *F1000Research* **2019**, *8*, 42. [[CrossRef](#)] [[PubMed](#)]
23. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)] [[PubMed](#)]
24. Medini, D.; Crabtree, J.; Mora, M.; Peterson, J.; Tettelin, H.; Angiuoli, S.; Ros, I.; Ward, N.; Donati, C.; Scarselli, M.; et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome”. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13950–13955.
25. De Carvalho, M.O.; Loreto, E.L.S. Methods for detection of horizontal transfer of transposable elements in complete genomes. *Genet. Mol. Biol.* **2012**, *35*, 1078–1084. [[CrossRef](#)] [[PubMed](#)]
26. Lawrence, J.G.; Ochman, H. Amelioration of bacterial genomes: Rates of change and exchange. *J. Mol. Evol.* **1997**, *44*, 383–397. [[CrossRef](#)]
27. Wiedenbeck, J.; Cohan, F.M. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol. Rev.* **2011**, *35*, 957–976. [[CrossRef](#)]
28. Vandroemme, J.; Cottyn, B.; Baeyen, S.; de Vos, P.; Maes, M. Draft genome sequence of *Xanthomonas fragariae* reveals reductive evolution and distinct virulence-related gene content. *BMC Genom.* **2013**, *14*. [[CrossRef](#)]

29. Shvetsov, A.V.; Lebedev, D.V.; Chervyakova, D.B.; Bakhlanova, I.V.; Yung, I.A.; Radulescu, A.; Kuklin, A.I.; Baitin, D.M.; Isaev-Ivanov, V.V. Structure of RecX protein complex with the presynaptic RecA filament: Molecular dynamics simulations and small angle neutron scattering. *FEBS Lett.* **2014**, *588*, 948–955. [[CrossRef](#)]
30. Cirz, R.T.; Chin, J.K.; Andes, D.R.; de Crécy-Lagard, V.; Craig, W.A.; Romesberg, F.E. Inhibition of mutation and combating the evolution of antibiotic resistance. *PLoS Biol.* **2005**, *3*, e176. [[CrossRef](#)]
31. Beaver, J.W.; Hochhut, B.; Waldor, M.K. SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* **2004**, *427*, 72–74. [[CrossRef](#)] [[PubMed](#)]
32. Studholme, D.J.; Kemen, E.; MacLean, D.; Schornack, S.; Aritua, V.; Thwaites, R.; Grant, M.; Smith, J.; Jones, J.D.G. Genome-wide sequencing data reveals virulence factors implicated in banana *Xanthomonas* wilt. *FEMS Microbiol. Lett.* **2010**, *310*, 182–192. [[CrossRef](#)] [[PubMed](#)]
33. Birch, R.G. *Xanthomonas albilineans* and the antipathogenesis approach to disease control. *Mol. Plant Pathol.* **2001**, *2*, 1–11. [[CrossRef](#)] [[PubMed](#)]
34. Park, H.-J.; Jung, B.; Lee, J.; Han, S.-W. Functional characterization of a putative DNA methyltransferase, EadM, in *Xanthomonas axonopodis* pv. *glycines* by proteomic and phenotypic analyses. *Sci. Rep.* **2019**, *9*, 2446. [[CrossRef](#)] [[PubMed](#)]
35. Rajeshwari, R.; Jha, G.; Sonti, R.V. Role of an in planta-expressed xylanase of *Xanthomonas oryzae* pv. *oryzae* in promoting virulence on rice. *Mol. Plant-Microbe Interact.* **2005**, *18*, 830–837. [[CrossRef](#)] [[PubMed](#)]
36. Wessel, M.; Klüsener, S.; Gödeke, J.; Fritz, C.; Hacker, S.; Narberhaus, F. Virulence of *Agrobacterium tumefaciens* requires phosphatidylcholine in the bacterial membrane. *Mol. Microbiol.* **2006**, *62*, 906–915. [[CrossRef](#)]
37. Farberg, A.M.; Hart, W.K.; Johnson, R.J. The unusual substrate specificity of a virulence associated serine hydrolase from the highly toxic bacterium, *Francisella tularensis*. *Biochem. Biophys. Rep.* **2016**, *7*, 415–422. [[CrossRef](#)]
38. Kulasakara, H.; Lee, V.; Brencic, A.; Liberati, N.; Urbach, J.; Miyata, S.; Lee, D.G.; Neely, A.N.; Hyodo, M.; Hayakawa, Y.; et al. Analysis of *Pseudomonas aeruginosa* diguanylate cyclases and phosphodiesterases reveals a role for bis-(3'-5')-cyclic-GMP in virulence. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 2839–2844. [[CrossRef](#)]
39. Akita, H.; Kimura, Z.I.; Hoshino, T. *Pseudomonas humi* sp. nov., isolated from leaf soil. *Arch. Microbiol.* **2019**, *201*, 245–251. [[CrossRef](#)]
40. Bai, Y.; Müller, D.B.; Srinivas, G.; Garrido-Oter, R.; Potthoff, E.; Rott, M.; Dombrowski, N.; Münch, P.C.; Spaepen, S.; Remus-Emsermann, M.; et al. Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* **2015**, *528*, 364–369. [[CrossRef](#)]
41. Hayes, F. Toxins-antitoxins: Plasmid maintenance, programmed cell death, and cell cycle arrest. *Science* **2003**, *301*, 1496–1499. [[CrossRef](#)] [[PubMed](#)]
42. Lee, M.W.; Tan, C.C.; Rogers, E.E.; Stenger, D.C. Toxin-antitoxin systems mqsR/ygiT and dinJ/reLE of *Xylella fastidiosa*. *Physiol. Mol. Plant Pathol.* **2014**, *87*, 59–68. [[CrossRef](#)]
43. Kaneko, T.; Nakamura, Y.; Sato, S.; Asamizu, E.; Kato, T.; Sasamoto, S.; Watanabe, A.; Idesawa, K.; Ishikawa, A.; Kawashima, K.; et al. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res.* **2000**, *7*, 331–338. [[CrossRef](#)] [[PubMed](#)]
44. Reddy, J.D.; Reddy, S.L.; Hopkins, D.L.; Gabriel, D.W. TolC is required for pathogenicity of *Xylella fastidiosa* in *Vitis vinifera* grapevines. *Mol. Plant Microbe Interact.* **2007**, *20*, 403–410. [[CrossRef](#)] [[PubMed](#)]
45. Guilhabert, M.R.; Kirkpatrick, B.C. Identification of *Xylella fastidiosa* antivirulence genes: Hemagglutinin adhesins contribute a biofilm maturation to *X. fastidiosa* and colonization and attenuate virulence. *Mol. Plant Microbe Interact.* **2005**, *18*, 856–868. [[CrossRef](#)] [[PubMed](#)]
46. Van Sluys, M.A.; de Oliveira, M.C.; Monteiro-Vitorello, C.B.; Miyaki, C.Y.; Furlan, L.R.; Camargo, L.E.A.; da Silva, A.C.R.; Moon, D.H.; Takita, M.A.; Lemos, E.G.M.; et al. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. *J. Bacteriol.* **2003**, *185*, 1018–1026. [[CrossRef](#)] [[PubMed](#)]
47. Roper, M.C.; Greve, L.C.; Warren, J.G.; Labavitch, J.M.; Kirkpatrick, B.C. *Xylella fastidiosa* requires polygalacturonase for colonization and pathogenicity in *Vitis vinifera* grapevines. *Mol. Plant-Microbe Interact.* **2007**, *20*, 411–419. [[CrossRef](#)]
48. Purcell, A.H.; Hopkins, D.L. Fastidious xylem-limited bacterial plant pathogens. *Annu. Rev. Phytopathol.* **1996**, *34*, 131–151. [[CrossRef](#)]
49. Achari, G.A.; Ramesh, R. Diversity, biocontrol, and plant growth promoting abilities of xylem residing bacteria from solanaceous crops. *Int. J. Microbiol.* **2014**. [[CrossRef](#)]
50. James, E.K.; Gyaneshwar, P.; Mathan, N.; Barraquio, W.L.; Reddy, P.M.; Iannetta, P.P.M.; Olivares, F.L.; Ladha, J.K. Infection and colonization of rice seedlings by the plant growth-promoting bacterium *Herbaspirillum seropedicae* Z67. *Mol. Plant-Microbe Interact.* **2002**, *15*, 894–906. [[CrossRef](#)]
51. Compant, S.; Reiter, B.; Sessitsch, A.; Nowak, J.; Clément, C.; Barka, E.A. Endophytic colonization of *Vitis vinifera* L. by plant growth-promoting bacterium *Burkholderia* sp. strain PsJN. *Appl. Environ. Microbiol.* **2005**, *71*, 1685–1693. [[CrossRef](#)] [[PubMed](#)]
52. Compant, S.; Kaplan, H.; Sessitsch, A.; Nowak, J.; Ait Barka, E.; Clément, C. Endophytic colonization of *Vitis vinifera* L. by *Burkholderia phytofirmans* strain PsJN: From the rhizosphere to inflorescence tissues. *FEMS Microbiol. Ecol.* **2008**, *63*, 84–93. [[CrossRef](#)] [[PubMed](#)]
53. Lacava, P.T.; Azevedo, J.L.; Miller, T.A.; Hartung, J.S. Interactions of *Xylella fastidiosa* and endophytic bacteria in citrus: A review. *Tree For. Sci. Biotechnol.* **2009**, *3*, 40–48.

54. Deyett, E.; Rolshausen, P.E. Temporal dynamics of the sap microbiome of grapevine under high Pierce's disease pressure. *Front. Plant Sci.* **2019**, *10*. [[CrossRef](#)]
55. Kung, S.H.; Retchless, A.C.; Kwan, J.Y.; Almeida, R.P.P. Effects of DNA size on transformation and recombination efficiencies in *Xylella fastidiosa*. *Appl. Environ. Microbiol.* **2013**, *79*, 1712–1717. [[CrossRef](#)]
56. Vanhove, M.; Retchless, A.C.; Sicard, A.; Rieux, A.; Coletta-Filho, H.D.; de La Fuente, L.; Stenger, D.C.; Almeida, R.P.P. Genomic diversity and recombination among *Xylella fastidiosa* subspecies. *Appl. Environ. Microbiol.* **2019**, *85*, e02972-18. [[CrossRef](#)] [[PubMed](#)]
57. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)]
58. Wasukira, A.; Tayebwa, J.; Thwaites, R.; Paszkiewicz, K.; Aritua, V.; Kubiriba, J.; Smith, J.; Grant, M.; Studholme, D.J. Genome-wide sequencing reveals two major sub-lineages in the genetically monomorphic pathogen *Xanthomonas campestris* pathovar *musacearum*. *Genes* **2012**, *3*, 361–377. [[CrossRef](#)]
59. Maglott, D.; Ostell, J.; Pruitt, K.D.; Tatusova, T. Entrez gene: Gene-centered information at NCBI. *Nucleic Acids Res.* **2011**, *39*, D52–D57. [[CrossRef](#)]
60. Altenhoff, A.M.; Gil, M.; Gonnet, G.H.; Dessimoz, C. Inferring hierarchical orthologous groups from orthologous gene pairs. *PLoS ONE* **2013**, *8*, e53786. [[CrossRef](#)]
61. Deorowicz, S.; Debudaj-Grabysz, A.; Gudys, A. FAMSA: Fast and accurate multiple sequence alignment of huge protein families. *Sci. Rep.* **2016**, *6*, 33964. [[CrossRef](#)] [[PubMed](#)]
62. Huerta-Cepas, J.; Serra, F.; Bork, P. ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **2016**, *33*, 1635–1638. [[CrossRef](#)] [[PubMed](#)]
63. Ondov, B.D.; Bergman, N.H.; Phillippy, A.M. Interactive metagenomic visualization in a Web browser. *BMC Bioinform.* **2011**, *12*, 385. [[CrossRef](#)] [[PubMed](#)]
64. Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A.C.; Kanehisa, M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **2007**, *35*, W182–W185. [[CrossRef](#)]
65. Teeling, H.; Waldmann, J.; Lombardot, T.; Bauer, M.; Glöckner, F.O. TETRA: A web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC Bioinform.* **2004**, *5*, 163. [[CrossRef](#)]