environmental microbiology reports

Environmental Microbiology Reports (2022) **14**(2), 274–285 doi:10.1111/1758-2229.13048

Brief Report

Complete genome assembly of the levan-positive strain PVFi1 of *Pseudomonas savastanoi* pv. *savastanoi* isolated from olive knots in Central Italy

Silvia Turco,[1] Mounira Inas Drais,[1] Luca Rossini,[1] Elena Chaboteaux,[1] Yaseen Jundi Rahi,[1,2] Giorgio Mariano Balestra,[1] Nicola Sante Iacobellis[3] and Angelo Mazzaglia[1]*

[1]*Dipartimento di Scienze Agrarie e Forestali, Università degli Studi della Tuscia, Via S. Camillo de Lellis snc, Viterbo, 01100, Italy.*

[2]*CIHEAM-Mediterranean Agronomic Institute of Bari, Via Ceglie 9, Valenzano, 70010, Italy.*

[3]*Former Professor of Plant Pathology at Università degli Studi della Basilicata, Italy.*

Summary

***Pseudomonas savastanoi* pv. *savastanoi*, the causal agent of olive knot disease, is a fluorescent Gram-negative bacterium classified, according to the specific LOPAT profile, as Ib. However, during the 90s, a number of atypical non-fluorescent levan-positive strains of *Pseudomonas savastanoi* pv. *savastanoi* have been unexpectedly isolated from olive knots in Central Italy. Since its first report, several studies were conducted on this species variant, but its genome sequence has never been reported. The complete genome sequence and two additional plasmids of PVFi1, a representative strain, were here obtained using a hybrid sequencing approach with both Oxford Nanopore Technology and Illumina sequencing. A thorough genomic analysis unravelled several genetic features of this peculiar strain, showing a transposase insertion downstream a fragmented copy of the levansucrase gene. The same features were previously reported on levan-negative *Pseudomonas savastanoi* pv. *savastanoi* strains. In addition, a second copy of the levansucrase gene fully equipped for a gene expression and comparable to the levan-positive *Pseudomonas savastanoi* pv. *glycinea*, may explain the levan-positive test. This result provides a solid genetic demonstration that the bacterial species *Pseudomonas savastanoi* contains either levan-positive or levan-negative strains, providing insights for an update of the related LOPAT classification.**

Received 6 December, 2021; accepted 23 January, 2022. *For correspondence. E-mail angmazza@unitus.it; Tel. +39 0761 357339; Fax +39 0761 357434.

Introduction

Pseudomonas savastanoi is a monomorphic pathogenic Gram-negative gamma-proteobacterium, belonging to the genomospecies 2 of *Pseudomonas syringae* complex (Gardan *et al*., 1992). The strains referable to this species are further classified in pathovars, according to their respective host range. *Pseudomonas savastanoi* pv. *savastanoi* (*Psv*), pv. *fraxini* (*Psf*), pv. *nerii* (*Psn*) and pv. *retacarpa* (*Psr*) attack mainly woody hosts belonging to different species of the families *Oleaceae*, *Fabaceae*, *Myrtaceae*, *Apocinaceae* and *Lythraceae* (Bradbury, 1986; Gardan *et al*., 1992; Iacobellis *et al*., 1998; Janse, 1982; Marchi *et al*., 2011; Moretti *et al*., 2017; de los Rios, 1999). *Pseudomonas savastanoi* pv. *glycinea* (*Psg*) and pv. *phaseolicola* (*Pph*) infect instead herbaceous hosts, soybean (Young *et al*., 1978) and beans or mulberry (Völksch and Weingart, 1997; Dworkin, 2006), respectively.

Within the species *Pseudomonas savastanoi*, the pathovar *savastanoi* is responsible for the olive knot disease, widely diffused in the Mediterranean basin where this species is considered endemic. Infected plants present galls or knots on young woody organs (i.e., twigs and young branches), whilst leaves or fruits are seldom attacked. Although the viability of adult plants is not endangered, the detrimental effect on vegetative growth and olive yield becomes significant as the incidence of the disease increases (Schroth, 1973; Osman *et al*., 1980; Quesada *et al*., 2010).

A recent population structure analysis revealed a highly complex relationship between and within pathovars, suggesting that further genomic studies are necessary to better characterize this species (Rahi *et al*., 2020).

As a member of the subgroup 1b of the LOPAT scheme (LOPAT meaning a series of determinative tests: L, levan production; O, oxidase production; P, pectinolytic activity; A, arginine dihydrolase production; and T, tobacco hypersensibility), *Pseudomonas savastanoi* has long been classified as levan-negative bacterium producing fluorescent pigments, in comparison with the other members of subgroup Ia, whom *Pseudomonas syringae* belongs, which are levan-positive instead (Lelliott *et al*., 1966).

However, during an epidemiological survey in Central Italy in the 90s, non-fluorescent levan-positive bacteria have been isolated from olive knots, whose identity was molecularly confirmed as *Pseudomonas savastanoi* pv *savastanoi* (Iacobellis *et al*., 1993). Over time this occurrence was repeatedly confirmed, and it is now considered as a well-established population of the pathogen and not a single isolated event (Marchi *et al*., 2005). These strains formed a separated cluster from the others when subjected to an AFLP analysis, either because of the same geographic origin or for being levan-positive (Sisto *et al*., 2007).

Together with alginate, levan is an important component of the extracellular matrix produced by some representatives of *Pseudomonas syringae* group. These two high-molecular-weight molecules represent an important food storage during starvation, but also a physical barrier against the plant defence compounds during the early stage of infection, escaping the pathogen recognition system (Laue et al., 2006; Mann and Wozniak, 2012). Levan is a fructose polymer whose main chain is composed by β2–6 linkages and branching with β2–1 bonds, derived from the sucrose transfructosylation catalysed by the levansucrase enzyme (EC 2.4.1.10, Lsc) that belongs to family 68 of glycoside hydrolases.

The gene responsible for the levansucrase production has been found in Gram positive bacteria like *Bacillus subtilis* or *Streptococcus mutans*, as well as in Gram negative bacteria like *Zymomona mobilis* and *Erwinia amilovora*. Interestingly, this gene has been found in all sequenced *Pseudomonas syringae* strains, with a copy number varying from one (in *Psv* NCPPB 3335 and *Pseudomonas syringae* pv. *oryzae* I_6) to three in *Pseudomonas syringae* pv. *tomato* DC3000 and T1 strains, in *Pseudomonas savastanoi* pv *phaseolicola* 1448A and in *Pseudomonas syringae* pv. *actinidiae* (O'Brien *et al*., 2011; Luti *et al*., 2021).

The genetic structure of the genes encoding for the levan production and its regulatory system in *Pseudomonas* has been thoroughly described in *Pseudomonas savastanoi* pv. *glycinea* PG4180 (Hettwer *et al*., 1998; Li and Ullrich, 2001; Srivastava *et al*., 2012; Khandekar *et al*., 2014) and more recently in *Pseudomonas syringae* pv. *actinidiae* (*Psa*) (Luti *et al*., 2021). It has been ascertained that three copies of the levansucrase gene are present within the genome of these strains, two of which expressed differentially (*lscB* and *lscC*) whilst *lscA* is not expressed at all. Notably, *lscA* in PG4180 likely derives from *Erwinia*, while variants of *lscB* and *lscC* were found only in *Pseudomonas syringae* pathovars (Srivastava *et al*., 2012). Further studies on *Pseudomonas savastanoi* pv. *glycinea* PG4180 showed that the main difference between *lscA* and *lscB*/*C* resides in the absence, in *lscA*, of a 48 bp region at the N-terminus of *lscB*/*C* ORF (MSTSSSAVSQLKNSPL) and the presence instead of a N-terminus of 21 bp (WSRADAL). In addition, a 450 bp conserved region upstream the *lscB*/*C* ORF is not present upstream of *lscA*. This entire region of about 530 bp shows significant similarity with a prophage borne DNA that prompted its naming as phage-associated promoter element (PAPE), also referred as *com* gene (Li and Ullrich, 2001; Srivastava *et al*., 2012). The transcriptional start site for *lscB*/*C* was exactly determined by Khandekar *et al*. (2014) and located at nucleotide position −339 bp upstream of the translational start codon of *lscB*/*C* and it was also demonstrated that the lack of this upstream region in *lscA* hampers the gene expression.

The pool of information about the levan production in *Pseudomonas savastanoi* pv. *glycinea* and *Pseudomonas syringae* pv. *actinidiae* leads to the hypothesis that similar structures can be found in the strains of the peculiar levan-positive *Pseudomonas savastanoi* pv. *savastanoi*, even though no information about the complete genome is present in the current literature.

With this work, hence, we aim to carefully reconstruct the complete genome sequence of one representative of the levan-positive *Pseudomonas savastanoi* pv. *savastanoi*, the strain PVFi1, through a hybrid sequencing approach using both Oxford Nanopore and Illumina technologies.

Thus, the genome sequence underwent a meticulous analysis of the gene regions responsible for levan production, i.e. the levansucrase gene and its regulatory system, to assess the presence of orthologous genes, their arrangement and potential functionality, as well as signals of co-evolution with related bacteria.

The genome was compared to the genomes available in NCBI database of other strains belonging to the same *savastanoi* pathovar, as well as to other representatives of pathovars *fraxini*, *nerii* and pv. *retacarpa* in order to obtain phylogenetic information and investigate the presence of enriched functions and single core genes between defined groups.

The phylogenetic analysis of the bacterium has been reinforced by the exact depiction of its secretion systems, type III secretion effectors, WHOP genes and other virulence related genes.

Results

Sequencing and annotation

Using the hybrid *de novo* assembly approach, the genome sequence of *Pseudomonas savastanoi* pv. *savastanoi* PVFi1 was arranged in one single circular chromosome of 6 253 686 bp, with a GC content of 57.98% and without any undetermined bases. In addition, two complete circular plasmids were assembled as well: the first one showing 72 349 bp long and the second one 41 097 bp (Fig. 1). The complete genome and the related plasmids have been deposited on the NCBI database under the accession numbers CP078139, CP078140 and CP078141, respectively.

The annotation using NCBI PGAP found a total of 5778 genes, 5695 of which are coding. The remaining 83 genes codify for RNA molecules: 16 for rRNAs, 63 for tRNAs and 4 noncoding RNAs. Instead, Prokka identified 5755 CDS on the chromosome, among which 2361 with hypothetical function, 63 tRNAs, 16 rRNAs and 1 tmRNA. Eighty-three and 53 CDS were found in the two plasmids. Among these CDS, the great part was annotated as a hypothetical protein (50 and 44 CDS, respectively), while most of the remaining others are related to transposase and transposon invertase (21 and 9 CDS, respectively).

Although registered using a different nomenclature, three different chromosomal regions associated with the levan production were identified by both Prokka and PGAP: *sacB_1*, *sacB_2* and *sacB_3* (according to *Bacillus subtilis* nomenclature) in Prokka annotation, while, PGAP annotated them as glycoside hydrolase family 68.

Phylogenetic analysis of 40 MLST genes

The phylogenetic relationships among the 21 *Pseudomonas savastanoi* strains are depicted in Fig. 2. The strains belonging to the *fraxini*, *nerii* and *retacarpa* pathovars are grouped into distinct clades. The nine strains belonging to the pathovar *savastanoi*, instead, are divided in two distinct sub-clades: one including seven strains (indicated in Fig. 2 as sav I) and a second one with only the strains NCPPB 3335 and ICMP 1411 (indicated in Fig. 2 as sav II), as previously found by Moreno-Pérez *et al*. (2020).

Comparative analysis of levansucrase genes in savastanoi *pathovars*

For a comparative analysis, the sequences of the levansucrase genes *lscA*, *lscB* and *lscC* of *Pseudomonas savastanoi* pv. *glycinea* were investigated as described in 'Identification of the levansucrase genes' section. The *lscA* alignment resulted in two different hits corresponding to the first and second half of the levansucrase gene among all the pathovar *savastanoi* and sharing an average similarity above 99% with *Pseudomonas savastanoi* pv. *glycinea lscA* (Table S1). Notably, a third hit with 86% of similarity was obtained from *lscA* alignment towards PVFi1 and ICMP 13519 genomes, which corresponds to the prokka annotated *sacB_1* and

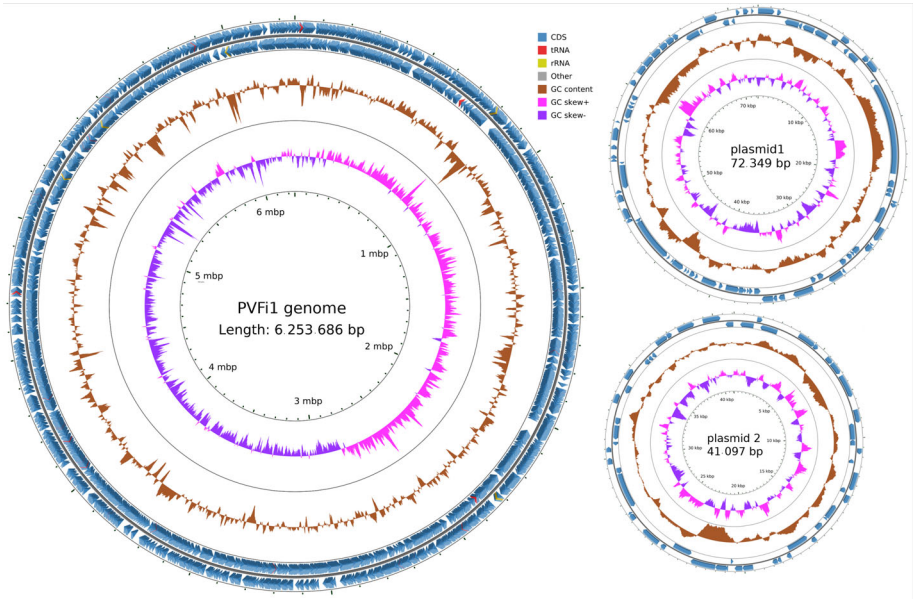

Fig. 1. A graphical circular map of the *de novo* assembled PVFi1 genome and related plasmids realized with CGview. From outside to centre, ring 1 and 2 show the CDS, tRNAs and rRNAs annotated by Prokka in forward and reverse strand, ring 3 and 4 the GC content percentage and GC skiew.

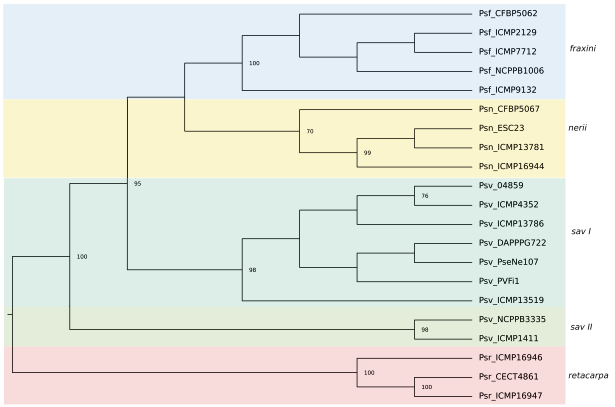

Fig. 2. Phylogenetic relationships among *Pseudomonas savastanoi* strains based on the alignment of 40 MLSA genes. Only the bootstraps value higher than 70 are shown. NCPPB 3335 and ICMP 1411 represent a distinct phylogenetic clade and thus are visualized with a different shade of green than the other *savastanoi* strains.

sacB_2, respectively. Instead, one single hit (and thus, one single copy) with 99% of similarity to either *Pseudomonas savastanoi* pv. *glycinea lscB/C* sequences were obtained from both PVFi1 and ICMP 13519. On the contrary, alignment of *lscB/C* to the other *savastanoi* strains showed a low similarity (84% and 87%), leading back to the same genomic region of *lscA* alignment, confirming that the levansucrase sequences they are bearing are indeed more related to this non-functional gene than the functional *lscB/C* (Table S1).

A deeper analysis of the PVFi1 and ICMP 13519 genomic structure confirmed a gene organization similar to the one of *lscB/C* in *glycinea*, which comprises all the following features: (i) the page-associated promoter PAGE including the transcription start site; (ii) the HexR repressor binding domain; (iii) the phagic *com* gene; (iv) the 48 bp region at the N-terminal extremity of *lsc/sacB_3* gene; and (v) 49–51 bp downstream the coding sequence possibly involved in the formation of a stem-loop structure for p-independent transcriptional termination (Fig. 3A). Furthermore, the region upstream the PAGE promoter appeared to have the EAL domain gene exactly as *glycinea lscB*, while the downstream region is similar to *glycinea* lscC for the presence of the glycoside idrolase 10 gene, also annotated as YddW protein. Three catalytic sites were also found within the coding sequence (Fig. 3A). The alignment of the regulatory *lscR* sequence of *glycinea* (accession number KX834132.1) and of the *HexR* sequence of *savastanoi* NCPPB 3335 (NZ_CP008742.1:173178–174044), necessary for levansucrase gene regulation, led to a sequence similarity of 97% and 100%, respectively (Table S1).

Thus, the overall structure surrounding the highly conserved coding sequence was reconstructed in a region of 5072 bp that, when blasted, resulted to be 99.90% similar to *Pseudomonas syringae* pv. *morsprunorum* strain CFBP 2116 (LT985192.1), *Pseudomonas amygdali* pv. *morsprunorum* strain R15244 (CP026558.1) and to *Pseudomonas syringae* pv. *morsprunorum* strain CFBP 3840 (LT963409.1) and only 99.45% and 98.36% similar to *lscB* and *lscC* of *glycinea,* respectively, which have been used for the initial alignment (Fig. 3B, Table S2).

In addition, by extending the alignment region upstream and downstream the levansucrase gene, all the genomes of the aforementioned *morsprunorum* strains, as well as the PVFi1 genome, appeared to bear the same transposase genes belonging either to the Tn3 and IS110 family.

Annotation comparison through MAUVE alignment

Pseudomonas savastanoi pv. *savastanoi* genomes and their relative annotations were reordered with Mauve towards the reference NCPPB 3335, as detailed in 'Comparative genomics analysis' section. However, the analysis highlighted several differences in features organization and nomenclature among the strains.

Some strains resulted to have two features annotated as *sacB_1* and *sacB_2* (i.e., DAPP PG722 and PseNe107), others showed only one feature as *sacB*, corresponding to a first fragment of the complete levansucrase gene (i.e., 04589, ICMP 13786, ICMP 1411 and ICMP 4352), while the second fragment was annotated as hypothetical protein. Instead, in the complete circular NCPPB 3335 strain, the two features *sacB_1* and *sacB_2* appeared interrupted by a transposase insertion (ISPa41), leading to a non-functional gene (Table S3 and Fig. S1).

As expected by the blast alignment in the 'Comparative analysis of levansucrase genes in *savastanoi* pathovars', PVFi1 strain showed three features related to levansucrase, with two of them (*sacB_2* and *sacB_3*) interspersed within a region enriched in transposase genes. Furthermore, ICMP 13519 showed a *sacB_2* feature that could be related to *sacB_1* of Psv PVFi1 (data not shown).

Levansucrase PCR validation

In order to verify the gene organization described above, standard PCR reactions were performed to amplify the *lscA/sacB_2–3* region and *lscBC/sacB_1* in PVFi1 DNA. The results confirmed the *in silico* reconstructed organization, with the *lscA* being interrupted by the transposase and giving a product of 889 bp, instead of the 365 bp expected in absence of the transposase as in *glycinea* or 1542 bp expected with a transposase insertion within the gene, as in NCPPB 3335. On the other hand, the *lscB/C* gene results to be placed downstream the *com* gene, as expected (Fig. S2).

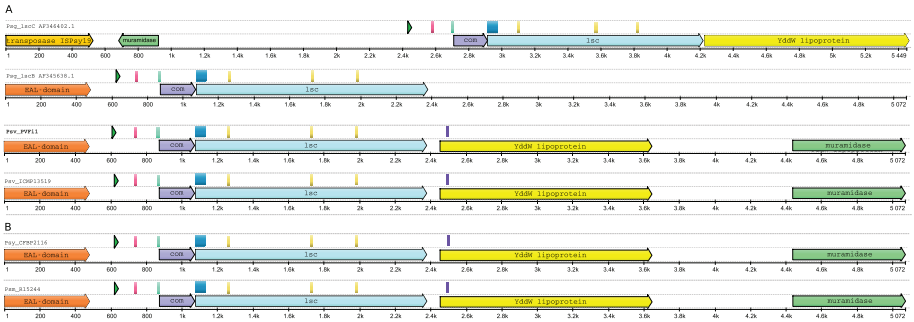

Fig. 3. Schematic representation of the genomic region carrying the levansucrase gene among *Pseudomonas* species.A. *LscB* and *lscC* gene in *Pseudomonas savastanoi* pv. *glycinea* PG4180, *Pseudomonas savastanoi* pv. *savastanoi* PVFi1 and *Pseudomonas savastanoi* pv. *savastanoi* ICMP 13519.B. *Pseuodomonas syringae* strain CFBP 2116 and *Pseudomonas amygdali* pv. *morsprunorum* strain R15244. Arrows represent coding sequences in the respective orientation, where the ORF has been annotated as 'hypothetical protein'. Coloured blocks represent common regions or regulatory elements: the green arrow indicates the position from where the lscB and lscC are aligned, pink block represent the translation start site TSS, cyano block represents the HexR binding sequence, blue block is the 48 N-terminal region and yellow blocks the catalytic sites, purple block represent the stem-loop for translational termination.

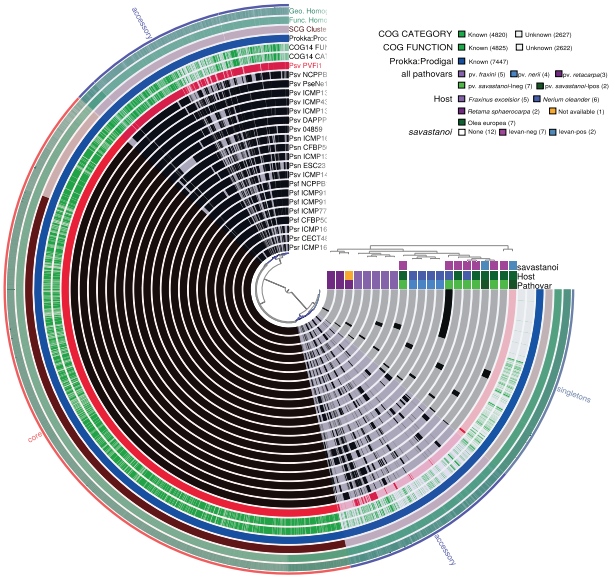

Fig. 4. Pangenome analysis of *Pseudomonas savastanoi* strains performed with anvi'o. From outside to the inside, the different layers represent the Geometrical and Functional Homogeneity, the single core genome (SCG) clusters, Prokka and COG functional and category annotation. The inner layers represent individual genomes organized regarding their phylogenetic relationships as indicated by the dendrogram, with the dark colour indicating the presence of the gene cluster and the light colour its absence. *Psv* PVFi1 is enlighten in red. The singletons, accessory and core gene clusters are indicated. Host, pathovar and levan groups used for the enrichment analysis are also indicated.

Pangenome analysis

The pangenome performed by Anvi'o resulted in a total of 7447 gene clusters with a total of 117 099 genes. Among the gene clusters, 3995 were defined as core genes with 3400 present as single copy per genome (SCG), while 2334 gene clusters were defined as accessory, and thus, present in 2–20 isolates (Fig. 4). Overall, 1115 gene clusters were present only in one isolate with 13 of them present only in Psv PVFi1, mostly with unknown function, besides one that was annotated as a sugar permease NicT. To further find out those functions that are peculiar to the defined levan-positive group and, thus, to PVFi1 strain, we used the *anvi-get-enriched-functions-per-pan-group* function within Anvi'o, with both COG functions and Prokka annotation (see supplementary material, Table S4). Interestingly, the assumed group of levan-positive *savastanoi* strains (PVFi1 and ICMP 13519) showed to be enriched in a cGAMP-activated phospholipase (capV), which in *Vibrio cholerae* confers resistance to bacteriophages. The same group, together with the *fraxini* and *retacarpa* isolates, resulted to be enriched in the lipoprotein YddW (COG category S), whose gene was mapped downstream the levan-coding gene (Fig. 3). Furthermore, transposase and related resolvase, phage-related proteins belonging to the mobilome COG category X in membrane transport proteins, antitoxin genes and catalytic enzymes related genes were also enriched (Table S4).

Secretion systems substrates

Gram-negative bacteria are known to release either into the surrounding environment or directly inside the host cells, specific proteins called substrates, through the different secretion systems. We assessed the presence of type II, III, IV and VI secretion systems by the alignment of the NCPPB 3335 annotated features from Rodriguez-Palenzuela *et al*. (2010) with both PVFi1 genome and annotation. A complete set of 28 genes representing the type II secreted system were identified, including the two distinct groups of *Sec* and *Tat* systems, already known among *Pseudomonas* spp. (Duong *et al*., 2001,

Table S5). The Type III secretion system was represented by 42 genes, including the peculiar gene cluster of 22 genes previously found in Psv NCPPB 3335, Pph 1448A, Pta 11528 and Por 1–6 (Rodriguez-Palenzuela *et al*., 2010). Surprisingly, among the 21 genes of NCPPB 3335 type IV secretion system, blastn was able to identify only one hit, the *TraA* gene. Thus, a deeper analysis was carried out using the SecreT4 database, which identified 22 genes related either to type IVA or IVB. Lastly, 36 genes were related to the type VI secretion system arranged into two expected main clusters (Table S5).

Type III secretion effectors

Recently, more than 14 000 putative T3SEs arranged in 70 different families have been catalogued in the *Pseudomonas syringae* complex (Dillon *et al*., 2019). Among those, a selection of 76 effectors belonging to the *Pseudomonas savastanoi* species, including each single gene variant to have a complete picture of the diversity, was blasted towards *Psv* PVFi1 genome, as well as to the other *savastanoi* strains used in this study. The results shown in Fig. 5 and detailed in Table S5 are in agreement with the clustering obtained in phylogenetic tree in Fig. 2. NCPPB 3335 and ICMP 1411 strains represent a stand-alone cluster, while PVFi1 results closely related to DAPP-PG722 and PseNe107 strains, as well as to the other *in silico* levan-positive ICMP 13589 strain. To be sure that nothing was left behind, the complete database with the 70 families identified by Dillon was blasted, and a further BastionHub prediction analysis was performed, but no more effectors than the *savastanoi*-related were identified in PVFi1 strain.

Other virulence-related genes

Pseudomonas species infecting woody plants and belonging to the phylogroup 1 and 3 are characterized by a genomic region defined as WHOP (Woody Host *Pseudomonas* spp.), as first identified by Rodriguez-Pelenzuela in NCPPB 3335 (2010) and further extended to 40 additional species by Caballo-Ponce *et al*. (2017). This WHOP region contains the antABC operon involved in anthranilate degradation into catechol, which is further degraded by the catABC operon product. To counteract the phenolic compounds produced by the plant as a mechanism of defence, this region carries also the genes necessary for their degradation (the ipoABC operon), which are regulated by the antR gene. Furthermore, in NCPPB 3335 the aerotaxis receptor (PSA3335_3206), the dhoAB operon and the benR regulator of the benABCD operon are associated with the WHOP region. Thus, the perfect alignment of 14 Kb WHOP region of NCPP 3335 downloaded from Genbank proved that the PVFi1 genome carries all the necessary tools to infect the vascular system of the woody plants. In addition, we found out that all the sequences annotated in the reference NCPPB 3335 genome as virulence-related like adhesins, transporters, motility, chemotaxis proteins, as well as genes involved in the production of the indoleacetic acid and quorum sensing genes, were detected in the PVFi1 strain (Table S5).

Discussion

The results of this study provided a complete genome sequence of one of the levan-positive *Pseudomonas savastanoi* pv. *savastanoi* strains that were unexpectedly identified during an epidemiological survey in Central Italy (Iacobellis *et al*., 1993). These strains have been previously characterized from a morphological, biochemical and molecular point of view (Marchi *et al*., 2005; Sisto *et al*., 2007), but without providing a complete picture of the genetic structure and the regulative elements involved in the levansucrase expression.

This work aimed to fill this gap in knowledge by providing the complete genome, arranged in a circular chromosome and two additional plasmids, obtained through a hybrid assembly of Illumina short reads and Oxford Nanopore Technology (ONT) long reads. Long reads, in fact, allows a better assembly performance of those regions of difficult exploration with only the short Illumina reads, such as those full of transposase or tandem repeats (Goodwin *et al*. 2016; van Dijk *et al*., 2018). On

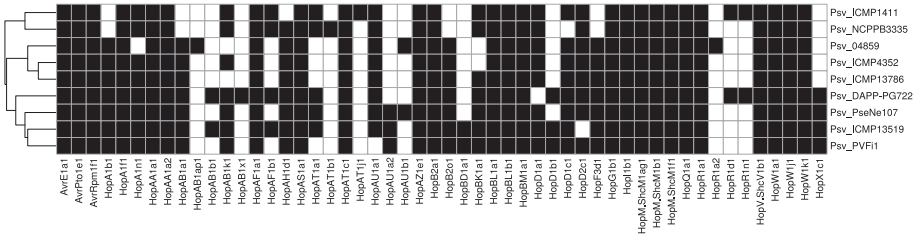

Fig. 5. Distribution of type III secretion effector genes and variants among the *Pseudomonas savastanoi* pv. *savastanoi* strains used in this study. The black block indicates the presence of the effector retrieved by blast alignment; the white block indicates its absence.

the other hand, the depth coverage and higher basecalling quality of the Illumina reads improves the solidity of the whole genetic reconstruction performed by the ONT reads (Chen *et al*., 2020).

The phylogenetic tree based on Multi Locus Sequence Typing (MLST), the core genome alignment and all the other features including secretion systems, effectors, phytohormones and quorum sensing, firmly place the levan-positive *savastanoi* strains within the *savastanoi* clades. The circular chromosome carries two copies of the levansucrase gene, one related to the *lscA* (*sacB_2* and *sacB_3*) and the other one to *lscB/C* (*sacB_1)* of *Pseudomonas savastanoi* pv. *glycinea* PG4180, while no levansucrase gene has been identified in the plasmids.

The first copy related to *lscA* has the Shine-Dalgarno sequence just before the starting codon and the N-terminal sequence WSRADAL, as observed in *Erwinia amylovora*, *Zymomona mobilis*, *Pseuodomonas savastanoi* pv. *glycinea* and *Pseudomonas syringae* pv. *phaseolicola* (Hettwer *et al*., 1998). The gene, expected to be 1248 bp, seems to be interrupted at position +728 from the translation start site, followed by a ISPa41 transposase gene insertion, as further confirmed by the PCR reactions. The same interruption has been found in the other *savastanoi* strains that are known to be levan negative and thus, suggesting that this fragmented gene is not functional.

The second copy of the levansucrase gene in PVFi1, as well as in ICMP 13519, is more related to the upstream region of *glycinea lscB* due to the presence of the EAL-domain, to the downstream region of *glycinea lscC* for the presence of the YddW lipoprotein, and shows a 99% of similarity with both of them in the 1296 bp coding sequence. In addition, following the lipoprotein sequence, both the strains carry a muramidase gene that *lscC* has 2 kb upstream the coding sequence, suggesting a similar genetic rearrangement.

In the pathovar *glycinea* several regulatory elements have been proved to be necessary for the gene expression and enzymatic activity Hettwer *et al*., 1998; Li and Ullrich, 2001; Srivastava *et al*., 2012; Khandekar *et al*., 2014; Mehmood *et al*., 2015; Abdallah *et al*., 2016), which have been also found in our PVFi1 strain: (i) the PAPE region with the *com* gene; (ii) the transcriptional start site; (iii) the N-terminal sequence of 48 bp; (iv) the hexose metabolism repressor *HexR*; (v) the prophage-borne transcriptional regulator *lscR*; (vi) the enzyme catalytic sites, defined as block I, block II and block III; and (vii) the stem-loop region for translation termination. This *in silico* reconstructed region of 5072 bp when blasted appeared to be more related to *Pseudomonas syringae* pv. *morsprunorum* than to *Pseudomonas savastanoi* pv. *glycinea*, suggesting a possible exchange of genetic material between the *morsprunorum* and the levan-positive *savastanoi* strains.

This hypothesis is further supported by the presence of sequences related to Tn3 and IS110 transposase gene families that surround the levansucrase genes in both species. Indeed, the transposase TnAS2 of the family Tn3 is found upstream the EAL domain of PVFil, while a ISPsy16 is present 8 Kb further downstream the muramidase gene (data not shown). On the other hand, as example *Pseudomonas syringae* pv. *morsprunorum* CFBP 2116 carries a transposase ISPsy16 just downstream of the muramidase gene.

These findings also sustain the hypothesis that an original pro-phage or transposase insertion may have contributed to the levansucrase gene spread among the *Pseudomonas* niche (Li and Ulrich, 2001; Srivastava *et al*., 2012). Indeed, horizontal gene transfer and transposable elements have already been linked to the exchange of avirulence factors and toxins among *Pseudomonas syringae* strains (Kim *et al*., 1998; Alarcón-Chaidez *et al*., 1999). Indeed, the high number of genes shared with the other pathovars used in this study, as well as the genes related to mobile element and genetic rearrangement revealed with the enrichment analysis, suggest that these strains are prone to exchange genetic material. Thus, further studies on the phylogeny of these transposases would better clarify the potential interaction among these species.

Nonetheless, given that to date only these Italian *P. savastanoi* strains have been reported as levan positive, it is worth to mention that *Psv* ICMP 13519, to only other carrying potentially functional levansucrase gene, was isolated by J. Young on 1997 in New Zealand from quarantine olive plants cv frantoio possibly of Italian origin, thus sustaining the geographic origin of these peculiar strains (Young *et al*., 2004; Iacobellis, personal communication).

In conclusion, these findings support, from a genomic point of view, the results of Iacobellis and Marchi (Iacobellis *et al*., 1993; Marchi *et al*., 2005), further suggesting that the LOPAT scheme for *Pseudomonas savastanoi* should be updated, since this bacterial species can be either levan positive or levan negative.

Experimental procedures

DNA isolation and genome sequencing

An aliquote of the lyophilized stock of the *Pseudomonas savastanoi* pv. *savastanoi* strain PVFi1, maintained in the bacterial collection of the Department of Agriculture and Forest Science of the University of Tuscia, was rehydrated and grown on King's B (KB) medium at constant

Table 1. List of *Pseudomonas savastanoi* strains involved in the comparative genomics and phylogenetic analysis.

Strain name	Pathovar	Host	Geographic origin	Accession number
Psf_CFBP5062	*fraxini*	*Fraxinus excelsior*	Netherlands	NZ_LIIC01000001.1
Psf_ICMP7712	*fraxini*	*Fraxinus excelsior*	Netherlands	NZ_RBSC01000183.1
Psf_ICMP9129	*fraxini*	*Fraxinus excelsior*	Netherlands	NZ_RBSA01000336.1
Psf_ICMP9132	*fraxini*	*Fraxinus excelsior*	Netherlands	NZ_RBSB01000011.1
Psf_NCPPB1006	*fraxini*	*Fraxinus excelsior*	United Kingdom	NZ_NIAW01000001.1
Psn_CFBP5067	*nerii*	*Nerium oleander*	Spain	NZ_LIHX01000001.1
Psn_ESC23	*nerii*	*Nerium oleander*	Italy	NZ_NIAY01000001.1
Psn_ICMP13781	*nerii*	*Nerium oleander*	Italy	NZ_RBTO01000205.1
Psn_ICMP16944	*nerii*	*Nerium oleander*	France	NZ_RBUB01000065.1
Psr_CECT4861	*retacarpa*	*Retama sphaerocarpa*	Spain	NZ_NBYW01000001.1
Psr_ICMP16946	*retacarpa*	*Retama sphaerocarpa*	Spain	NZ_RBQM01000278.1
Psr_ICMP16947	*retacarpa*	*Retama sphaerocarpa*	Not available	NZ_RBNM01000403.1
Psv_04859	*savastanoi*	*Nerium oleander*	Not available	NZ_RBNY01000535.1
Psv_DAPPPG722	*savastanoi*	*Olea europea*	Italy	NZ_JOJV01000001.1
Psv_ICMP13519	*savastanoi*	*Olea europea*	New Zealand	NZ_RBNW01000134.1
Psv_ICMP13786	*savastanoi*	*Nerium oleander*	Italy	NZ_RBTN01000340.1
Psv_ICMP1411	*savastanoi*	*Olea europea*	California	NZ_RBPF01000190.1
Psv_ICMP4352	*savastanoi*	*Olea europea*	Yugoslavia	NZ_LJRJ01000054.1
Psv_NCPPB3335	*savastanoi*	*Olea europea*	France	CP008742.1
Psv_PseNe107	*savastanoi*	*Olea europea*	Nepal	NZ_JYHF01000001.1
Psv_PVFi1	*savastanoi*	*Olea europea*	Italy	CP078139

Table 2. List of primers and their relative sequences involved in this study.

Primer name	Sequence 5′-3′	References
lscAF	AAGACTTTACGGAGGTGAAAACACT	This study
lscAR	GTCCTGAAAGACATAATGCGGGCGT	This study
transR	CACCATTATTCATGCCCCAAG	This study
comF	CCTCCCAAAGCTATTTAAAGG	This study
lscCf	ATCAGGGCGTGGAGTTGA	Marchi *et al*. (2005)
lscCr	GTCTGGAAAGGCTGCTCG	Marchi *et al*. (2005)

temperature of 26°C. High-molecular-weight (HMW) DNA was extracted according to Mayjonade *et al*. (2016). DNA integrity and DNA purity were evaluated in a 0.5% agarose gel electrophoresis run and with Thermo Scientific Multiskan GO (Thermo Fischer Scientific), respectively. The same DNA was quantified by an Invitrogen Qubit fluorimeter (Thermo Fisher Scientific, Massachusetts).

DNA sequencing approach was based on a double sequencing using both ONT, that was accomplished with a MinION Mk1C device (ONT, UK) using a R9.4.1 flowcell (ONT), after library preparation using the SQK-LSK109 kit (ONT), and Illumina technology with a NovaSeq 6000 S2 system, using paired-end sequencing at Eurofins Genomics (Eurofins Genomics GmbH, Konstanz, Germany).

Genome assembly and annotation

The quality of the paired-end Illumina reads was evaluated using FastQC (Andrews, 2010), the adapters of the NovaSeq 6000 were removed using Trimmomatic (Bolger *et al*., 2014) and the low-quality reads were removed by Sickle (Joshi and Fass, 2011). The basecalling of the ONT long reads was performed using Guppy within the MK1C device, while reads quality and sequencing statistics were evaluated with NanoPlot web-service v. 1.30.1 (De Coster *et al*., 2018).

The *de novo* hybrid assembly of both ONT and Illumina reads was performed using Unicycler v0.4.9b (Wick *et al*., 2017) with default parameters, except the minimum fasta length that was set to 5.000. Quality assessment of genome assembly was carried out using QUAST 5.0 (Gurevich *et al*., 2013) and the genomes were graphically visualized using CGview (Petkau *et al*., 2010). Annotation was performed with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP, Tatusova *et al*., 2016) and with Prokka for an unbiased comparative genomics analysis (Seemann, 2014).

Phylogenetic analysis

Phylogenetic and comparative genomics analysis were conducted retrieving from NCBI Genomes the complete or draft sequence of eight strains of *Pseudomonas*

savastanoi pv. *savastanoi* and 12 *Pseudomonas savastanoi* pathovars (*fraxini*, *nerii* and *retacarpa*). The 20 genomes downloaded from the NCBI database and the genome sequenced in this work are listed in Table 1, together with additional information.

To evaluate the phylogenetic relationship among the strains, a MLSA was applied on a set of 40 concatenated housekeeping genes that were aligned with MAFFT using default parameters (Katoh *et al*., 2018), according to the method used by Moreno-Pérez *et al*. (2020). The tree was obtained using the neighbour-joining method, Jukes–Cantor as substitution model, 100 bootstraps replicates and visualized in a dendrogram using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree).

Identification of the levansucrase genes

To identify and characterize the levansucrase gene, the sequences of *lscA*, *lscB* and *lscC* genes of *Pseudomonas savastanoi* pv. *glycinea* PG4180 (accession numbers AF037443.1, AF345638.1 and AF346402.1) were aligned to the genomes under investigation using NCBI blastn. The alignment was further visualized with Ugene (Altschul *et al*., 1990; Okonechnikov *et al*., 2012).

The validation of the genomic region containing *lscA* and *lscB*/*C* sequences provided by the *in silico* analysis, was performed by a series of standard PCR reactions using both the *lscCf*/*lscCr* primer pair from Marchi *et al*. (2005) and three additional pairs specifically designed in this study (Table 2): *lscAF*/*lscAR*, *lscAF*/*transR* and *comF* (to be used with *lscCr*).

PCR reactions were prepared as follows: 2 μl of DNA was incubated with 0.5 μM of forward primer and 0.5 μM of reverse primer in 1x master mix buffer, in a final volume of 20 μl. The amplifications were performed with a C1000 thermocycler (Biorad Laboratories, CA., USA) with the following program: initial denaturation for 5 min at 95°C, 35 cycles of denaturation for 30 s at 95°C, annealing for 30 s at 62°C, and extension for 60 s at 72°C, plus a final elongation step for 5 min at 72°C. The PCR products were separated by Midori-stained agarose gel electrophoresis (1%) in TAE buffer (TRIS-Acetate-EDTA, pH 8).

Comparative genomics analysis

All the 21 *Pseudomonas* genomes listed in Table 1 were first annotated with Prokka v1.14.5 (Seemann, 2014) for an unbiased comparison. The genomes and related annotations were then reordered with MAUVE using *savastanoi* NCPPB 3335 as reference since, at the time of this research, it was the only one complete assembly. Progressive MAUVE algorithm with default parameters was subsequently used for the alignment (Darling *et al*., 2004).

The annotated features of NCPPB 3335 strain, including secretion systems, virulence-related features, phytohormones and WHOP genes (Rodriguez-Palenzuela *et al*. 2010; Caballo-Ponce *et al*., 2017), were aligned towards the nine *savastanoi* strains using NCBI blastn, as well as effectors belonging to the *Pseudomonas savastanoi* species (Dillon *et al*., 2019).

SecreT4 database was used to identify the type IV secretion system related genes (Bi *et al*., 2013), while BastionHub was used to predict any further type III effectors (Wang *et al*., 2021).

Pangenome analysis

Prokka annotation of the 21 *Pseudomonas* genomes was used as input for the pangenome analysis performed within Anvi'o v.7 docker container (Eren *et al*., 2015).

Genomes were also annotated within Anvi'o using NCBI's Clusters of Orthologous Groups (COG) database for a functional categories comparison (Tatusov *et al*., 2000). The pangenome was created setting the following parameters: (i) NCBI blastp for amino acid sequence similarity search, (ii) the default minbit euristic set to 0.5 and (iii) the MCL inflation parameter set to 10, as recommended for the comparison of strains belonging to the same species (https://merenlab.org/2016/11/08/pangenomics-v2/)

Acknowledgements

The research was carried out in the frame of the Italian MIUR (Ministry for education, University and Research) initiative 'Department of Excellence' (Law 232/2016). All the bioinformatics calculations and analyses were performed at DAFNE HPC scientific computing centre of the Università degli Studi della Tuscia.

Data availability statement

The complete genome and the related plasmids have been deposited on the NCBI database under the accession numbers CP078139, CP078140 and CP078141, respectively. The fastq data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Abdallah, K., Hartman, K., Pletzer, D., Zhurina, D., and Ullrich, M.S. (2016) The bacteriophage-derived transcriptional regulator, LscR, activates the expression of levansucrase genes in *Pseudomonas syringae*. *Mol Microbiol* **102**: 1062–1074.

Alarcón-Chaidez, F.J., Peñaloza-Vázquez, A., Ullrich, M., and Bender, C.L. (1999) Characterization of plasmids encoding the phytotoxin coronatine in *Pseudomonas syringae*. *Plasmid* **42**: 210–220.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. URL https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bi, D., Liu, L., Tai, C., Deng, Z., Rajakumar, K., and Ou, H.Y. (2013) SecReT4: a web-based bacterial type IV secretion system resource. *Nucleic Acids Res* **41**: D660–D665.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Bradbury, J.F. (1986) *Guide to Plant Pathogenic Bacteria*. Slough, UK: CAB International Mycological Institute.

Caballo-Ponce, E., Murillo, J., Martínez-Gil, M., Moreno-Pérez, A., Pintado, A., and Ramos, C. (2017) Knots untie: molecular determinants involved in knot formation induced by *Pseudomonas savastanoi* in woody hosts. *Front Plant Sci* **8**: 1089.

Chen, Z., Erickson, D.L., and Meng, J. (2020) Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing. *BMC Genomics* **21**: 631.

Darling, A.C.E., Mau, B., Blattner, F.R., and Perna, N.T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**: 1394–1403.

De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., and Van Broeckhoven, C. (2018) NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**: 2666–2669.

De Los Rios, J.E.G. (1999) *Retama sphaerocarpa* (L.) Boiss., a new host of *Pseudomonas savastanoi*. *Phytopathol Mediterr* **38**: 54–60.

Dillon, M.M., Almeida, R.N.D., Laflamme, B., Martel, A., Weir, B.S., Desveaux, D., and Guttman, D.S. (2019) Molecular evolution of *Pseudomonas syringae* type III secreted effector proteins. *Front Plant Sci* **10**: 418.

Duong, F., Bonnet, E., Géli, V., Lazdunski, A., Murgier, M., and Filloux, A. (2001) The AprX protein of *Pseudomonas aeruginosa*: a new substrate for the Apr type I secretion system. *Gene* **262**: 147–153.

Dworkin, M. (2006) The prokaryotes. In *Proteobacteria: Gamma Subclass*, 3rd ed, Vol. **6**. New York, NY: Springer-Verlag.

Eren, A.M., Esen, O.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., and Delmont, T.O. (2015) Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**: 1–29.

Gardan, L., Bollet, C., Ghorrah, M.A., Grimont, F., and Grimont, P.A.D. (1992) DNA relatedness among the pathovar strains of *pseudomonas syringae* subsp. *savastanoi* Janse (1982) and proposal of *pseudomonas savastanoi* sp. nov. *Int J Syst Bacteriol* **42**: 606–612.

Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**: 333–351.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**: 1072–1075.

Hettwer, U., Jaeckel, F.R., Boch, J., Meyer, M., Rudolph, K., and Ullrich, M.S. (1998) Cloning, nucleotide sequence, and expression in *Escherichia coli* of Levansucrase genes from the plant pathogens *Pseudomonas savastanoi* pv. *glycinea* and *P. syringae* pv. *phaseolicola*. *Appl Environ Microbiol* **64**: 3180–3187.

Iacobellis, N.S., Caponero, A., and Evidente, A. (1998) Characterization of *Pseudomonas syringae* ssp. *savastanoi* strains isolated from ash. *Plant Pathology* **47**: 73–83.

Iacobellis, N.S., Sisto, A., and Surico, G. (1993) Occurrence of unusual strains of *Pseudomonas syringae* subsp. *savastanoi* on olive in Central Italy. *EPPO Bulletin* **23**: 429–435.

Janse, J.D. (1982) *Pseudomonas syringae* subsp. *savastanoi* (ex Smith) subsp. nov., nom. rev., the bacterium causing excrescences on *Oleaceae* and *Nerium oleander* L. *Int J Syst Evol Mycrobiol* **32**: 166–169.

Joshi, N. & Fass, J. (2011). Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. URL https://github.com/najoshi/sickle

Katoh, K., Rozewicki, J., and Yamada, K.D. (2018) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* **20**: 1160–1166.

Khandekar, S., Srivastava, A., Pletzer, D., Stahl, A., and Ullrich, M.S. (2014) The conserved upstream region of lscB/C determines expression of different levansucrase genes in plant pathogen Pseudomonas syringae. *BMC Microbiol* **14**: 79.

Kim, J.F., Charkowski, A.O., Alfano, J.R., Collmer, A., and Beer, S.V. (1998) Sequences related to transposable elements and bacteriophages flank avirulence genes of *Pseudomonas syringae*. *Mol Plant Microb Interact* **11**: 1247–1252.

Laue, H., Schenk, A., Li, H., Lambertsen, L., Neu, T.R., Molin, S., and Ullrich, M.S. (2006) Contribution of alginate and levan production to biofilm formation by *Pseudomonas syringae*. *Microbiology* **152**: 2909–2918.

Lelliott, R.A., Billing, E., and Hayward, A.C. (1966) A determinative scheme for the fluorescent plant pathogenic *Pseudomonads*. *J Appl Bacteriol* **29**: 470–489.

Li, H., and Ullrich, M.S. (2001) Characterization and mutational analysis of three allelic *lsc* genes encoding levansucrase in *Pseudomonas syringae*. *J Bacteriol* **183**: 3282–3292.

Luti, S., Campigli, S., Ranaldi, F., Paoli, P., Pazzagli, L., and Marchi, G. (2021) Lscβ and lscγ, two novel levansucrases of *Pseudomonas syringae* pv. *actinidiae* biovar 3, the causal agent of bacterial canker of kiwifruit, show different enzymatic properties. *Int J Biol Macromol* **179**: 279–291.

Mann, E.E., and Wozniak, D.J. (2012) Pseudomonasbiofilm matrix composition and niche biology. *FEMS Microbiol Rev* **6**: 893–916.

Marchi, G., Cinelli, T., and Surico, G. (2011) A review on *Pseudomonas savastanoi* genetic traits involved in disease development and in symptom induction. In *Olive Disease and Disorders*: Kerala, India: Transworld Research Network.

Marchi, G., Viti, C., Giovannetti, L., and Surico, G. (2005) Spread of levan-positive populations of *Pseudomonas savastanoi* pv. *savastanoi*, the causal agent of olive knot, in Central Italy. *Eur J Plant Pathol* **112**: 101–112.

Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., *et al*. (2016) Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *Biotechniques* **61**: 203–205.

Mehmood, A., Abdallah, K., Khandekar, S., Zhurina, D., Srivastava, A., Al-Karablieh, N., *et al*. (2015) Expression of extra-cellular levansucrase in *Pseudomonas syringae* is controlled by the in planta fitness-promoting metabolic repressor HexR. *BMC Microbiol* **15**: 48.

Moreno-Pérez, A., Pintado, A., Murillo, J., Caballo-Ponce, E., Tegli, S., Moretti, C., *et al*. (2020) Host range determinants of *Pseudomonas savastanoi* pathovars of woody hosts revealed by comparative genomics and cross-pathogenicity tests. *Front Plant Sci* **11**: 1–19.

Moretti, C., Vinatzer, B.A., Onofri, A., Valentini, F., and Buonaurio, R. (2017) Genetic and phenotypic diversity of Mediterranean populations of the olive knot pathogen, Pseudomonas savastanoi pv. savastanoi. *Plant Pathol* **66**: 595–605.

O'Brien, H.E., Desveaux, D., and Guttman, D.S. (2011) Next-generation genomics of *Pseudomonas syringae*. *Curr Opin Microbiol* **14**: 24–30.

Okonechnikov, K., Golosova, O., Fursov, M., Varlamov, A., Vaskin, Y., Efremov, I., *et al*. (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**: 1166–1167.

Osman, W.A., Tarabeih, A.M., and Michail, S.H. (1980) Studies on the distribution of olive knot disease induced by Pseudomonas savastanoi in Iraq. *Mesopotamia J Agric* **15**: 245–261.

Petkau, A., Stuart-Edwards, M., Stothard, P., and Van Domselaar, G. (2010) Genome analysis interactive microbial genome visualization with GView. *Bioinformatics* **26**: 3125–3126.

Quesada, J.M., Penyalver, R., Pérez-Panadés, J., Salcedo, C.I., Carbonell, E.A., and López, M.M. (2010) Dissemination of *pseudomonas savastanoi* pv. *savastanoi* populations and subsequent appearance of olive knot disease. *Plant Pathol* **59**: 262–269.

Rahi, Y.J., Turco, S., Taratufolo, M.C., Tatì, M., Cerboneschi, M., Tegli, S., *et al*. (2020) Genetic diversity and population structure of *Pseudomonas savastanoi*, an endemic pathogen of the Mediterranean area, revealed up to strain level by the MLVA assay. *J Plant Pathol* **102**: 1051–1064.

Rodríguez-Palenzuela, P., Matas, I.M., Murillo, J., López-Solanilla, E., Bardaji, L., Pérez-Martínez, I., *et al*. (2010) Annotation and overview of the *Pseudomonas savastanoi* pv. *savastanoi* NCPPB 3335 draft genome reveals the virulence gene complement of a tumour-inducing pathogen of woody hosts. *Environ Microbiol* **12**: 1604–1620.

Schroth, M.N. (1973) Quantitative assessment of the effect of the olive knot disease on olive yield and quality. *Phytopathology* **63**: 1064.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069.

Sisto, A., Cipriani, M.G., Tegli, S., Cerboneschi, M., Stea, G., and Santilli, E. (2007) Genetic characterization by fluorescent AFLP of *pseudomonas savastanoi* pv. *Savastanoi* strains isolated from different host species. *Plant Pathol* **56**: 366–372.

Srivastava, A., Al-Karablieh, N., Khandekar, S., Sharmin, A., Weingart, H., and Ullrich, M.S. (2012) Genomic distribution and divergence of Levansucrase-coding genes in *Pseudomonas syringae*. *Genes* **3**: 115–137.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**: 33–36.

Tatusova, T., Dicuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., *et al*. (2016) NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* **44**: 6614–6624.

van Dijk, E.L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018) The third revolution in sequencing technology. *Trends Genet* **34**: 666–681.

Völksch, B., and Weingart, H. (1997) Comparison of ethylene-producing *Pseudomonas syringae* strains isolated from kudzu (*Pueraria lobata*) with *Pseudomonas syringae* pv. *phaseolicola* and *Pseudomonas syringae* pv. *glycinea*. *Eur J Plant Pathol* **103**: 795–802.

Wang, J., Li, J., Hou, Y., Dai, W., Xie, R., Marquez-Lago, T. T., *et al*. (2021) BastionHub: a universal platform for integrating and analyzing substrates secreted by gram-negative bacteria. *Nucleic Acids Res* **49**: 651–659.

Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017) Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **13**: e1005595.

Young, J.M., Dye, D.W., Bradbury, J.F., Panagopoulos, C. G., and Robbs, C.F. (1978) A proposed nomenclature and classification for plant pathogenic bacteria. *N Z J Agric Res* **21**: 153–177.

Young, J.M., Wilkie, J.P., Fletcher, M.J., Park, D.C., Pennycook, S.R., Triggs, C.M., *et al*. (2004) Relative tolerance of nine olive cultivars to *Pseudomonas savastanoi* causing bacterial knot disease. *Phytopathol Mediterr* **43**: 395–402.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Mauve alignment of the levansucrase genomic region of the nine *Pseudomonas savastanoi* pv. *savastanoi* used in this study. Prokka annotation of the levansucrase genes and surrounding ISPa41 transposase and pdtaR are indicated. The red vertical bars represent the contigs interruptions and the coloured green blocks represent genetic similarity.

Fig. S2. Levansucrase gene validation through Polymerase Chain Reaction (PCR). A) Schematic representation of the PCR primers designed on the *lscA* and *lscB*/*C* genes. Without any transposase insertion, the primers on the *lscA* gene (lscAF-lscAR) would led to a PCR product of 365 bp. On the contrary, if *lscA* is interrupted by the transposase as in *Psv* NCPPB 3335, the same primers would led to a PCR product of 1,542 bp. Instead, lscAF in combination with a primer

designed on the transposase itself, would produce a product of 889 bp. The presence of lscB can be detected using lscCF-lscCR designed by Marchi *et al*. 2005, while the com gene can be detected with primers comF-lscCR, leading to products of 549 and 1,128, respectively.

B) Gel electrophoresis showing the levansucrose genes arrangements in *Psv* PVFi. Lane M: 100 bp +3Kb Smobio ladder (DM2300); Lane 1: lscAF in combination with lscAR; Lane 2: lscAF in combination with transR. Lane 3: Marchi's primers lscCF-lscCR on the levansucrase gene. Lane 4: comF in combination with a reverse primer on the lscB/C gene. Lane 5: No template control.

Table S1. Blastn results of the levansucrase genes lscA, lscB and lscC alignment on the Pseudomonas savastanoi strains under comparison.

Table S2. Blastn highest similarities of the reconstructed levansucrase region of PVFi1.

Table S3. Comparison of the of the different fragments of the levansucrase region among the different Pseudomonas savastanoi strains.

Table S4. Enrichment analysis performed with Anvi'o.

Table S5. Presence/Absence matrix of the virulence-related features blasted on the Pseudomonas savastanoi strains under comparison.