

Genome analysis of *Pseudomonas syringae* pv. *lachrymans* strain 814/98 indicates diversity within the pathovar

Renata Słomnicka · Helena Olczak-Woltman · Michał Oskiera ·
Małgorzata Schollenberger · Katarzyna Niemirowicz-Szczytt ·
Grzegorz Bartoszewski 

Accepted: 10 December 2017 / Published online: 20 December 2017
© The Author(s) 2017. This article is an open access publication

Abstract Although many *Pseudomonas syringae* strains have already been determined, only a few genomes of strains belonging to pathovar *lachrymans* have been sequenced so far. In this study we report the genome sequence of *P. syringae* pv. *lachrymans* strain 814/98, which is highly virulent to cucumber. The genome size was estimated to be 6.58 Mb, with 57.97% GC content. In total, 6024 genes encoding proteins and 92 genes encoding RNAs were identified in this genome. Comparisons with the available sequenced genomes of pathovar *lachrymans* as well as with other *P. syringae* pathovars were conducted, revealing the presence of three unique plasmids and 24 type III effector proteins (TTEs) in strain 814/98. The phylogenetic

analyses of MLST loci and TTEs clearly showed the existence of two distinct clusters of strains within pathovar *lachrymans*, which were grouped into either phylogroup 1 or 3, supporting non-monophyly within this pathovar.

Keywords Angular leaf spot · Next-generation sequencing · Plasmids · Virulence effectors

Introduction

Since the application of Next Generation Sequencing (NGS) in microbiology, thousands of bacterial genomes have been sequenced. Among these analyzed species is *Pseudomonas syringae*, and at present ca. 180 genomic sequences have been assembled for *P. syringae* (NCBI 2017), the plant pathogenic bacteria species that cause diseases in many agriculturally important crops and which has been divided into different pathovars.

One of the pathovars of *P. syringae*, namely *lachrymans*, is mainly a pathogen of the cucumber (*Cucumis sativus* L.), to which it causes serious damage and yield loss due to the presence of water-soaked lesions on the leaves that later become necrotic, thus reducing the photosynthetic capacity of the infected foliage (Olczak-Woltman et al. 2008; Lamichhane et al. 2015). The disease caused by this pathogen, i.e. bacterial angular leaf spot, is distributed worldwide and appears on other cucurbit species as well. Novel haplotypes of *P. syringae* were observed to be common on multiple cucurbit hosts, thus illustrating this species' large ecological diversity

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10658-017-1401-8>) contains supplementary material, which is available to authorized users.

R. Słomnicka · H. Olczak-Woltman ·
K. Niemirowicz-Szczytt · G. Bartoszewski (✉)
Department of Plant Genetics Breeding and Biotechnology,
Faculty of Horticulture Biotechnology and Landscape
Architecture, Warsaw University of Life Sciences,
Nowoursynowska 159, 02-787 Warsaw, Poland
e-mail: grzegorz_bartoszewski@sggw.pl

M. Oskiera
Microbiology Laboratory, Research Institute of Horticulture,
Konstytucji 3 Maja 1/3, 96-100 Skierniewice, Poland

M. Schollenberger
Department of Plant Pathology, Faculty of Horticulture
Biotechnology and Landscape Architecture, Warsaw University of
Life Sciences, Nowoursynowska 159, 02-787 Warsaw, Poland

(Newberry et al. 2016). Moreover, pathovar *lachrymans* is particularly detrimental because it can facilitate infection by *Pseudoperonospora cubensis*, which is the most destructive cucumber pathogen that causes downy mildew (Olczak-Woltman et al. 2008). Recently, outbreaks of angular leaf spot were reported in several Chinese provinces, where the disease affected 15–50% of growing fields, causing between 30% and 50% of yield reduction (Meng et al. 2016). This is an economically important pathogen, as cucumber is grown over an area of 2.1 million hectares with total production at 71.3 million tons, mainly in China but also in the EU and USA (FAO 2017).

Current genomic technologies provide the means not only for efficient genome sequencing but also for comparative genome analyses from which structural, phylogenetic or evolutionary conclusions can be drawn. Genome sequencing is expected to provide relevant tools in bacterial taxonomy and in an in-depth characterization of bacterial pathogens. To date, the genomes of seven strains belonging to pathovar *lachrymans* have been sequenced and are available as drafts or early drafts (Baltrus et al. 2011; Jeong et al. 2015; Mott et al. 2016; NCBI 2017). However, among the seven genomes, only two strains, MAFF301315 and MAFF302278, were described in detail and aligned with other representative strains of *P. syringae* by Baltrus et al. (2011). These strains, unlike other *P. syringae* strains, have only a low percentage of novel Type Three Effectors (TTEs), although MAFF301315 possessed a relatively higher number of TTEs. Moreover, MAFF301315 possessed a megaplasmid pMPPla107, approximately 1 Mb in size encoding 776 hypothetical proteins. This megaplasmid was found to be present in the very closely related strain N7512 but was absent in other *lachrymans* strains. It was inferred that the plasmid was a recent acquisition since the collected pathovar *lachrymans* strains possess nearly identical sequences at their MLST loci and only these two strains possess the megaplasmid (Baltrus et al. 2011). It was later shown that pMPPla107 is self-transmissible across a variety of diverse pseudomonad strains with conjugation dependent on a Type Four Secretion System (T4SS). However, its role in virulence remains elusive (Romanchuk et al. 2014). Recently, Baltrus et al. (2017) described four different Type Three Secretion Systems (T3SS) in *P. syringae* pathovars: canonical, rhizobial, single and atypical. The canonical system present in *P. syringae* pv. *tomato* DC3000 is

required for virulence *in planta* in every pathogenic strain investigated so far, and its presence is strongly correlated with pathogenic potential on agriculturally relevant plants. On the other hand, the rhizobial system does not appear to be required for virulence *in planta* but was acquired via multiple horizontal gene transfers by strains within the *P. syringae* complex (Baltrus et al. 2017).

In this paper we report the genome sequence of pathovar *lachrymans* strain 814/98 which is highly virulent to cucumber. This sequence was compared with other *P. syringae* strains with a special focus on the strains of pathovar *lachrymans*.

Material and methods

The bacterial strain Both virulence and genetic diversity of the strains collected at the Department of Plant Genetics, Breeding and Biotechnology of WULS were described previously (Olczak-Woltman et al. 2007; Słomnicka et al. 2015). Based on those studies, pathovar *lachrymans* strain 814/98, recognized as the most virulent strain to cucumber, was chosen for genome sequencing. The strain is of Dutch origin and was obtained from a collection maintained in the Pathogen Bank of the Institute of Plant Protection, Poland.

Bacterial growth and DNA isolation methods The bacterial culture of strain 814/98 was initiated from a single colony and grown for 24 h in Luria Broth liquid medium on a rotary shaker at 28 °C and 200 rpm. Total genomic DNA was extracted using the DNA Genomic-tips 100/G kit (Qiagen, Germany), as per the manufacturer's instructions. The DNA concentration was estimated by using a NanoDrop2000 spectrophotometer (Thermo Scientific, USA) and by electrophoresis on agarose gel stained with ethidium bromide. Finally, quality of the sample was verified by chip electrophoresis using the Experion™ Automated Electrophoresis System (Bio-Rad, USA); ca. 70 µg DNA of high purity was provided for sequencing.

Whole genome sequencing An Illumina HiSeq 2000 platform was used for sequencing. Briefly, two types of DNA paired-end libraries with an insert size of 500 bp and 6500 bp were constructed according to manufacturer's recommendations (Illumina, USA) to generate >100× genome coverage (Table S1). The DNA was sonicated, end-

repaired and ‘A’ was added at 3’ ends using T4 polynucleotide kinase. Further adapters were ligated, size-selected DNA was enriched by PCR and used for library preparation. Later, commercial sequencing was performed at BGI Tech Solutions (Hong Kong, China).

De novo genome assembly and structural annotation Raw Illumina reads were filtered to remove adapters and low quality bases. Clean data as FastQ files were assembled using SOAPdenovo (Li et al. 2008) into contigs and scaffolds (assembly P814h, NCBI GenBank Accession NBLF00000000, BioProject PRJNA380232, raw sequence read archive SRA SUB2542424). These were structurally analyzed and the number and length of the contigs and scaffolds (Table S2) as well as repetitive fragments (Table S3) were described. Tandem repeats were identified using a Tandem Repeat Finder (Benson 1999). Minisatellite and microsatellite DNA were classified based on the number and length of repeat units (15–65 bp for minisatellite DNA and 2–10 bp for microsatellite DNA). Sequences flanking microsatellite loci (100 bp up- and downstream) were compared using the BLASTX algorithm to sequences deposited at: NCBI (<http://www.ncbi.nlm.nih.gov/>), JGI (<http://jgi.doe.gov/>) and *Pseudomonas* Genome DB (<http://www.pseudomonas.com/>). The BLASTX hits were identified (E-value < $1e^{-5}$ with similarities of at least 85%) and the summary results are presented in Table S3.

Gene prediction and functional annotation Genes encoding proteins were predicted from the genome assembly using Glimmer v.3.02 (Delcher et al. 2007). Genes encoding rRNA and tRNA were identified using RNAmmer v.1.2 (Lagesen et al. 2007) and tRNAscan-SE v.1.3.1 (Lowe and Eddy 1997). sRNA genes were predicted using the Rfam database (<http://rfam.xfam.org/>). Functional gene annotation was done by analyzing protein sequences (Table S4). Genes were aligned with several databases to obtain their corresponding annotations. The following were searched: KEGG v.59 (Kanehisa et al. 2006) and COG v.20090331 (Tatusov et al. 2003); Swiss-Prot v.2011_10_19 and GO v.1.419 (Bard and Winter 2000). To ensure biological meaning, high-quality alignment results were chosen for gene annotation.

Phylogenetic and comparative analyses A dataset of 26 sequenced *Pseudomonas* spp. strains was constructed in order to perform comparative analysis (Table 1). It

consisted of strain 814/98, other previously sequenced strains of pathovar *lachrymans* and published in NCBI, sequenced strains belonging to other *P. syringae* pathovars, and one strain each of: *P. aeruginosa*, *P. cichorii*, *P. fluorescens* and *P. putida*. The constructed dataset of sequenced *Pseudomonas* spp. strains was enriched with 15 *P. syringae* strains derived from our collection in order to perform MLST analysis. Of the analyzed MLST loci: *cts*, *gapA*, *pgi*, *rpoD*, *gyrB*, *pfk* and *acn* (Sankar and Gutman, 2004; Hwang et al. 2005), three sequences, i.e. *acn*, *gyrB* and *pgi* were chosen for phylogenetic analysis because the set of sequences without unknown nucleotides for all of the analyzed strains was found only for these genes. The corresponding nucleotide sequences were extracted for each of the genomes (Table S5a–c). Genome sequences were examined for the full length of the three gene sequences without the unknown nucleotides. Furthermore, sequences set of concatenated partial genes sequences were processed with BLASTclust (<http://toolkit.tuebingen.mpg.de/blastclust>) with clustering level of 100 % sequence identity, then the obtained sequences were assembled with ClustalX and manually trimmed in CLC Genomic Workbench v.9.0 (CLC Bio, Denmark). Final block alignment was prepared using GBlocksServer with a less stringent selection option (http://molevol.cmima.csic.es/castresana/Gblocks_server.html). In the final alignment, concatenated sequences showing no differentiation were removed, except for sequences that belonged to strains of different species, such as *P. fluorescens* ICMP7711, *P. savastanoi* pv. *neri* ICMP16943 and *P. savastanoi* pv. *savastanoi* NCPPB3335. Nucleotide sequences of *P. aeruginosa*, *P. cichorii*, *P. fluorescens* and *P. putida* were used as the out-group. Final sequence alignment for 40 selected strains consisted of 1672 nucleotides of concatenated partial genes sequences and partial genes alignments consisted of 521, 672, 475 nucleotides for *acn*, *gyrB*, and *pgi*, respectively. The substitution model, nucleotide frequencies and substitution values were estimated with the jModel Test v.0.1.1 (Darriba et al. 2012) and with the AICc selection criterion (model GTR + I + G). Metropolis-coupled Markov chain Monte Carlo (MCMCMC) analysis was performed with two runs for 5 million generations with four chains, with the heating coefficient $\lambda = 0.1$ with MrBayes v.3.2.2 × 64 (Ronquist et al. 2012).

Plasmid identification Plasmid identification was performed using several different methods. Bioinformatic identification using CLC Genomic Workbench v.9.0

was done after mapping the raw clean reads on already assembled contigs and scaffolds. The calculated coverage was the basis for plasmid identification because short and multicopy plasmid sequences have higher coverage than chromosome sequences (CLC Genomics Manual) (Table S2). The other method was to search for the *ori* sequence. In order to perform this approach, proper sequences identified in other strains were downloaded from NCBI and *Pseudomonas* DB databases (Table S6). Finally, the scaffolds and contigs of 814/98 were surveyed with BLASTN for the presence of other structural genes or unique sequences associated with *P. syringae* plasmids. A whole scaffold or contig with a BLAST hit (E-value = 0.0; similarity length at least 1 kb) identifying the fragment as coming from a plasmid was compared to the NCBI database, and plasmid sequences available in the NCBI database were compared to the 814/98 scaffolds and contigs (BLAST and reciprocal-BLAST). The BLAST results are summarized in Table S6. BLAST results for gene sequences present on the plasmids and their descriptions are presented in Table S7. Plasmid identification on Eckhardt gels was performed previously (Słomnicka et al. 2015).

Identification of TTEs A constructed dataset of 26 sequenced *Pseudomonas* spp. strains was used to identify the presence of known TTEs (Table 1). Additionally, strain B728a of *P. syringae* pv. *syringae* was added. The genome sequence of each strain was surveyed with the TBLASTN algorithm in CLC Genomic Workbench v.9.0. The set of known effectors was constructed based on an extended table of the Baltrus et al. (2011) concept and on validated TTE family members deposited on the PPI website (<http://www.pseudomonas-syringae.org/>). Each strain was considered to possess TTEs if a majority of the protein sequences had significant BLAST hits ($<10^{-5}$) with an identity of at least 80%. The binary matrix of either the presence or absence of TTEs for the *Pseudomonas* spp. collection was created (Table S8). Finally, this matrix was converted into a genetic distance matrix in NTSYSpc 2.1 software (Exeter Software, USA) using the Jaccard coefficient, and the dendrogram was constructed in Mega 7.0 software (Center for Evolutionary Medicine and Informatics, USA) using the NJ clustering method. Additionally, protein sequences were evaluated using Effective T3 database v.1.0.1 (<http://effectivedb.org/>).

Results

Strain 814/98's genome sequencing and de novo assembly

In order to sequence the genome of strain 814/98, paired-end de novo sequencing on an Illumina HiSeq2000 platform was conducted to produce 1056 Mb sequence reads from short and long insert libraries, resulting in $160\times$ genome coverage (Table S1). Sequence reads were de novo assembled into 102 contigs and 35 scaffolds. The longest scaffold was 1,437,579 bp in length and the longest contig was 443,518 bp in length with N50 1,124,118 bp for scaffolds and 141,450 bp for contigs (Table 2). Nineteen short contigs/scaffolds were identified to be either duplicates or almost identical to parts of longer contigs/scaffolds. Some of them were duplicated several times. After rejection of short duplicated contigs and scaffolds, the total number of scaffolds was reduced to 22. The main six scaffolds with a size of over 400 kb constituted the core of the 814/98 chromosome. The size of the other six scaffolds ranged from 10 kb to 100 kb. Only 10 scaffolds were shorter than 2 kb. The entire genome size was estimated to be 6,579,377 bp in length and the GC content was 57.97% (Table 2).

Functional annotation of the 814/98 genome

In total, 6024 genes encoding proteins were identified in strain 814/98 (Table 3, S4). The average gene length was estimated to be 934 bp and 400 genes were longer than 2 kb. In total, 92 genes encoding RNAs were identified: 62 genes encoding tRNAs, 16 genes encoding rRNAs (including 7 *rrn5*, 5 *rrn16* and 4 *rrn23*) and 14 genes encoding sRNAs. The length of the non-coding RNAs was estimated to be 838,126 bp. The total length of the gene-encoding sequences was 5,629,212 bp, which constituted 85.56% of the genome, with the intergenic regions constituting 14.44%. GC content within the genic regions was 58.84%.

The gene functional annotation was done by aligning the protein predictions with selected databases. The consistent results were obtained using KEGG and COG databases. Alignment with the COG database allowed for differentiation of 22 general gene classes. The largest number of genes was associated with

Table 1 *Pseudomonas* spp. strains, their NCBI accession numbers and references used in the bioinformatic and phylogenetic analyses. Details are listed in Supplementary Table S5a-c

No.	Species and strain name	NCBI accession number	Reference
1	<i>Pseudomonas syringae</i> pv. <i>lachrymans</i> 814/98	NBLF00000000	This manuscript
2	<i>Pseudomonas syringae</i> pv. <i>lachrymans</i> MAFF301315	NZ_AEAF00000000	Baltrus et al. 2011
3	<i>Pseudomonas syringae</i> pv. <i>lachrymans</i> 98-744A	NZ_LCWT00000000	Jeong et al. 2015
4	<i>Pseudomonas syringae</i> pv. <i>lachrymans</i> 107	NZ_LGLK00000000	Mott et al. 2016
5	<i>Pseudomonas syringae</i> pv. <i>lachrymans</i> YM7902	NZ_LGLI00000000	Mott et al. 2016
6	<i>Pseudomonas syringae</i> pv. <i>lachrymans</i> MAFF302278	NZ_AEAM00000000	Baltrus et al. 2011
7	<i>Pseudomonas syringae</i> pv. <i>lachrymans</i> 3988	NZ_LGLJ00000000	Mott et al. 2016
8	<i>Pseudomonas syringae</i> pv. <i>phaseolicola</i> 1448A	NC_005773	Joardar et al. 2005
9	<i>Pseudomonas syringae</i> pv. <i>tomato</i> DC3000	NC_004578	Buell et al. 2003
10	<i>Pseudomonas syringae</i> pv. <i>aesculi</i> 0893_23	NZ_AEAD00000000	Baltrus et al. 2011
11	<i>Pseudomonas syringae</i> pv. <i>mori</i> 301,020	NZ_AEAG00000000	Baltrus et al. 2011
12	<i>Pseudomonas syringae</i> pv. <i>morsprunorum</i> M302280	NZ_AEAE00000000	Baltrus et al. 2011
13	<i>Pseudomonas syringae</i> pv. <i>sesami</i> ICMP763	NZ_CM000959	Thakur et al. 2016
14	<i>Pseudomonas syringae</i> pv. <i>tabaci</i> ATCC11528	NZ_AEAP00000000	Baltrus et al. 2011
15	<i>Pseudomonas syringae</i> pv. <i>ulmi</i> ICMP3962	NZ_LJRQ01000000	Thakur et al. 2016
16	<i>Pseudomonas savastanoi</i> pv. <i>fraxini</i> ICMP7711	NZ_LLJL00000000	Thakur et al. 2016
17	<i>Pseudomonas syringae</i> pv. <i>glycinea</i> B076	NZ_AEGG00000000	Qi et al. 2011
18	<i>Pseudomonas savastanoi</i> pv. <i>neri</i> ICMP16943	NZ_LJQW01000000	Thakur et al. 2016
19	<i>Pseudomonas savastanoi</i> pv. <i>savastanoi</i> NCPPB3335	NZ_ADMI00000000	Rodríguez-Palenzuela et al. 2010
20	<i>Pseudomonas syringae</i> pv. <i>actinidiae</i> M302091	NZ_AEAL00000000	Baltrus et al. 2011
21	<i>Pseudomonas syringae</i> pv. <i>maculicola</i> ES4326	NZ_AEAK00000000	Baltrus et al. 2011
22	<i>Pseudomonas syringae</i> pv. <i>pisi</i> 1704B	NZ_AEAI00000000	Baltrus et al. 2011
23	<i>Pseudomonas aeruginosa</i> PAO1	NC_002516	Stover et al. 2000
24	<i>Pseudomonas cichorii</i> JBC1	NZ_CP007039	Ramkumar et al. 2015
25	<i>Pseudomonas fluorescens</i> SBW25	NC_012660	Silby et al. 2009
26	<i>Pseudomonas putida</i> KT2440	NC_002947	Belda et al. 2016

Table 2 *Pseudomonas syringae* pv. *lachrymans* strain 814/98's genome assembly summary. The number of core scaffolds is presented in brackets

Parameter	Contig	Scaffold
Total amount	102	35 (22)
Total length (bp)	6,492,840	6,579,377
N50 (bp)	141,450	1,124,118
N90 (bp)	46,110	412,490
The longest one (bp)	443,518	1,437,579
The shortest one (bp)	239	576
GC content (%)	57.97	57.97

membrane transport and metabolism. The predicted functions of 304 and 145 genes were related to transcription and posttranslational modification, and protein turnover, respectively (Fig. S1A). The KEGG database allowed for identification of 885 genes involved in membrane transport. A high number of the identified genes was also involved in metabolism, i.e. 503, 420 and 191 genes were related to amino acid, carbohydrate and energy metabolism, respectively (Fig. S1B). Moreover, a similar number of genes involved in DNA replication, recombination and repair processes and the signal transduction mechanism were identified in both databases (276 and 274 genes, respectively).

Table 3 Functional annotation summary of *Pseudomonas syringae* pv. *lachrymans* strain 814/98's genome

Feature	Number
Genome size (bp)	6,579,377
Gene number	6024
Total gene length (bp)	5,629,212
Gene average length (bp)	934
Gene length / genome (%)	85.56
Total intergenic region length (bp)	950,165
Intergenic region length / genome (%)	14.44
Tandem repeat number	208
Total tandem repeat length (bp)	50,367
Tandem repeat length/genome (%)	0.7655
Minisatellite loci number	77
Microsatellite loci number	21
rRNA number	16
-5S	7
-16S	5
-23S	4
tRNA number	62
sRNA number	14
Total ncRNA length (bp)	838,126
Total ncRNA length/genome (%)	0.4094

Tandem repeats in the 814/98 genome

Genome analysis of 814/98 revealed 208 tandem repeats (TR) – structural genomic components (repeat length from 4 to 1860 bp). These represented ca. 50 kb, i.e. less than 1% of the genome, and consisted of three classes of TR: long tandem repeats, minisatellites and microsatellites (Table 3). Out of 208 TR, 77 were classified as minisatellite DNA and 21 as microsatellite DNA. No CRISPR repeats or spacers were found. The repeat size of minisatellite DNA was from 15 to 63 bp and the total length was about 7 kb (about 0.1% of the genome). Microsatellite DNA possessed repeat size from 4 to 10 bp and a total length of 703 bp (about 0.01% of the genome). Microsatellite DNA may have different repeat unit size and repeat frequency, so it can be useful in molecular diagnostics (Table S3). BLAST analysis was performed for sequences flanking the microsatellite loci, and similarities to protein sequences for most of the loci were found (Table S3).

Phylogenetic placement of strain 814/98

A phylogenetic dendrogram of *P. syringae* strains, with *P. aeruginosa*, *P. cichorii*, *P. fluorescens* and *P. putida* species used as outgroups, consists of three main clusters (Fig. 1). The first large cluster, which corresponds to phylogroup 3 according to Hwang et al. (2005) and genomospecies 2 (Gardan et al. 1999), includes strains that primarily belong to *P. syringae* and *P. savastanoi* species, and it is divided further into subclusters. This large cluster united strains belonging to woody plant pathogens of *Pseudomonas* (pathovars *nerii*, *savastanoi*, *fraxini*, *aesculi*, *ulmi* and *mori*) and herbaceous plant pathogens (pathovars *glycinea*, *phaseolicola*, *tabaci*, *sesami* and *lachrymans*). The five strains of pathovar *lachrymans*, namely 107, 814/98, YM7902, 98-744A and MAFF301315, grouped in phylogroup 3, showed high genetic similarity to one another. The second main cluster corresponded to phylogroup 2 and genomospecies 1 and contained mainly strains of *P. syringae* pv. *syringae*. The third main cluster corresponded to phylogroup 1 and genomospecies 3 and included strains belonging to pathovars *tomato*, *morsprunorum*, *actinidiae*, *maculicola* and several *lachrymans* strains; here were grouped *lachrymans* strains 3988, BG966, and LMG5070.

Comparison of distinct genomes of pathovar *lachrymans* strains

Phylogenetic reconstruction of sequence data clearly showed that strains assigned to *P. syringae* pv. *lachrymans* are grouped into two genetically divergent groups, i.e. phylogroups 3 and 1 (Fig. 1). This divergence is visible in the virulence test on susceptible cucumber accession line B10 (Fig. 2). Strain 814/98 (phylogroup 3) produced necrotic angular leaf spot symptoms (a), whereas strain BG 966 (b) placed in phylogroup 1 produced only weak symptoms. The existence of two distinct clusters within pathovar *lachrymans* was subsequently confirmed at both the nucleotide and amino acid level. 814/98 contigs and scaffolds were subsequently compared to all sequenced genomes of pathovar *lachrymans*, i.e. 98A-744, 107, YM7902, MAFF301315, MAFF302278, 3988 and ICMP3507. We found that the 814/98 genome exhibited the highest similarity to strain 98A-744, than to strains 107 and YM7902, and to MAFF301315 (all in

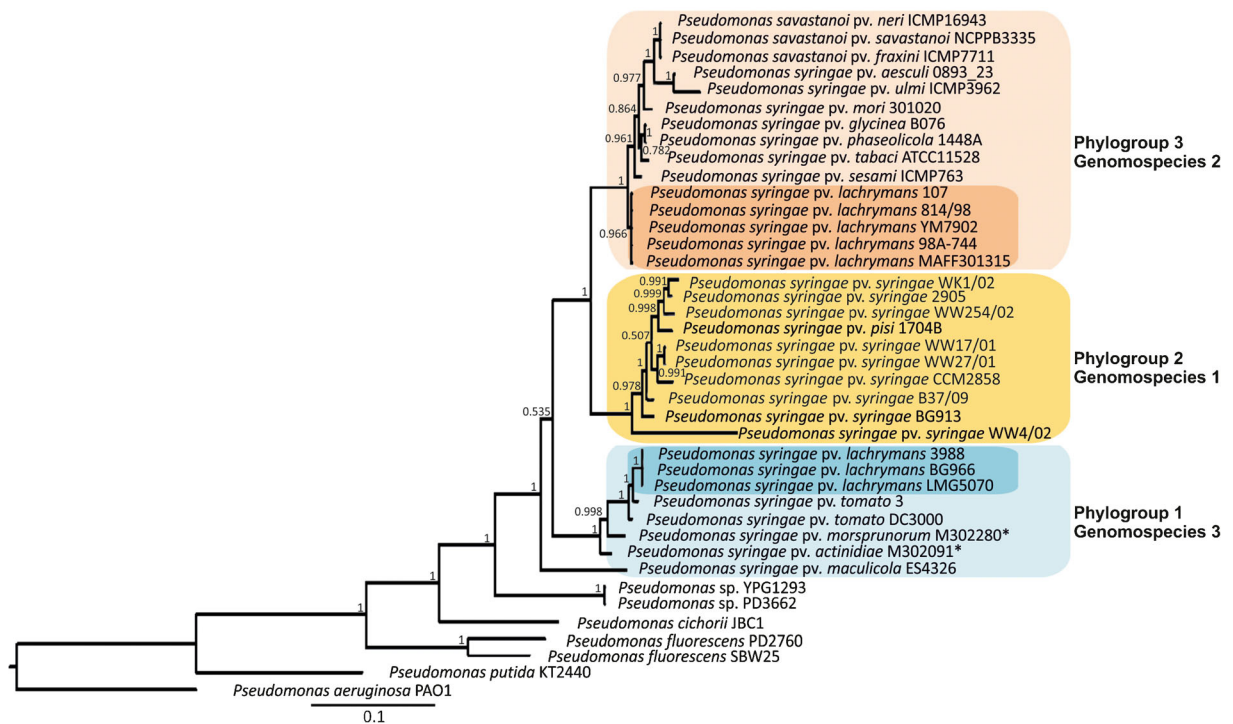


Fig. 1 Phylogenetic relationship among *Pseudomonas* spp. strains based on MLST analysis. A Bayesian dendrogram was constructed based on three housekeeping gene fragments, i.e. *acn*, *gyrB* and *pgi* genes. Nucleotide sequences of *P. aeruginosa*, *P. cichorii*, *P. fluorescens*, and *P. putida* were used as the out-group. Final sequence alignment for 40 selected strains consisted of 1672 nucleotides with concatenated partial gene sequences

(*acn*, *gyrB*, and *pgi*) and partial gene alignments consisted of 521, 672, 475 nucleotides for *acn*, *gyrB*, and *pgi*, respectively. The phylogroups (genomospecies) are indicated in colors. Strains indicated by asterisk (*) were classified into genomospecies 8 by Marcelletti and Scortichini (2014), which is closely-related to genomospecies 3

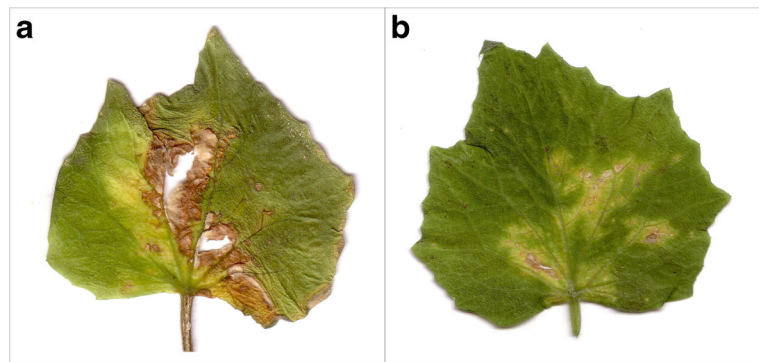
phylogroup 3). In contrast, there was limited similarity to strains MAFF302278, ICMP3507 and 3988 which belonged to phylogroup 1 (data not shown). At the amino acid level the genome of strain 814/98 was compared with *lachrymans* genomes belonging to phylogroup 3 (MAFF301315) and phylogroup 1 (MAFF302278). The analysis detected the evolution of homologous genomes as indicated by variation in the location of gene clusters with similar function. This analysis of representative genomes confirmed the existence of two strain types. Strain 814/98 was very similar to MAFF301315 (phylogroup 3), except for the mega-plasmid sequence which was absent in 814/98 (Fig. 3).

Plasmid identification in strain 814/98

After mapping raw reads on the assembled contigs, a higher than average level of coverage with reads (180–520×) that characterize plasmids was identified for 21

contigs as compared to the average (140–160×) (Table S2). Five groups of contigs were formed according to coverage level. These groups could represent either potential plasmids or repetitive regions, therefore they were further investigated and surveyed for the presence of structural genes or unique sequences associated with *P. syringae* plasmids (Table S6). This survey revealed similarities to plasmids of pathovars *actinidiae*, *tomato*, *maculicola*, *syringae*, *phaseolicola* and *P. fluorescens*. Scaffold 7 showed similarity to the plasmid of *P. fluorescens*. Scaffold 8 showed similarity to the plasmid of *P. syringae* pv. *actinidiae* ICMP 18884, to the large plasmid of pathovar *phaseolicola* strain 1448A, and to both plasmids of the *tomato* DC3000 strain. Up to 20% of scaffold 8's length displayed 100% sequence similarity to plasmids in the bacteria as listed above. Scaffold 9 exhibited 93–100% similarity in 40% of its length to several plasmids, namely to the small plasmid of pathovar *phaseolicola* strain 1448A and to the plasmids of strains

Fig. 2 Differentiation in virulence of *P. syringae* pv. *lachrymans* strains 814/98 (a) and BG 966 (b) on susceptible cucumber line B10 leaves 7 days after inoculation



belonging to pathovar *maculicola* and *actinidiae* (Table S6). The results obtained in the BLAST search were confirmed by reciprocal BLAST. Subsequently, after surveying the genome against *ori* sequences deposited in the databases, a similarity of over 90% was shown on scaffolds 8 and 9, confirming the existence of at least two plasmids in the 814/98 genome. Further blasting against the unique plasmid *repA* gene confirmed that the sequence of plasmid *repA* is present on scaffolds 8 and 9. The length of *repA* was ca. 1300 bp on both scaffolds and similarity was 91% and 87%, respectively. In total, 203 genes were found on the plasmids (Table S7). Most of the genes were identified on scaffold 7 (87) than on scaffold 8 (69) and scaffold 9 (47). However, only single type III effector gene *hopAF1* was found on scaffold 9, where also *repA*, genes encoding plasmid stability protein *SibB*, conjugal transfer and *VirB* proteins related to T4SS and *GntR* were found.

The plasmid genes on scaffold 8 showed similarity to *repA*, *parA*, *parB*, *mobA*, *mobB*, *mobC*, *gntR*, MFS transporter, several *tra* genes and also to conjugal transfer protein genes. This indicates that scaffolds 8 and 9 are probably conjugative plasmids as they possess many genes encoding T4SS and conjugation-related proteins. Genes on scaffold 7 showed similarity to genes encoding proteins related to pillus formation and hypothetical plasmid proteins (Table S7); thus BLAST analysis confirmed the existence of three plasmids in the 814/98 genome.

Identification of type III effector proteins (TTEs)

The genomes of the *Pseudomonas* spp. strains collected in our dataset (Table 1) were characterized for the presence of TTEs by using the approach of Baltrus et al. (2011). Of the 90 examined TTEs, 78 were identified in 22 of the

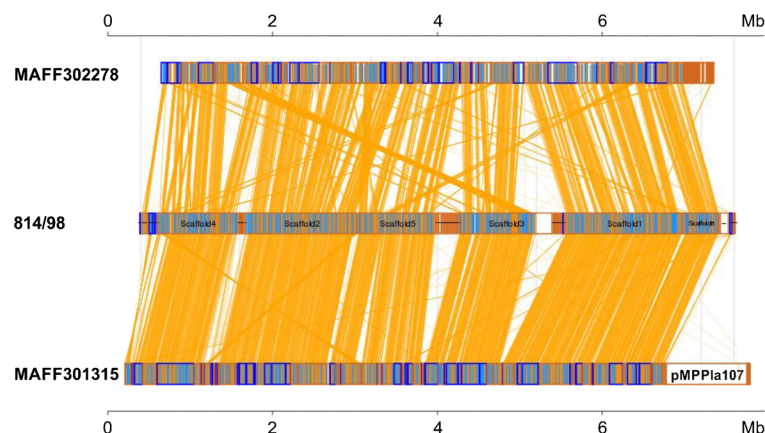


Fig. 3 Linear synteny of *P. syringae* pv. *lachrymans* genomes at the amino acid level. Strain 814/98 was compared with strains MAFF301315 (phylogroup 3) and MAFF302278 (phylogroup 1). In the box of sequences, the orange region stands for the amino acid sequence in the forward chain of the genome sequence and the blue region stands for the amino acid sequence in the reverse

chain of this genome sequence. The orange lines stand for forward alignment of two sequences. Higher synteny is observed between pathovar *lachrymans* strains 814/98 and MAFF301315 than between 814/98 and MAFF302278. Box pMPPla107 indicates 1 Mb megaplasmid present in MAFF301315

analyzed genomes and 24 of them were identified in the 814/98 genome (Table S8). The same effectors were also present in the sequenced genomes of three pathovar *lachrymans* strains, i.e. 98A-744, 107 and YM7902. A different set of TTEs was present in *lachrymans* strains 3998 and MAFF302278. None of the effectors was found in *P. aeruginosa* PAO1, *P. fluorescens* SBW25, *P. cichorii* JBC1 or *P. putida* KT2440. A dendrogram consisting of two major clusters was built based on the TTEs' presence (Fig. 4). The main cluster included strains belonging primarily to the *P. syringae* and *P. savastanoi* species. This cluster was divided into subclusters and contained strains belonging to woody plant pathogens of *Pseudomonas* spp. (pathovars *neri*, *mori*, *fraxini*, *aesculi*, *ulmi* and *savastanoi*) and pathovars which infect a diverse group of plant species (*actinidiae*, *maculicola*, *morsprunorum*, *sesami*, *tabaci* as well as *phaseolicola* and *glycinea*), including a subcluster of five *lachrymans* strains: 814/98, 107, 98A-744, YM7902 and MAFF301315 (phylogroup 3). A second major cluster includes pathovars *tomato*, *syringae*, *pisi* and two *lachrymans* strains from phylogroup 1, namely 3998 and MAFF302278.

After the TTEs analysis in many strains representing different pathovars, we concluded that a core set of eight TTEs was conserved in many of the *P. syringae* strains,

including three fully sequenced genomes of *P. syringae* pv. *tomato* DC3000, *P. syringae* pv. *syringae* B728a and *P. syringae* pv. *phaseolicola* 1488A (bolded in Fig. 5). In addition to this conserved core set, each strain contained several unique TTEs. We compared the TTEs of 814/98 and 3998, the two strains of *P. syringae* pv. *lachrymans* which are located in different phylogroups and showed that besides the core set there were seven additional TTEs common for these strains (Fig. 5). An additional nine TTEs were present only in 814/98. A significantly different TTEs profile was found in strain 3998, as this *lachrymans* strain contained an additional 17 TTEs which are also present in *P. syringae* pv. *tomato* DC3000 (Fig. 5, Table S8). Interestingly, there were two TTEs, i.e. *hopAW1* and *hopBD1*, which were present in *lachrymans* strains belonging to both phylogroups but were absent in the strains of other pathovars and therefore are possibly unique type III effectors for the *lachrymans* pathovar irrespectively of the phylogroup location.

Discussion

Rapid technological progress and cost reduction achieved in DNA sequencing technology have both been

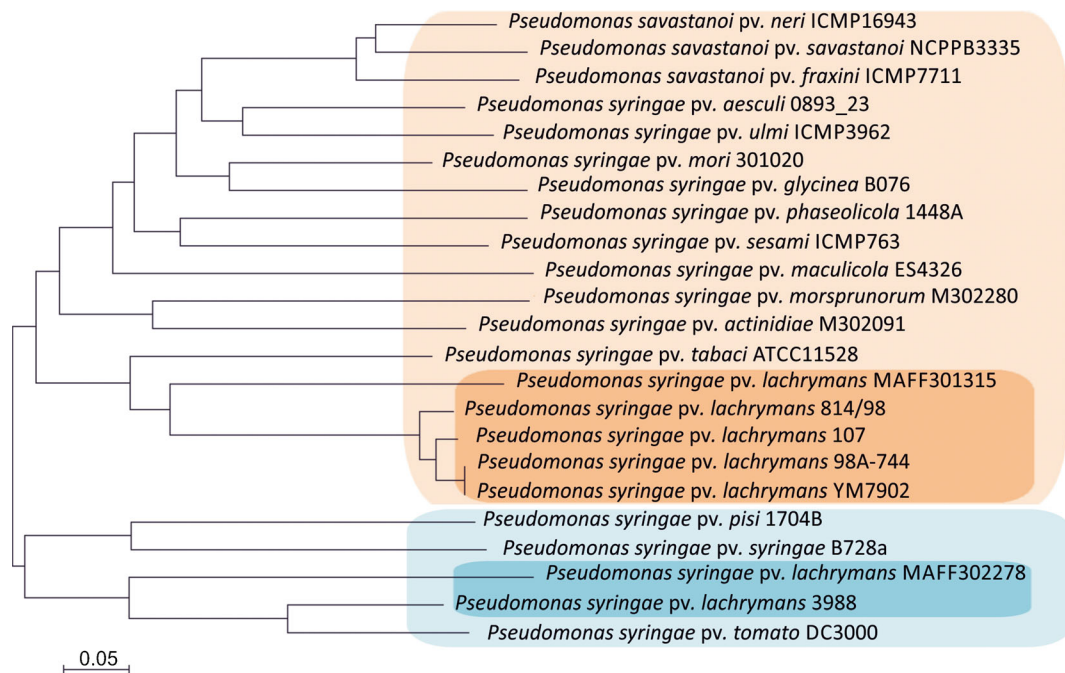


Fig. 4 Relationships among phytopathogenic strains of *Pseudomonas* spp. with a focus on pathovar *lachrymans*, based on the presence of TTEs. The division into phylogroups is indicated in

orange and blue. The binary matrix of the TTEs' presence or absence, which was the basis for construction of the tree, is presented in Table S8

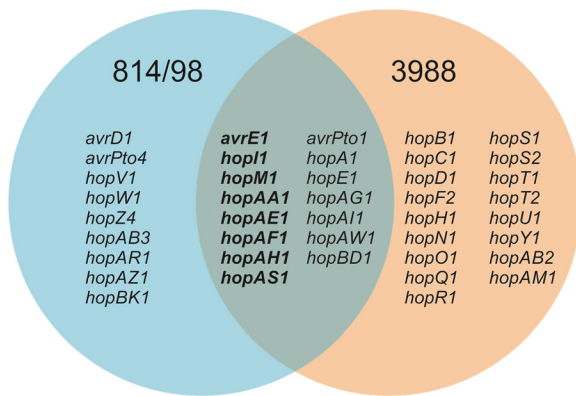


Fig. 5 TTEs present in two strains of *P. syringae* pv. *lachrymans*, namely in strains 814/98 and 3988, which represent two distinct phylogroups. TTEs indicated and bolded in the center of the diagram are conserved among all of the analyzed strains (core set of TTEs). The diagram was constructed based on the TTEs' binary matrix (Table S8)

instrumental in causing the increase in the number of drafts and complete genome sequences. In particular, a plethora of draft genomes were published recently for plant pathogenic bacteria (NCBI 2017). This abundance of information allows to make even more efficient comparisons of related strains, although the quality of the genomes varies. A draft genome sequence of strain 814/98 of *P. syringae* pv. *lachrymans* is presented here. Strain 814/98 is well-characterized phenotypically as highly virulent with the ability to cause large, water-soaked lesions on cucumber leaves that become necrotic after a few days (Fig. 2). The symptoms caused by strain 814/98 on cucumber leaves never failed to develop in every single test over the course of many years of testing. This strain did not produce fluorescent pigment on King's medium B but displayed the typical phenotypic and biochemical characteristics of *P. syringae* in LOPAT tests (Olczak-Woltman et al. 2007; Olczak-Woltman et al. 2008; Słomnicka et al. 2015).

The genome size of strain 814/98 was estimated to be 6.58 Mb in length and the GC content was 57.97%. A genome size of ca. 6–6.5 Mb is typical for the whole *P. syringae* genome, and also typical for pathovar *lachrymans* (Baltrus et al. 2011). The de novo assembly of 814/98 seems to be one of the most complete among the sequenced pathovar *lachrymans* strains so far, as the whole genome is represented in 22 scaffolds. The number of scaffolds in the sequenced genome is dependent on the method used in sequencing and assembling. It is often larger than 30 and increases up to several hundred, whereas the number of contigs increases even up to 50 hundred

(Baltrus et al. 2011; Martínez-García et al. 2015; NCBI 2017). NGS assembly often suffers from short reads and repetitive fragments which influence the analysis of genome coverage with reads. Paired-end sequencing combined with two types of libraries (500 bp and 6500 bp inserts) was used as a solution as it partly compensated for the lack of long reads (Zhang et al. 2011). A hybrid de novo assembly method using a combination of long reads and Illumina short reads data was shown to reduce the contig number after assembly (Boetzer and Pirovano 2014). Different technologies used in genome sequencing and assembly influence sequence quality and cause comparative analysis results to not be convincingly conclusive, e.g. miss-assembly or gaps in the genome sequence of MAFF302278 meant that we were not able to place it on the phylogenetic tree (Fig. 1).

An adequate strain, pathovar and species classification is an important goal. We previously attempted to describe the collected cucumber strains (Olczak-Woltman et al. 2007; Słomnicka et al. 2015); however, a more precise methodology and new data are beginning to challenge old conclusions. Baltrus (2016) proposed a sequence-based classification system that is unambiguous and would enable microbial classification without abandoning previous taxonomic systems. Establishing such a system is imperative because the existence of distinct clusters of strains within pathovars is reported often after genome sequencing, which forces nomenclature revision. Here we present strong evidence that there are two significantly divergent clusters of strains within pathovar *lachrymans* and grouped into different phylogroups. Strains from phylogroup 3 (Fig. 1) form a “necrotic type” of *lachrymans* strains capable of producing symptoms of angular leaf spot disease on cucumber leaves as exemplified by strain 814/98 (Fig. 2a). On the other hand, strains from phylogroup 1 produce only weak symptoms (Fig. 2b). It is interesting to note that Newberry et al. (2016) demonstrated that some strains associated with angular leaf spot in cucurbits are phylogenetically distinct from pathovar *lachrymans*. Strain classification is also an issue in other pathovars. Gironde and Manceau (2012) identified the pathovar *tomato* as a controversial one because of the phenotypic diversity of the strains, particularly at the level of pathogenicity. There are two populations within that pathovar that are pathogenic on different host species and should probably be classified as different pathovars. Moreover, strain DC3000, currently in the pathovar *tomato*, should probably be reclassified into pathovar

maculicola, or the two pathovars should be grouped together (Gironde and Manceau 2012). Similarly, two very distinct clusters of strains in *P. avellanae* led to the claim that nomenclatural revision should be made (Scortichini et al. 2013). The genetic differences within *P. syringae* pv. *actinidiae* and differences in pathogenicity between strains were sufficient to define a new pathovar called *P. syringae* pv. *actinidifoliorum* (Cunty et al. 2015). The results as listed above indicate that the strains should be classified carefully because of genetic and phenotypic diversity. Here we present results which clearly show that there are two groups of *P. syringae* pv. *lachrymans* strains and that perhaps a new pathovar should also be defined. However, we believe that more evidence is needed to define the new pathovar. The strains representing the two *lachrymans* groups should be carefully tested on different cucurbit species and more genomes of *lachrymans* strains have to be analyzed.

Bacterial genomes consist of a chromosome and may contain one or more plasmids, and strains may vary in the number and size of the plasmids (Buell et al. 2003; Feil et al. 2005; Joardar et al. 2005; Zhao et al. 2005; Baltrus et al. 2011). In order to identify plasmids in strain 814/98, we simultaneously incorporated various methodologies, e.g. a search for similarity to known plasmids, bioinformatic analysis of structural plasmid genes, genome coverage with reads and laboratory plasmid isolation (Słomnicka et al. 2015). The bioinformatic analysis showed that strain 814/98 contains at least three plasmids. A small one, approximately 40–60 kb in size, was formed by scaffold 9. It has almost 400× coverage with reads, indicating that it is a medium-copy-number plasmid. A plasmid of this size was previously suggested in strain 814/98 based on electrophoretic plasmid separation using Eckhardt gels (Słomnicka et al. 2015). This plasmid shows sequence similarities to several other plasmids, including the small plasmid of pathovar *phaseolicola* strain 1448A (Joardar et al. 2005), the plasmid of pathovar *syringae* strain UMAF0158, which contains *ruAB*, *repA*, *virB* and *virD* genes, and to scaffold 29 of strain MAFF301315 (Cazorla et al. 2008; Baltrus et al. 2011; Martínez-García et al. 2015). The second, large plasmid of 814/98 is represented by scaffold 8 and possibly other smaller contigs of similar cover with reads (200–250×) that together

are ca. 200 kb in length. This size corresponds to the largest plasmid detected by Słomnicka et al. (2015). This plasmid shows high similarity to plasmids of pathovar *tomato* strain DC3000 and to the large plasmid of pathovar *phaseolicola* strain 1448A (Buell et al. 2003; Joardar et al. 2005). On the other hand, it shows negligible similarity to strain MAFF301315 scaffolds (Baltrus et al. 2011). These two plasmids, represented by scaffolds 8 and 9, possess a large number of genes encoding proteins connected with T4SS and conjugation (Table S7). The presence of plasmids related to conjugation might have resulted in effective acquisition of virulence genes by strain 814/98. The third plasmid detected on Eckhardt gels by Słomnicka et al. (2015) was ca. 100 kb in size and corresponded to scaffold 7, which showed similarity to *P. fluorescens* SBW25 plasmids (Silby et al. 2009) carrying genes connected with pillus formation (Table S7) and no similarity to any sequences of already sequenced pathovar *lachrymans* genomes. This suggests that it may be a unique, relatively large, low-copy-number plasmid of strain 814/98. Unfortunately, we were not able to find the *ori* sequence on scaffold 7. The largest and best described plasmid family in *P. syringae* is pPT23A, also called PFP (Zhao et al. 2005; Ma et al. 2007). PFP plasmids appear to originate from a common ancestor and share homologous RepA-PFP related to Cole2 (Bardaji et al. 2017). The PFP plasmids are from 35 to over 100 kb in size. Based on our results, we concluded that the 814/98 plasmids most likely belong to this family.

The functional analysis of the 814/98 genome revealed similarity to B728a, 1448A and DC3000 strains, although a slightly smaller number of genes was identified. A high degree of conservation between 814/98 and the other strains belonging to the *P. syringae* complex was observed with respect to both the gene and the TTE numbers and classes (Martínez-García et al. 2015). The analysis of TTEs in the 814/98 genome resulted in identification of a total of 24 effectors. The identified TTEs are members of different TTE families as described by Baltrus et al. (2011), belonging either to the core TTEs found in all pathogenic *P. syringae* strains or to TTEs that are diverse in sequence and are present in a wide variety of

genomic locations. The TTE analysis of the *P. syringae* genome presented here is in agreement with previously published results and confirms TTE-based phylogenetic strain grouping (Baltrus et al. 2011; O'Brien et al. 2011; Lindeberg et al. 2012). Moreover, the phylogenetic grouping conducted here corresponds well with phylogroups according to Hwang et al. (2005) and genomospecies according to Marcelletti and Scortichini (2014) (Fig. 4). Almost all of the strains in phylogroup 1 (genomospecies 3), represented by the pathovars *actinidiae*, *morsprunorum*, *tomato* and a few *lachrymans* strains, carry at least 30 functional TTEs. In phylogroup 2 (genomospecies 1 – represented by pathovar *syringae*) and phylogroup 3 (genomospecies 2), containing, among others, “necrotic type” *lachrymans* strains, the number of effectors ranges between 19 and 30. Phylogenetic analysis of the TTEs suggests that some of the effectors were acquired, lost and reacquired (O'Brien et al. 2011; Lindeberg et al. 2012). This pattern indicates a heterogeneous distribution of TTEs within *P. syringae*. This observation may support the idea that there is strong selection pressure for the loss of effectors that can be recognized by host plants. In the case of pathovar *lachrymans*, the existence of two different strain sub-clusters was suggested previously (Ślomsnicka et al. 2015). Huge differences in *lachrymans* strains clustered in different phylogroups indicate that they may have evolved separately from different species.

Acknowledgments This study was supported by a research program of the Polish Ministry of Agriculture and Rural Development titled “Basic research for biological progress in crop production”. We thank Karolina Kaźmińska for her help in the bioinformatics analysis.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

Human and animal rights The research did not involve human participants and/or animals.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided

you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Baltrus, D. A. (2016). Divorcing strains classification from species names. *Trends in Microbiology*, 24, 431–439.
- Baltrus, D. A., Nishimura, M. T., Romanchuk, A., et al. (2011). Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. *PLoS Pathogens*, 7, e1002132.
- Baltrus, D. A., McCann, H. C., & Guttman, D. S. (2017). Evolution, genomics and epidemiology of *Pseudomonas syringae*. *Molecular Plant Pathology*, 18, 152–168.
- Bard, J., & Winter, R. (2000). Gene ontology tool for the unification of biology. *Nature Genetics*, 25, 25–29.
- Bardaji, L., Añorga, M., Ruiz-Masó, J. A., del Solar, G., & Murillo, J. (2017). Plasmid replicons from *Pseudomonas* are natural chimeras of functional, exchangeable modules. *Frontiers in Microbiology*, 8, 190.
- Belda, E., van Heck, R. G., Lopez-Sanchez, M. J., et al. (2016). The revisited genome of *Pseudomonas putida* KT2440 enlightens its value as a robust metabolic chassis. *Environmental Microbiology*, 18, 3403–3424.
- Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, 27, 573–580.
- Boetzer, M., & Pirovano, W. (2014). SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics*, 15, 211.
- Buell, C. R., Joardar, V. R., Lindeberg, M., & et al. (2003). The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 10181–10186.
- Cazorla, F. M., Codina, J. C., Abad, C., et al. (2008). 62-kb plasmids harboring *ruAB* homologues confer UV-tolerance and epiphytic fitness to *Pseudomonas syringae* pv. *syringae* mango isolates. *Microbial Ecology*, 56, 286–291.
- Cunty, A., Poliakov, F., Rivoal, C., et al. (2015). Characterization of *Pseudomonas syringae* pv. *actinidiae* (Psa) isolated from France and assignment of Psa biovar 4 to a *de novo* pathovar: *Pseudomonas syringae* pv. *actinidifoliorum* pv. Nov. *Plant Pathology*, 64, 582–596.
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, 9, 772–772.
- Delcher, A. L., Bratke, K. A., Powers, E. C., & Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with glimmer. *Bioinformatics*, 23, 673–679.
- FAO website (2017) <http://www.fao.org/faostat/en/> (accessed February 2017).
- Feil, H., Feil, W. S., Chain, P. S., et al. (2005). Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 11064–11069.
- Gardan, L., Shafik, H., Belouin, S., Broch, R., Grimont, F., & Grimont, P. A. D. (1999). DNA relatedness among the

- pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremiae* sp. nov. and *Pseudomonas cannabina* sp. nov. (ex Sutic and Dowson, 1959). *International Journal of Systematic Bacteriology*, 49, 469–478.
- Gironde, S., & Manceau, C. (2012). Housekeeping gene sequencing and multilocus variable-number tandem-repeat analysis to identify subpopulations within *Pseudomonas syringae* pv. *maculicola* and *Pseudomonas syringae* pv. *tomato* that correlate with host specificity. *Applied and Environmental Microbiology*, 78, 3266–3279.
- Hwang, M. S., Morgan, R. L., Sarkar, S. F., Wang, P. W., & Guttman, D. S. (2005). Phylogenetic characterization of virulence and resistance phenotypes of *Pseudomonas syringae*. *Applied and Environmental Microbiology*, 71, 5182–5191.
- Jeong, H., Kloepper, J. W., & Ryu, C.-M. (2015). Genome sequences of *Pseudomonas amygdali* pv. *tabaci* strain ATCC 11528 and pv. *lachrymans* strain 98A-744. *Genome Announcements*, 3, e00683–e00615.
- Joardar, V., Lindeberg, M., Jackson, R. W., et al. (2005). Whole-genome sequence analysis of *Pseudomonas syringae* pv. *phaseolicola* 1448A reveals divergence among pathovars in genes involved in virulence and transposition. *Journal of Bacteriology*, 187, 6488–6498.
- Kanehisa, M., Goto, S., Hattori, M., et al. (2006). From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Research*, 34, D354–D357.
- Lagesen, K., Hallin, P. F., Rødland, E., Staerfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, 35, 3100–3108.
- Lamichhane, J. R., Messéan, A., & Morris, C. E. (2015). Insights into epidemiology and control of diseases of annual plants caused by the *Pseudomonas syringae* species complex. *Journal of General Plant Pathology*, 81, 331–350.
- Li, R., Li, Y., Kristiansen, K., & Wang, J. (2008). SOAP: Short oligonucleotide alignment program. *Bioinformatics*, 24, 713–714.
- Lindeberg, M., Cunnac, S., & Collmer, A. (2012). *Pseudomonas syringae* type III effector repertoires: Last words in endless arguments. *Trends in Microbiology*, 20, 199–208.
- Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25, 955–964.
- Ma, Z., Smith, J. J., Zhao, Y., Jackson, R. W., Arnold, D. L., Murillo, J., & Sundin, G. W. (2007). Phylgenetic analysis of the pPT23A plasmid family of *Pseudomonas syringae*. *Applied and Environmental Microbiology*, 73, 1287–1295.
- Marcelletti, S., & Scortichini, M. (2014). Definition of plant-pathogenic *Pseudomonas* genomospecies of the *Pseudomonas syringae* complex through multiple comparative approaches. *Phytopathology*, 104, 1274–1282.
- Martínez-García, P., Rodríguez-Palenzuela, P., Arrebola, E., et al. (2015). Bioinformatics analysis of the complete genome sequence of the mango tree pathogen *Pseudomonas syringae* pv. *syringae* UMAF0158 reveals traits relevant to virulence and epiphytic lifestyle. *PLoS One*, 10, e0136101.
- Meng, X., Xie, X., Shi, Y., et al. (2016). Evaluation of a loop-mediated isothermal amplification assay based on *hrpZ* gene for rapid detection and identification of *Pseudomonas syringae* pv. *lachrymans* in cucumber leaves. *Journal of Applied Microbiology*, 122, 441–449.
- Mott, G. A., Thakur, S., Smakowska, E., et al. (2016). Genomic screens identify a new phytochemical microbe-associated molecular pattern and the cognate *Arabidopsis* receptor-like kinase that mediates its immune elicitation. *Genome Biology*, 17, 98.
- NCBI website (2017) <http://www.ncbi.nlm.nih.gov/assembly/organism/136849/all/> (accessed February 2017).
- Newberry, E. A., Jardini, T. M., Rubio, I., et al. (2016). Angular leaf spot of cucurbits is associated with genetically diverse *Pseudomonas syringae* strains. *Plant Disease*, 100, 1397–1404.
- O'Brien, H. E., Thakur, S., & Guttman, D. S. (2011). Evolution of plant pathogenesis in *Pseudomonas syringae*: A genomics perspectives. *Annual Review of Phytopathology*, 49, 269–289.
- Olczak-Woltman, H., Masny, A., Bartoszewski, G., Phucienniczak, A., & Niemirowicz-Szczytt, K. (2007). Genetic diversity of *Pseudomonas syringae* pv. *lachrymans* strains isolated from cucumber leaves collected in Poland. *Plant Pathology*, 56, 373–382.
- Olczak-Woltman, H., Schollenberger, M., Mądry, W., & Niemirowicz-Szczytt, K. (2008). Evaluation of cucumber (*Cucumis sativus* L.) cultivars grown in Eastern Europe and progress in breeding for resistance to angular leaf spot (*Pseudomonas syringae* pv. *lachrymans*). *European Journal of Plant Pathology*, 122, 385–393.
- Qi, M., Wang, D., Bradley, C. A., & Zhao, Y. (2011). Genome sequence analyses of *Pseudomonas savastanoi* pv. *glycinea* and subtractive hybridization-based comparative genomics with nine pseudomonads. *PLoS One*, 6, e16451.
- Ramkumar, G., Lee, S. W., Weon, H.-Y., Kim, B.-Y., & Lee, Y. H. (2015). First report on the whole genome sequence of *Pseudomonas cichorii* strain JBC1 and comparison with other *Pseudomonas* species. *Plant Pathology*, 64, 63–70.
- Rodríguez-Palenzuela, P., Matas, I. M., Murillo, J., 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. *Environmental Microbiology*, 12, 1604–1620.
- Romanchuk, A., Jones, C. D., Karkare, K., et al. (2014). Bigger is not always better: Transmission and fitness burden of approximately 1MB *Pseudomonas syringae* megaplasmid pMPPla107. *Plasmid*, 73, 16–25.
- Ronquist, F., Teslenko, M., van der Mark, P., et al. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61, 539–542.
- Sarkar, S. F., & Guttman, D. S. (2004). Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Applied and Environmental Microbiology*, 70, 1999–2012.
- Scortichini, M., Marcelletti, S., Ferrante, P., & Firrao, G. (2013). A genomic redefinition of *Pseudomonas avellanae* species. *PLoS One*, 8, e75794.
- Silby, M. W., Cerdano-Tarraga, A. M., Vernikos, G. S., et al. (2009). Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. *Genome Biology*, 10, 1–R51.16.
- Stomnicka, R., Olczak-Woltman, H., Bartoszewski, G., & Niemirowicz-Szczytt, K. (2015). Genetic and pathogenic

- diversity of *Pseudomonas syringae* strains isolated from cucurbits. *European Journal of Plant Pathology*, 141, 1–14.
- Stover, C. K., Pham, X. Q., Erwin, A. L., et al. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*, 406, 959–964.
- Tatusov, R. L., Fedorova, N. D., Jackson, J. D., et al. (2003). The COG database: An updated version includes eukaryotes. *BMC Bioinformatics*, 4, 41.
- Thakur S, Weir BS, Guttman DS (2016) Phytopathogen genome announcement: Draft genome sequences of 62 *Pseudomonas syringae* type and pathotype strains. Mol plant–microbe interact MPMI01160013TA.
- Zhang, W., Chen, J., Yang, Y., Tang, Y., Shang, J., & Shen, B. (2011). A practical comparison of *de novo* genome assembly software tools for next-generation sequencing technologies. *PLoS One*, 6, e17915.
- Zhao, Y., Ma, Y. Z., & Sundin, G. W. (2005). Comparative genomic analysis of the pPT23A plasmid family of *Pseudomonas syringae*. *Journal of Bacteriology*, 187, 2113–2126.