Functional annotation of the genome of plant pathogenic bacteria *Pseudomonas syringae* pv. *lachrymans* strain 814/98

RENATA SŁOMNICKA, HELENA OLCZAK-WOLTMAN, GRZEGORZ BARTOSZEWSKI

Department of Plant Genetics, Breeding and Biotechnology
Faculty of Horticulture, Biotechnology and Landscape Architecture, Warsaw University of Life Sciences-SGGW
Nowoursynowska 159, 02-787 Warsaw, Poland

Introduction

Over 50 pathovars are distinguished in the plant pathogenic bacterium species *Pseudomonas syringae*, including *P. syringae* pv. *lachrymans* which causes angular leaf spot disease in cucumber (*Cucumis sativus* L.). The bacterial strains of pathovar *lachrymans* display within-pathovar variability and vary in both the virulence and DNA level (Olczak-Woltman et al. 2007, Słomnicka et al. 2015a). Recently, highly virulent strain's genome was sequenced and the draft genome sequence was assembled into 35 scaffolds built of 102 contigs. The size of the genome was estimated to be 6.58 Mb, with a GC content of 57.97% (Słomnicka et al. 2015b).

The aim of this study was to functionally annotate the draft genome sequence of this highly virulent strain Psl814/98.

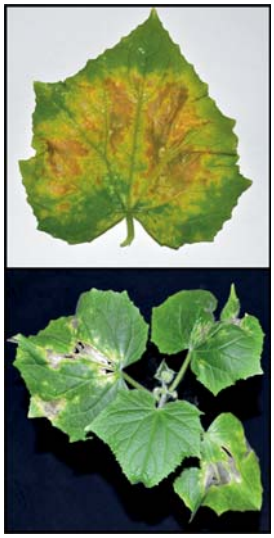


Figure 1. Angular leaf spot symtopms on cucumber plants (line B10), caused by *Pseudomonas syringae* pv. *lachrymans* highly virulent strain 814/98, evaluated seven days post inoculation.

Materials & Methods

Genome sequence. The draft genome sequence of strain Psl814/98 that consisted of Illumina HiSeq2000 paired-end reads, assembled using a SOAPdenovo assembler into 35 scaffolds and 102 contigs, was analyzed (Słomnicka et al. 2015b).
Gene prediction. Genes coding proteins were predicted from the genome assembly using Glimmer v. 3.02 (Delcher et al. 2007) which is developed for microorganisms such as bacteria. Genes coding rRNA- and tRNA were identified using RNAmmer v.1.2 (Lagesen et al. 2007) and tRNAscan-SE v.1.3.1 (Lowe et al. 1997). sRNA genes were predicted using the Rfam database (Gardner et al. 2009).
Gene annotation. Functional gene annotation was done via analysis of protein sequences. Genes were aligned with several databases to obtain their corresponding annotations. The following databases were searched: KEGG v.59 (Kanehisa et al. 2006) and COG v. 20090331 (Tatusov et al. 2003), Swiss-Prot v. 2011_10_19 (Magrane 2011) and GO v.1.419 (Bard and Winter 2000). To ensure biological meaning, only the highest quality alignment result was chosen for gene annotation.
Identification of Type III Effectors (TTEs). Pre-identification of TTEs for strain Psl814/98 was performed by comparing protein predictions deposited at the Pseudomonas Plant Interaction PPI database (http://www.pseudomonas-syringae.org/) using the BLAST algorithm (Altschul et al. 1990). It was considered that Psl814/98 possesses an effector if the protein sequence had a significant BLAST hit with an identity of at least 90%. Additionally, the protein sequences were evaluated using Effective T3 v.1.0.1. software (Jehl et al. 2011).

Results - gene prediction

In total, 78 genes encoding RNAs were identified: 62 genes encoding tRNAs and 16 genes encoding rRNAs, including 7 rrn5, 5 rrn16 and 4 rrn23 genes. Generally, in prokaryotic rRNA, 5S consists of 120 nucleotides, 16S consists of 1540 nucleotides and 23S consists of 2900 nucleotides. The number of genes encoding tRNA and rRNA in Psl814/98 was exactly the same as in the reference *P. syringae* strains (Martínez-García et al. 2015). Moreover, 14 genes encoding sRNA (length range from 50 nt to 500 nt) were identified.

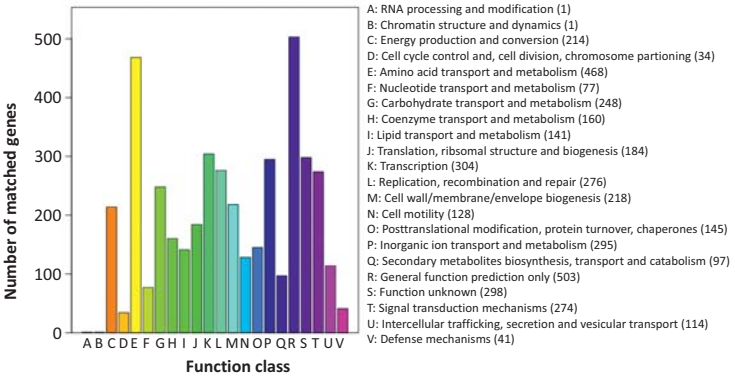
Table 1. Statistics of ncRNA genes for the Psl814/98 strain.

Type		Number	Average length (bp)	Total length (bp)	In genome (%)
tRNA		62	78	4855	0.074
rRNA	5S	7	116	812	
	16S	5	1527	7635	0.304
	23S	4	2891	11564	
sRNA		14	148	2072	0.032

Results - gene annotation

Gene function annotation was done by aligning the protein predictions with selected databases. The most consistent results were obtained with the KEGG and COG databases. Aligning with the KEGG and COG databases allowed to identify a large number of genes involved in membrane transport, amino acid metabolism, carbohydrate metabolism, and energy metabolism (Figure 2). The results of the number and classes of genes, if compared to the reference *P. syringae* strains, shows huge similarity to *P. syringae* strains B728a, 1448A and DC3000 representing pathovars: *syringae*, *phaseolicola*, and *tomato*, although a slightly smaller number of genes was identified in each class (Martínez-García et al. 2015).

Figure 2. COG-based functional annotation of Psl814/98 genome-predicted proteins.



Results - identification of Type III Effectors (TTEs)

The preliminary search for TTEs, which are components related to pathogenicity, allowed to identify so far a total of 6 effectors (*avrE1, hopI1, hopM1, hopAA1, hopAF1, hopAH1*) typical for *Pseudomonas* sp. In the Psl814/98 genome 24 effectors were identified: 6 core TTEs, 10 TTEs identified also in reference *Pseudomonas* genomes (B728a, 1448A or DC3000) and 8 TTEs distinctive Psl814/98 strain.

The identified TTEs belong to both the core TTEs found within all pathogenic *P. syringae* strains and to flexible TTEs located in a wide variety of genomic locations that are very diverse in sequence (Baltrus et al. 2011).

Summary

The bioinformatic functional annotation of the genome for *P. syringae* pv. *lachrymans* strain Psl814/98, a pathogen of cucurbits, revealed a high degree of conservation with other pseudomonads belonging to *P. syringae* complex. This conservation is observed in both the gene number and classes, and in the TTE number and types. This work provides the basis for further functional analysis of the factors governing host specificity in *P. syringae* pathovar *lachrymans* - cucumber interaction.

Acknowledgments

This research was supported by a Polish Ministry of Agriculture and Rural Development project.

Literature

Baltrus DA, Nishimura MT, Romanchuk A, Chang JH, Mukhtar MS, Cherkis K, Roach J, Grant SR, Jones CD, Dang JL (2011): Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. PLoS ONE 7, e1002132
Olczak-Woltman H, Masny A, Bartoszewski G, Płucienniczak A, Niemirowicz-Szczytt K (2007): Genetic diversity of *Pseudomonas syringae* pv. *lachrymans* strains isolated from cucumber leaves collected in Poland. Plant Pathol. 56, 373382
Słomnicka R, Olczak-Woltman H, Bartoszewski G, Niemirowicz-Szczytt K (2015a): Genetic and pathogenic diversity of *Pseudomonas syringae* strains isolated from cucurbits. Eur. J. Plant Pathol. 141, 114
Słomnicka R, Olczak-Woltman H, Kondzior E, Niemirowicz-Szczytt K, Bartoszewski G (2015b): Draft genome sequence of *Pseudomonas syringae* pv. *lachrymans* - the causal agent of cucumber angular leaf spot disease. DIAGMOL 2015: XVI Conference on "Molecular biology in diagnostics of infectious diseases and biotechnology" in memory of Professor Władysław J.H. Kunicki-Goldfinger, Warsaw University of Life Science-SGGW, Warsaw, Poland, 28 November 2015, Warsaw, 2015, pp: 140143
Martínez-García P, Rodríguez-Palenzuela P, Arrebola E, Carrión V, Gutiérrez-Barranquero J, Pérez-García A, Ramos C, Cazorla F, Vicente A (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