Czajkowski et al.,

IFB running title: Complete genome of Dickeya solani strain IPO 2019

High-quality complete genome resource of plant pathogenic bacterium *Dickeya solani* strain IPO 2019 isolated from *Hyacinthus orientalis*

Robert Czajkowski^{1 †}, Lukasz Rabalski², Przemysław Bartnik¹, Sylwia Jafra³

¹ Division of Biologically Active Compounds, Intercollegiate Faculty of Biotechnology UG and MUG, University of Gdansk, Antoniego Abrahama 58, 80-307 Gdansk, Poland

² Division of Recombinant Vaccines, Intercollegiate Faculty of Biotechnology UG and MUG, University of Gdansk, Antoniego Abrahama 58, 80-307 Gdansk, Poland

³ Division of Biological Plant Protection, Intercollegiate Faculty of Biotechnology UG and MUG, University of Gdansk, Antoniego Abrahama 58, 80-307 Gdansk, Poland

[†] to whom correspondence should be addressed:

Robert Czajkowski, Division of Biologically Active Compounds, Intercollegiate Faculty of Biotechnology UG and MUG, University of Gdansk, Antoniego Abrahama 58, 80-307 Gdansk, Poland, phone: 0048 58 5236333, e-mail: robert.czajkowski@ug.edu.pl

Czajkowski et al.,

Abstract

Dickeya solani is an emerging plant pathogenic bacterium, causing disease symptoms in a variety of agriculturally relevant crop species worldwide. To date a number of *D. solani* genomes have been sequenced and characterized, the great majority of these genomes have however come from *D. solani* strains isolated from potato (*Solanum tuberosum* L.) and not from other plant hosts. Herewith, we present the first complete, high-quality genome of *D. solani* strain IPO 2019 (LMG 25990) isolated from ornamental plant *Hyacinthus orientalis*. The genome of *D. solani* strain IPO 2019 consists of one chromosome of 4,919,542 bp., with a GC content of 56.2% and no plasmids. The genome contains 4502 annotated features, 22 rRNA genes, 73 tRNA genes and 1 CRISPRS. We believe that the information of this high-quality, complete, closed genome of *D. solani* strain isolated from host plant different than potato (i.e. hyacinth) will provide resources for comparative genomic studies as well as for analyses targeting adaptation and ecological fitness mechanisms present in *Dickeya solani* species.

Key words: *Erwinia chrysanthemi*, soft rot, blackleg, potato, plant pathogen, adaptation, genomics

Genome announcement

Dickeya solani (former *Dickeya* spp. biovar 3) is an emerging plant pathogenic bacterium, causing disease symptoms in a variety of agricultural plant species worldwide (Toth et al., 2011; van der Wolf et al., 2014). The pathogen was first reported in potato in early 2000s as a distinct genetic clade of biovar 3, causing severe outbreaks in potato cultivation in different European countries (Czajkowski et al., 2009; Slawiak et al., 2009; Laurila et al., 2010). Until now the pathogen has been repeatably isolated from symptomatic as well as from symptomless potato plants grown in agricultural fields across Europe and worldwide (van der Wolf et al., 2014). As *D. solani* was reported to cause severe

Czajkowski et al.,

disease symptoms under hot climatic conditions as well as there is an observed relationship between *D. solani* high disease incidences in crops and the presence of hot growing seasons in temperate countries (Degefu et al., 2013; Tsrer et al., 2013) it has been speculated that the bacterium has been introduced from subtropical and/or tropical agricultural regions outside Europe to Europe (Slawiak et al., 2009; Toth et al., 2011). Likewise, it has been suggested that *D. solani* might have first appeared in ornamental plants cultivated commercially outside Europe and/or in plants grown in Europe in glasshouses under elevated temperature regime and from there spread to potato ecosystem in temperate regions (Slawiak et al., 2009). At the moment, due to the limited number of *D. solani* strains isolated from different (e.g. (sub)tropical, ornamental) plant hosts, there are not enough data supporting either of these hypotheses. It is well accepted however that the pathogen has been only recently introduced to potato ecosystem, as most strains of *D. solani* analyzed belong to the same haplotype and expresses only a low number of genetic variations (Khayy et al., 2016).

Since the establishment of *D. solani* species in 2014 (van der Wolf et al., 2014), the strain IPO 2222 (DSM 28711, LMG 25993, NCPPB 4479) is recognized as a type strain. A complete, closed, high quality genome sequence of *D. solani* strain IPO 2222 isolated from potato together with the corresponding phylogenetic and phenotypic information about the IPO 2222 has been published (Khayy et al., 2016). Likewise, to date, a number of *D. solani* genomes have been sequenced and characterized (<https://www.ncbi.nlm.nih.gov/genome/14829>). At the moment of manuscript preparation (February 2021) the great majority of these (completed and draft) genome sequences however have come from *D. solani* strains isolated exclusively from potato and not from other plant hosts.

D. solani strain IPO 2019 was isolated from *Hyacinthus orientalis* (common name: garden hyacinth, Dutch hyacinth) in 2009 in the Netherlands. The strain has been deposited in the Belgian Co-ordinated Collection of Micro-organisms (BCCM) by Johan van Vaerenbergh (ILVO Eenheid Plant, Onderzoeksdomein Gewasbescherming, Belgium) under collection number LMG 25990. In our

Czajkowski *et al.*,

comparative phenotypic studies involving *D. solani* strains IPO 2019 and IPO 2222, we detected that IPO 2019 but not IPO 2222 was able to utilize a unique panel of saccharides including dextrin, D-maltose, D-trehalose, gentobiose, D-turanose, L-rhamnose and sorbitol as carbon sources (Bartnik, Czajkowski, unpublished). To further explore the hypothesis that IPO 2019 isolated from hyacinth may show distinct phenotypic and genetic characteristics from *D. solani* type strain IPO 2222 isolated from potato and to link the preliminarily observed phenotypes with the genomic data, the genome of IPO 2019 strain has been sequenced and annotated.

For isolation of genomic DNA, *D. solani* IPO 2019 was grown at 28 °C in tryptone soya broth (TSB, Oxoid) for 16 h with shaking (150 rpm). Bacterial DNA was isolated using a Wizard Genomic DNA Purification Kit (Promega) according to the protocol provided by the manufacturer. The IPO 2019 genome was sequenced by a hybrid approach with long reads generated using Oxford Nanopore Technologies (ONT) and short reads generated using Illumina technology. DNA sequencing libraries were prepared according to manufacturer (for ONT it was LSK kit and for Illumina it was Nextera XT kit). Long reads were generated by single Flongle run on MinION and short reads were generated during Midoutput run on Illumina Miniseq. *De novo* assembly of long reads was performed using Flye 2.8.2. (<https://github.com/fenderglass/Flye>) with final mean coverage x 28. Initial genome polishing was conducted by tool integrated in the assembler. Pilon 1.23 (<https://github.com/broadinstitute/pilon>) was used to further error corrections. Final mean coverage of short reads to polished genome was 748.9. The combined procedure produced a single contig of 4,919,542 bp., which was further manually inspected and curated. Repetitive genetic elements were searched using Geneious (Kearse et al., 2012). Finally, BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), InterProScan (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>) and HMMER (phmmer, UniProtKB) (Finn et al., 2011) (<http://hmmer.org/>) were used to annotate the IPO 2019 genome. The obtained structural and functional annotations were retested using RAST (Rapid Annotation using Subsystem Technology

Czajkowski *et al.*,

(<http://rast.nmpdr.org/>) (Aziz *et al.*, 2008) and DFAST (<http://dfast.ddbj.nig.ac.jp/>) (Tanizawa *et al.*, 2018). Preliminary comparative genomics analyses (IPO 2019 vs. IPO 2222) were done using EDGAR (Blom *et al.*, 2009; Blom *et al.*, 2016). The presence of prophages (bacteriophage sequences) in the genome of IPO 2019 was predicted using PHASTER (Arndt *et al.*, 2016) and the presence of gene clusters involved in biosynthesis of secondary metabolites was predicted using antiSMASH 5.0 database (Blin *et al.*, 2019).

The complete, closed, high-quality genome of *D. solani* strain IPO 2019 has been uploaded to NCBI Genbank and received the accession number: CP071062. Upon submission to NCBI Genbank, the genome was also annotated by the NCBI Prokaryotic Genome Annotation Pipeline (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/) (Tatusova *et al.*, 2016). The genome is double-stranded 4,919,542 bp. DNA sequence, possessing 4502 annotated features and an average GC content of 56.2% with no plasmids. The average gene length was predicted to be ca. 1200 bp. and the average protein length was predicted to be ca. 333 amino acids. 85.8% of the genome consists of coding regions. The *D. solani* IPO 2019 genome encodes 22 rRNAs, 73 tRNAs and 1 CRISPRs. We also identified 11 gene clusters responsible for synthesis of secondary metabolites. Six of these matched known clusters for the biosynthesis of NRPS (non-ribosomal peptide synthases), batalactone, cyanobacin, arylpolyene and thiopeptide. The investigation of IPO 2019 genome for presence of viral sequences done with PHASTER did not result in prediction of regions containing intact (complete) viral genomes, three regions in the genome were however found containing incomplete prophage genomes.

The first, completed, high-quality genome of *D. solani* strain isolated from hyacinth plant may prove useful in (comparative) adaptation studies and ecological fitness analyses done to better understand the origin, spread and virulence traits of this new, emerging *Dickeya* species.

Czajkowski et al.,

1 **Authors' statement**

2 The authors declare than no conflict of interest exists.

4 **Acknowledgements**

5 The authors would like to thank Dr. Jan M. van der Wolf (Plant Research International, Wageningen
6 University and Research Center, Wageningen, The Netherlands) for providing *D. solani* strain IPO
7 2019 for this study. The work was financially supported by Polish Ministry of Higher Education
8 (Ministers two Nauki i Szkolnictwa Wyzszego, Polska) funds DS 531-N104-D800-21 to RC and by
9 Polish Ministry of Science and Higher Education (Ministerstwo Nauki i Szkolnictwa Wyzszego,
10 Polska) funds DS 531-M105-D786-21 to SJ.

12 **Literature cited**

13 Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. 2016. PHASTER: a
14 better, faster version of the PHAST phage search tool. Nucl. Acids Res. 44:W16-W21.
15 Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S.,
16 Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A.,
17 McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R.,
18 Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. 2008. The RAST Server: Rapid
19 Annotations using Subsystems Technology. BMC Genomics 9:75-75.
20 Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., and Weber, T.
21 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucl.
22 Acids Res. 47:W81-W87.

Czajkowski et al.,

- 1 Blom, J., Albaum, S., Doppmeier, D., Puhler, A., Vorholter, F.-J., Zakrzewski, M., and Goesmann, A.
- 2 2009. EDGAR: A software framework for the comparative analysis of prokaryotic genomes.
- 3 BMC Bioinformatics 10:154.
- 4 Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., and Goesmann, A. 2016. EDGAR 2.0:
- 5 an enhanced software platform for comparative gene content analyses. Nucl. Acids Res.
- 6 44:W22-W28.
- 7 Czajkowski, R., van Veen, J.A., and van der Wolf, J.M. 2009. New biovar 3 *Dickeya* spp. strain (syn.
- 8 *Erwinia chrysanthemi*) as a causative agent of blackleg in seed potato in Europe.
- 9 Phytopathology 99:27.
- 10 Degefu, Y., Potrykus, M., Golanowska, M., Virtanen, E., and Lojkowska, E. 2013. A new clade of
- 11 *Dickeya* spp. plays a major role in potato blackleg outbreaks in North Finland. Annal. Appl.
- 12 Biol. 162:231-241.
- 13 Finn, R.D., Clements, J., and Eddy, S.R. 2011. HMMER web server: interactive sequence similarity
- 14 searching. Nucl. Acids Res. 39:W29-W37.
- 15 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A.,
- 16 Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. 2012.
- 17 Geneious Basic: An integrated and extendable desktop software platform for the organization
- 18 and analysis of sequence data. Bioinformatics 28:1647-1649.
- 19 Khayi, S., Blin, P., Chong, T.M., Chan, K.-G., and Faure, D. 2016. Complete genome anatomy of the
- 20 emerging potato pathogen *Dickeya solani* type strain IPO 2222(T). Stand. Genomic Sci. 11:87.
- 21 Laurila, J., Hannukkala, A., Nykyri, J., Pasanen, M., Hélias, V., Garlant, L., and Pirhonen, M. 2010.
- 22 Symptoms and yield reduction caused by *Dickeya* spp. strains isolated from potato and river
- 23 water in Finland. Eur. J. Plant Pathol. 126:249-262.

Czajkowski et al.,

- 1 Slawiak, M., van Beckhoven, J.R.C.M., Speksnijder, A.G.C.L., Czajkowski, R., Grabe, G., and van der
- 2 Wolf, J.M. 2009. Biochemical and genetical analysis reveal a new clade of biovar 3 *Dickeya*
- 3 spp. strains isolated from potato in Europe. Eur. J. Plant Pathol. 125:245-261.
- 4 Tanizawa, Y., Fujisawa, T., and Nakamura, Y. 2018. DFAST: a flexible prokaryotic genome
- 5 annotation pipeline for faster genome publication. Bioinformatics 34:1037-1039.
- 6 Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze,
- 7 A., Pruitt, K. D., Borodovsky, M., Ostell, J. 2016. NCBI prokaryotic genome annotation
- 8 pipeline. Nucleic Acids Res. 19:6614-6624
- 9 Toth, I.K., van der Wolf, J.M., Saddler, G., Lojkowska, E., Hélias, V., Pirhonen, M., Tsrer, L., and
- 10 Elphinstone, J.G. 2011. *Dickeya* species: An emerging problem for potato production in
- 11 Europe. Plant Pathol. 60:385-399.
- 12 Tsrer, L., Ben-Daniel, B., Chalupowicz, L., van der Wolf, J., Lebiush, S., Erlich, O., Dror, O., Barel,
- 13 V., Nijhuis, E., and Manulis-Sasson, S. 2013. Characterization of *Dickeya* strains isolated from
- 14 potato grown under hot-climate conditions. Plant Pathol. 62:1097-1105.
- 15 van der Wolf, J.M., Nijhuis, E.H., Kowalewska, M.J., Saddler, G.S., Parkinson, N., Elphinstone, J.G.,
- 16 Pritchard, L., Toth, I.K., Lojkowska, E., Potrykus, M., Waleron, M., de Vos, P., Cleenwerck, I.,
- 17 Pirhonen, M., Garlant, L., Hélias, V., Pothier, J.F., Pflüger, V., Duffy, B., Tsrer, L., and
- 18 Manulis, S. 2014. *Dickeya solani* sp. nov., a pectinolytic plant pathogenic bacterium isolated
- 19 from potato (*Solanum tuberosum*). Int. J. Syst. Evol. Microbiol. 64:768-774.

Czajkowski et al.,

1 **Table 1. General features of the *D. solani* strain IPO 2019 complete genome**

	IPO 2019 genome
Genome size (bp.)	4,919,542
GC content (%)	56.2
No. contigs	1
N50	1
Coding ratio (%)	85.8
No. total coding sequences (features)	4502
Average gene length (bp.)	1200
Presence of plasmids	no
rRNAs	22
tRNAs	73
CRISPRs	1
No. intact prophage genomes (according to PHASTER)	0
GenBank accession number	CP071062

2

3

4

5

6

Czajkowski et al.,

- 1 **Figure 1. Features of the *D. solani* strain IPO 2019 genome and its comparison with the IPO 2222**
- 2 **genome.** Circular plot of the IPO 2019 and IPO 2222 genomes (A) and core genome shared between *D.*
- 3 *solani* strains IPO 2019 and IPO 2222 (B). The comparative analyses were done using EDGAR – an
- 4 enhanced software platform for comparative gene content analyses (Blom et al., 2016).

5