Updated Genome Sequence and Annotation for the Full Genome of *Pseudomonas protegens* CHA0

Theo H. M. Smits,ᵃ Fabio Rezzonico,ᵃ David Frasson,ᵇ Pilar Vesga,ᶜ Jordan Vacheron,ᵈ Jochen Blom,ᵉ Joël F. Pothier,ᵃ Christoph Keel,ᵈ Monika Maurhofer,ᶜ Martin Sieversᵇ

ᵃEnvironmental Genomics and Systems Biology Research Group, Institute of Natural Resource Sciences, Zurich University of Applied Sciences (ZHAW), Wädenswil, Switzerland
ᵇMicrobiology and Molecular Biology Research Group, Institute of Chemistry and Biotechnology, Zurich University of Applied Sciences (ZHAW), Wädenswil, Switzerland
ᶜETH Zür, Plant Pathology, Institute of Integrative Biology, Zurich, Switzerland
ᵈDepartment of Fundamental Microbiology, Quartier UNIL-Sorge, University of Lausanne, Lausanne, Switzerland
ᵉBioinformatics and Systems Biology, Justus-Liebig-Universität, Giessen, Germany

ABSTRACT Minor differences in the previously obtained genome of *Pseudomonas protegens* CHA0 were detected after resequencing the strain. Based on this, the genome size slightly increased. Additionally, we performed a manual annotation of genes involved in biocontrol and insect pathogenicity. This annotation version will be the basis for upcoming genome studies.

*P*seudomonas protegens CHA0 has a long history as a biocontrol agent. Many features and the regulation thereof have been described for this strain that are involved in the process of biocontrol against fungal and oomycete plant pathogens (1–3). Additionally, the organism has been recognized as an effective insect pathogen that can kill larvae of different plant pest insects after oral infection by the bacterium (4, 5). In addition, several features contributing to insect pathogenicity were described already (2, 4, 6). While the genome sequence of this strain was published a few years ago (7), the pathogen control and insect pathogenicity determinants were only poorly included.

We have resequenced the genome of *P. protegens* CHA0 starting with genomic DNA from its accession number CCOS 2 at the Culture Collection of Switzerland (CCOS). The strain was grown in LB broth at 28°C for 1 day. Total DNA was extracted from the pure culture using the DNeasy blood and tissue kit (Qiagen, Hilden, Germany). Genomic library preparation and genome sequencing were outsourced to GATC Biotech, AG (Constance, Germany). Libraries were prepared using a SPRIworks fragment library system I (Beckman Coulter, Brea, CA), following the manufacturer's instructions. The TruSeq paired-end (PE) cluster kit v3-cBot-HS (Illumina, San Diego, CA) was used for cluster generation. Sequencing was performed on a HiSeq 2000 Illumina sequencer with 2×50-bp paired-end reads using the TruSeq SBS kit v3-HS (Illumina). A total of 97,357,690 quality-filtered reads were obtained from GATC, giving an approximate coverage of 700×. For *de novo* assembly using SeqMan NGen v12.2 (DNASTAR, Madison, WI) with standard settings, only 8,500,000 reads (55× coverage) were used. Repeated cycles of read mapping with the SeqMan NGen software and inspection in different subroutines of the Lasergene package (DNASTAR) yielded a complete genome of 6,868,156 bp with a G+C content of 63.39%. Based on a small region close to an rRNA region, the genome was 176 bp larger than the previous version (7). Additionally, few indels mainly in homopolymer regions were observed.

Annotation of the genome sequence was done in GenDB (8), with manual improvements. Comparative genomics was done in EDGAR v2.3 (9) using the parameters defined by Smits et al. (10). A comparison of the annotations indicated a difference

between the two genome sequences. The two genomes share 5,995 genes, whereas 142 and 120 singletons were observed for our version, respectively, to the previous one, with the majority representing hypothetical proteins. Furthermore, manual annotation of all known biocontrol features and insect pathogenicity factors was performed. This version of the genome is now suited for use in comparative genomics studies to study different aspects of biocontrol and insect pathogenicity.

Data availability. The genome sequence of *P. protegens* CHA0 (CCOS 2) was deposited at DDBJ/EMBL/GenBank under the BioProject number PRJEB28440 and accession number LS999205. The version described in this paper is version LS999205.1. Raw sequence reads (Illumina) have been deposited at the EMBL under accession number ERR3047520.

REFERENCES

1. Haas D, Défago G. 2005. Biological control of soil-borne pathogens by fluorescent pseudomonds. Nat Rev Microbiol 3:307–319. https://doi.org/10.1038/nrmicro1129.
2. Flury P, Aellen N, Ruffner B, Péchy-Tarr M, Fataar S, Metla Z, Dominguez-Ferreras A, Bloemberg G, Frey J, Goesmann A, Raaijmakers JM, Duffy B, Höfte M, Blom J, Smits THM, Keel C, Maurhofer M. 2016. Insect pathogenicity in plant-beneficial pseudomonads: phylogenetic distribution and comparative genomics. ISME J 10:2527–2542. https://doi.org/10.1038/ismej.2016.5.
3. Kupferschmied P, Maurhofer M, Keel C. 2013. Promise for plant pest control: root-associated pseudomonads with insecticidal activities. Front Plant Sci 4:287. https://doi.org/10.3389/fpls.2013.00287.
4. Péchy-Tarr M, Bruck DJ, Maurhofer M, Fischer E, Vogne C, Henkels MD, Donahue KM, Grunder J, Loper JE, Keel C. 2008. Molecular analysis of a novel gene cluster encoding an insect toxin in plant-associated strains of *Pseudomonas fluorescens*. Environ Microbiol 10:2368–2386. https://doi.org/10.1111/j.1462-2920.2008.01662.x.
5. Ruffner B, Péchy-Tarr M, Ryffel F, Hoegger P, Obrist C, Rindlisbacher A, Keel C, Maurhofer M. 2013. Oral insecticidal activity of plant-associated pseudomonads. Environ Microbiol 15:751–763. https://doi.org/10.1111/j.1462-2920.2012.02884.x.
6. Kupferschmied P, Péchy-Tarr M, Imperiali N, Maurhofer M, Keel C. 2014. Domain shuffling in a sensor protein contributed to the evolution of insect pathogenicity in plant-beneficial *Pseudomonas protegens*. PLoS Pathog 10:e1003964. https://doi.org/10.1371/journal.ppat.1003964.
7. Jousset A, Schuldes J, Keel C, Maurhofer M, Daniel R, Scheu S, Thuermer A. 2014. Full-genome sequence of the plant growth-promoting bacterium *Pseudomonas protegens* CHA0. Genome Announc 2:e00322-14. https://doi.org/10.1128/genomeA.00322-14.
8. Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R, Pühler A. 2003. GenDB–an open source genome annotation system for prokaryote genomes. Nucleic Acids Res 31:2187–2195. https://doi.org/10.1093/nar/gkg312.
9. Blom J, Kreis J, Spänig S, Juhre T, Bertelli C, Ernst C, Goesmann A. 2016. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. Nucleic Acids Res 44:W22–W28. https://doi.org/10.1093/nar/gkw255.
10. Smits THM, Rezzonico F, Kamber T, Blom J, Goesmann A, Frey JE, Duffy B. 2010. Complete genome sequence of the fire blight pathogen *Erwinia amylovora* CFBP 1430 and comparison to other *Erwinia* spp. Mol Plant Microbe Interact 23:384–393. https://doi.org/10.1094/MPMI-23-4-0384.