The Draft Genome Sequence of *Paenibacillus polymyxa* Strain CCI-25 Encompasses High Potential for Secondary Metabolite Production

Gajender Aleti, Livio Antonielli, Erika Corretto, Branislav Nikolić, Angela Sessitsch, Günter Brader

Bioresources Unit, Health and Environment Department, Austrian Institute of Technology (AIT), Tulln, Austria

We report here the draft genome sequence of *Paenibacillus polymyxa* strain CCI-25, which displays strong antifungal and antibacterial activities *in vitro*. The genome encompasses nonribosomal peptide synthetases predicted to encode a tridecaptin, polymyxin, fusaricidin, an iturin-like synthetase, a lantibiotic similar to paenicidin A, as well as a type 1 polyketide synthase.

Plant-associated *Paenibacillus polymyxa* strains are well noted for their production of a wide range of secondary metabolites (1–3), predominantly lipopeptides and polyketides involved in plant growth promotion and biocontrol of fungi (4–6). Here, we highlight the secondary metabolite capacity of *P. polymyxa* strain CCI-25 isolated from vermicompost. Both colonies and lipopeptide and polyketide crude extracts (7, 8) exhibited strong antimicrobial activity against *Escherichia coli* and fungi, including *Fusarium oxysporum* ACC01, *Botrytis cinerea* ofi 501-E (Austrian Institute of Technology [AIT] collection), and *Rhizoctonia solani* CBS101769, on plate assays.

To evaluate the molecular basis for secondary metabolite production, genomic DNA was isolated by phenol-chloroform extraction, and a library was prepared, according to the manufacturer's protocol, using the Nextera XT kit (Illumina, San Diego, CA). Library sequencing was performed using an Illumina MiSeq platform (MiSeq reagent kit version 3). Sequencing generated 2,213,773 paired-end reads with 124 ± 53-fold coverage after PhiX sequence removal by Bowtie2 (9). Adapter and quality trimming were performed using Trimmomatic-0.32 (10). Overlapping reads were merged with FLASH (11), and paired-end reads were assembled by SPAdes 3.1.0 (12). Quality control of mapping data was carried out by Qualimap 2.2 (13), and assembly quality was estimated by QUAST 3.2 (14). Assembly resulted in 117 contigs >1,000 bp, with an N_{50} size of 95,765 bp. The draft genome size is 5.61 Mb, with a G+C content of 44.95%. The identification of 40 highly conserved single-copy marker genes in the assembly by PhyloSift version 1.0.1 (15) indicated completeness of the genome and excluded contaminant sequences. Genomic BLAST showed similarities to *P. polymyxa* CR1. The NCBI Prokaryotic Genome Annotation Pipeline (PGAP) identified 5,146 genes, 4,953 coding sequences (CDSs), 15 complete 5S rRNAs, 30 partial 16S rRNAs, 37 partial 23S rRNAs (for a total of 15 putative rRNA operons), 107 tRNAs, 4 noncoding RNAs (ncRNAs), and 241 pseudogenes. The rRNAs were further confirmed by RNAmmer 1.2 (16). Prediction of secondary metabolite-encoding sequences was performed by antiSMASH (17).

The CCI-25 draft genome encompasses nonribosomal peptide synthetases with sequence similarities to published genes (1, 18–22), and the prediction includes the encoding of a tridecaptin with valine instead of isoleucine at the 13th position compared to *Paenibacillus terrae* NRRL B-30644 and fusaricidin C, and a polymyxin with leucine instead of phenylalanine at the 6th position compared to *P. polymyxa* M1. In addition, an iturin- and paenilarvin-like compound with altered monomer composition (D-Gly-D-Orn-D-Glu-D-nrp-L-nrp-L-Ile-L-Val) compared to the published metabolites from *Bacillus amyloliquefaciens* FZB42 (<58% identity) and *Paenibacillus larvae* DSM 25430 (<40% identity) (1, 23) has been predicted. CCI-25 contains a lantibiotic gene similar to paenicidin A and a predicted polyketide synthase with a different number of acyl carrier domains with 61% identity to bacillaene synthase from *B. amyloliquefaciens* FZB42 and 87% identity to *P. polymyxa* M-1 polyketide synthase (23). Given the fact that about 370 kb (6.6% of the total genome) is dedicated to secondary metabolite biosynthesis, CCI-25 has high potential to be exploited for medical or agricultural applications.

Nucleotide sequence accession number. The nucleotide sequences have been deposited at the DDBJ/EMBL/GenBank under the accession no. LTYJ00000000. The version described in this paper is the first version.

REFERENCES

1. **Niu B, Rueckert C, Blom J, Wang Q, Borriss R**. 2011. The genome of the plant growth-promoting rhizobacterium *Paenibacillus polymyxa* M-1 contains nine sites dedicated to nonribosomal synthesis of lipopeptides and polyketides. J Bacteriol **193**:5862–5863. http://dx.doi.org/10.1128/JB.05806-11.
2. **Catch JR, Jones TSG, Wilkinson S**. 1949. The chemistry of polymyxin A. Ann N Y Acad Sci **51**:917–923. http://dx.doi.org/10.1111/j.1749-6632.1949.tb27318.x.

3. He Z, Kisla D, Zhang L, Yuan C, Green-Church KB, Yousef AE. 2007. Isolation and identification of a *Paenibacillus polymyxa* strain that coproduces a novel lantibiotic and polymyxin. Appl Environ Microbiol **73:** 168–178. http://dx.doi.org/10.1128/AEM.02023-06.

4. Beatty PH, Jensen SE. 2002. *Paenibacillus polymyxa* produces fusaricidin-type antifungal antibiotics active against *Leptosphaeria maculans*, the causative agent of blackleg disease of canola. Can J Microbiol **48:**159–169. http://dx.doi.org/10.1139/w02-002.

5. Raza W, Yang X, Wu H, Wang Y, Xu Y, Shen Q. 2009. Isolation and characterisation of fusaricidin-type compound-producing strain of *Paenibacillus polymyxa* SQR-21 active against *Fusarium oxysporum* f. sp. *nevium*. Eur J Plant Pathol **125:**471–483. http://dx.doi.org/10.1007/s10658-009-9496-1.

6. Dijksterhuis J, Sanders M, Gorris LG, Smid EJ. 1999. Antibiosis plays a role in the context of direct interaction during antagonism of *Paenibacillus polymyxa* towards *Fusarium oxysporum*. J Appl Microbiol **86:**13–21. http://dx.doi.org/10.1046/j.1365-2672.1999.t01-1-00600.x.

7. Aktuganov G, Jokela J, Kivelä H, Khalikova E, Melentjev A, Galimzianova N, Kuzmina L, Kouvonen P, Himanen JP, Susi P, Korpela T. 2014. Isolation and identification of cyclic lipopeptides from *Paenibacillus ehimensis*, strain IB-X-b. J Chromatogr B **973:**9–16. http://dx.doi.org/10.1016/j.jchromb.2014.09.042.

8. Smyth TJP, Perfumo A, McClean S, Marchant R, Banat IM. 2010. Isolation and analysis of lipopeptides and high molecular weight biosurfactants, p 3687–3704. *In* Timmis KN (ed), Handbook of hydrocarbon and lipid microbiology. Springer, Heidelberg, Germany.

9. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9:**357–359. http://dx.doi.org/10.1038/nmeth.1923.

10. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30:**2114–2120.

11. Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics **27:**2957–2963. http://dx.doi.org/10.1093/bioinformatics/btr507.

12. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single cell sequencing. J Comput Biol **19:**455–477. http://dx.doi.org/10.1089/cmb.2012.0021.

13. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, M, Götz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. 2012. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics **28:** 2678–2679. http://dx.doi.org/10.1093/bioinformatics/bts503.

14. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics **29:**1072–1075. http://dx.doi.org/10.1093/bioinformatics/btt086.

15. Darling AE, Jospin G, Lowe E, Matsen FA IV, Bik HM, Eisen JA. 2014. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ **2:** http://dx.doi.org/10.7717/peerj.243.

16. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res **35:**3100–3108. http://dx.doi.org/10.1093/nar/gkm160.

17. Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T. 2013. antiSMASH 2.0 a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res **41:** W204–W212 http://dx.doi.org/10.1093/nar/gkt449.

18. Choi SK, Park SY, Kim R, Kim SB, Lee CH, Kim JF, Park SH. 2009. Identification of a polymyxin synthetase gene cluster of *Paenibacillus polymyxa* and heterologous expression of the gene in *Bacillus subtilis*. J Bacteriol **191:**3350–3358. http://dx.doi.org/10.1128/JB.01728-08.

19. Choi SK, Park SY, Kim R, Lee CH, Kim JF, Park SH. 2008. Identification and functional analysis of the fusaricidin biosynthetic gene of *Paenibacillus polymyxa* E681. Biochem Biophys Res Commun **365:**89–95. http://dx.doi.org/10.1016/j.bbrc.2007.10.147.

20. Lohans CT, Huang Z, van Belkum MJ, Giroud M, Sit CS, Steels EM, Zheng J, Whittal RM, McMullen LM, Vederas JC. 2012. Structural characterization of the highly cyclized lantibiotic paenicidin A via a partial desulfurization/reduction strategy. J Am Chem Soc **134:**19540–19543. http://dx.doi.org/10.1021/ja3089229.

21. Aleti G, Sessitsch A, Brader G. 2015. Genome mining: prediction of lipopeptides and polyketides from *Bacillus* and related *Firmicutes*. Comput Struct Biotechnol J **13:**192–203.

22. Lohans CT, Van Belkum MJ, Cochrane SA, Huang Z, Sit CS, McMullen LM, Vederas JC. 2014. Biochemical, structural, and genetic characterization of tridecaptin A$_1$, an antagonist of *Campylobacter jejuni*. Chembiochem **15:**243–249. http://dx.doi.org/10.1002/cbic.201300595.

23. Sood S, Steinmetz H, Beims H, Mohr KI, Stadler M, Djukic M, von der Ohe W, Steinert M, Daniel R, Müller R. 2014. Paenilarvins: iturin family lipopeptides from the honey bee pathogen *Paenibacillus larvae*. Chembiochem **15:**1947–1955. http://dx.doi.org/10.1002/cbic.201402139.