Genome Sequence of *Streptomyces* sp. Strain GQFP Isolated from Soil Near the Roots of Pharmaceutical Plant *Elaeagnus pungens*

Jie Zhu[a]

aState Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences, School of Life Sciences & Biotechnology, Shanghai Jiao Tong University, Shanghai, China

ABSTRACT   Here, we report the complete and linear genome sequence of *Streptomyces* sp. strain GQFP, isolated from the soil microbiome of the root of pharmaceutical plant *Elaeagnus pungens*. The strain contained one 8,306,813-bp chromosome, with a GC content of 69.8%.

Microorganisms often play diverse and vital roles in the growth of pharmaceutical plant hosts, such as stress tolerance, resistance to pathogens, promotion of plant growth, and uptake of a diversity nutrient substances (1–4). The genomic information of individual members of microorganisms facilitated our understanding toward coassociations between individual symbiotic microorganisms and pharmaceutical plant hosts.

Here, we have isolated the microbes from the soil sticking to the root of *Elaeagnus pungens*, whose leaf is one of the antiasthmatic traditional Chinese medicines (5). A strain of the genus *Streptomyces* was isolated from the soil sample. The strain was cultured in ISP2 agar plates at 30°C for 7 days. The ISP2 agar plates contained 4 g/L yeast extract, 10 g/L malt extract, 4 g/L glucose (pH 7.5), and 20 g/L agar. The strain was then transferred into trypticase soy broth with yeast (TSBY) liquid medium in 250-mL flasks at 30°C and 220 rpm for 5 days. The TSBY liquid medium contained 30 g/L tryptone soya broth, 5 g/L yeast extract, and 103 g/L sucrose (pH 7.5). The mycelia were harvested through centrifugation and washed twice in distilled H_2O. After centrifugation, the mycelia were stored in −80°C for genome sequencing.

Genomic DNA was extracted from the mycelia using the cetyltrimethylammonium bromide method (6). Sequencing of genomic DNA was conducted at Personalbio (Shanghai Personal Biotechnology Co., Ltd., Shanghai, China). To construct libraries with DNA fragments of different lengths, a whole-genome shotgun strategy was performed first. Genomic sequencing was conducted using both the Pacific Biosciences (PacBio) and Illumina HiSeq platforms. The PacBio sequencing platform produced 104,162 sequences and 905,073,920 bp of high-quality bases. PacBio sequences were assembled with HGAP4 (7) and CANU v. 1.6 (8). For Illumina sequencing, the genomic DNA was randomly fragmented to 400 bp. Through end polishing, A-tailing, and ligation with adapters, the library was sequenced on an Illumina HiSeq platform, generating 2×150-bp paired-end reads. Raw data were further trimmed using AdapterRemoval v. 2.1.7 (9) to remove adapters and polished using SOAPec v. 2.0 (10). The Illumina HiSeq platform generated 24,032,166 high-quality reads. Subsequently, the high-quality reads from the Illumina HiSeq platform were used to polish the assembly generated by the PacBio platform with Pilon v. 1.22 (11). The total genome was determined to comprise 8,306,813 bp with a GC content of 69.8%.

To obtain more information of the genome, further analysis was performed. A total of 62 tRNAs were predicted by tRNAscan-SE v. 1.3.1 (12). The protein-coding gene prediction was conducted using GeneMarkS v. 4.32 (13), thereby resulting in a total of

TABLE 1 Genome features of *Streptomyces* sp. GQFP isolated from soil near the roots of pharmaceutical plant *Elaeagnus pungens*

Parameter	Feature
GenBank accession no. for genome assembly	CP089312
No. of high-quality paired-end Illumina reads	24,032,166
Total sequence no. for the PacBio sequencing	104,162
Maximum sequence length of the PacBio sequencing (bp)	43,568
Minimum sequence length of the PacBio sequencing (bp)	200
N_{50} length of the PacBio sequencing (bp)	11,173
N_{90} length of the PacBio sequencing (bp)	5,303
Assembly size (bp)	8,306,813
G+C content (%)	69.8
Total gene length (bp)	7,151,034
Total no. of genes	7,479
Avg gene length (bp)	969.24
Gene percentage of genome (%)	86.09
No. of tRNA genes	62
Secondary metabolites gene clusters	24

7,479 protein-coding genes with the average length of 969.24 bp. Further functional analyses of these genes were performed based on different databases, including the NCBI nonredundant (NR) database (14), Nonsupervised Orthologous Groups (NOG) (15), Kyoto Encyclopedia of Genes and Genomes (KEGG) (16), Swiss-Prot, and Gene Ontology (GO) (17). To gain insight into the secondary metabolism of the strain, antiSMASH v. 6.0 (18) was used to analyze the secondary metabolite biosynthetic gene clusters with the "bacterial taxon" option. The analyses revealed a total of 24 putative secondary metabolite biosynthetic gene clusters, which could generate siderophores (19–21), lanthipeptides (22, 23), and terpenes (24–26). Default parameters were used for all software.

The complete genome sequence revealed biological activities and potential secondary metabolites of the strain, which could provide valuable understanding toward the coassociation between the strain with the plant host.

Data availability. The genome sequence has been deposited in GenBank under the accession number CP089312, BioSample accession number SAMN24594572, and BioProject accession number PRJNA794000. The raw sequencing data were deposited in the SRA database under the accession number SRR18249970 and SRR18249971. The features of the genome are summarized in Table 1.

REFERENCES

1. Pieterse CM, Zamioudis C, Berendsen RL, Weller DM, Van Wees SC, Bakker PA. 2014. Induced systemic resistance by beneficial microbes. Annu Rev Phytopathol 52:347–375. https://doi.org/10.1146/annurev-phyto-082712-102340.

2. Trivedi P, Trivedi C, Grinyer J, Anderson IC, Singh BK. 2016. Harnessing host-vector microbiome for sustainable plant disease management of phloem-limited bacteria. Front Plant Sci 7:1423. https://doi.org/10.3389/fpls.2016.01423.

3. Backer R, Rokem JS, Ilangumaran G, Lamont J, Praslickova D, Ricci E, Subramanian S, Smith DL. 2018. Plant growth-promoting rhizobacteria: context, mechanisms of action, and roadmap to commercialization of biostimulants for sustainable agriculture. Front Plant Sci 9:1473. https://doi.org/10.3389/fpls.2018.01473.

4. Richardson AE, Simpson RJ. 2011. Soil microorganisms mediating phosphorus availability update on microbial phosphorus. Plant Physiol 156:989–996. https://doi.org/10.1104/pp.111.175448.

5. Ge Y, Liu J, Su D. 2009. *In vivo* evaluation of the anti-asthmatic, antitussive and expectorant activities of extract and fractions from *Elaeagnus pungens* leaf. J Ethnopharmacol 126:538–542. https://doi.org/10.1016/j.jep.2009.08.042.

6. Möller EM, Bahnweg G, Sandermann H, Geiger HH. 1992. A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. Nucleic Acids Res 20:6115–6116. https://doi.org/10.1093/nar/20.22.6115.

7. Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GR, Delledonne M, Luo C, Ecker JR, Cantu D, Rank DR, Schatz MC. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods 13:1050–1054. https://doi.org/10.1038/nmeth.4035.

8. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.

9. Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes 9:88. https://doi.org/10.1186/s13104-016-1900-2.

10. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. Gigascience 1:18. https://doi.org/10.1186/2047-217X-1-18.

11. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated

tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.

12. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964. https://doi.org/10.1093/nar/25.5.955.

13. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29:2607–2618. https://doi.org/10.1093/nar/29.12.2607.

14. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.

15. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. Mol Biol Evol 34:2115–2122. https://doi.org/10.1093/molbev/msx148.

16. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 35:W182–W185. https://doi.org/10.1093/nar/gkm321.

17. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang HY, Dosztányi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto SC, Wu CH, Xenarios I, Yeh LS, Young SY, Mitchell AL. 2017. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res 45:D190–D199. https://doi.org/10.1093/nar/gkw1107.

18. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res 49:W29–W35. https://doi.org/10.1093/nar/gkab335.

19. Tanabe T, Funahashi T, Nakao H, Miyoshi S, Shinoda S, Yamamoto S. 2003. Identification and characterization of genes required for biosynthesis and transport of the siderophore vibrioferrin in Vibrio parahaemolyticus. J Bacteriol 185:6938–6949. https://doi.org/10.1128/JB.185.23.6938-6949.2003.

20. Zhu D, Liu J, Han R, Shen G, Long Q, Wei X, Liu D. 2014. Identification and characterization of ectoine biosynthesis genes and heterologous expression of the ectABC gene cluster from Halomonas sp. QHL1, a moderately halophilic bacterium isolated from Qinghai Lake. J Microbiol 52:139–147. https://doi.org/10.1007/s12275-014-3389-5.

21. Funabashi M, Funa N, Horinouchi S. 2008. Phenolic lipids synthesized by type III polyketide synthase confer penicillin resistance on Streptomyces griseus. J Biol Chem 283:13983–13991. https://doi.org/10.1074/jbc.M710461200.

22. Mohimani H, Kersten RD, Liu WT, Wang M, Purvine SO, Wu S, Brewer HM, Pasa-Tolic L, Bandeira N, Moore BS, Pevzner PA, Dorrestein PC. 2014. Automated genome mining of ribosomal peptide natural products. ACS Chem Biol 9:1545–1551. https://doi.org/10.1021/cb500199h.

23. Baindara P, Chaudhry V, Mittal G, Liao LM, Matos CO, Khatri N, Franco OL, Patil PB, Korpole S. 2016. Characterization of the antimicrobial peptide penisin, a class Ia novel lantibiotic from Paenibacillus sp. strain A3. Antimicrob Agents Chemother 60:580–591. https://doi.org/10.1128/AAC.01813-15.

24. Bentley SD, Chater KF, Cerdeño-Tárraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. 2002. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 417:141–147. https://doi.org/10.1038/417141a.

25. Braga D, Hoffmeister D, Nett M. 2016. A non-canonical peptide synthetase adenylates 3-methyl-2-oxovaleric acid for auriculamide biosynthesis. Beilstein J Org Chem 12:2766–2770. https://doi.org/10.3762/bjoc.12.274.

26. Oliynyk M, Samborskyy M, Lester JB, Mironenko T, Scott N, Dickens S, Haydock SF, Leadlay PF. 2007. Complete genome sequence of the erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338. Nat Biotechnol 25:447–453. https://doi.org/10.1038/nbt1297.