The Draft Genome Sequence of *Xanthomonas* sp. Strain Mitacek01 Expands the Pangenome of a Genus of Plant Pathogens

M. B. Couger, a Radwa A. Hanafy, a Rachel M. Mitacek, a Connie Budd, a Donald P. French, b Wouter D. Hoff, a Mostafa S. Elshahed, a Noha H. Youssef a

Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, Oklahoma, USA; Department of Integrative Biology, Oklahoma State University, Stillwater, Oklahoma, USA

We report the draft genome sequence of *Xanthomonas* sp. strain Mitacek01, isolated from an indoor environment vending machine surface with frequent human use in Stillwater, Oklahoma, USA, as part of the Student-Initiated Microbial Discovery project. The genome has a total size of 3,617,426 bp and a contig N₅₀ of 1,906,967 bp.

*Xanthomonas* spp. infect over 300 host plants, including many important crops such as rice (1). Such infections can result in reduction in crop yields or outright crop failures and hence could have a major impact on economic development and global food supply (2). Genomic analysis of strains belonging to the genus *Xanthomonas* can contribute to understanding the molecular mechanisms of pathogenesis and subsequently reduce the occurrence and/or mitigate the severity of such infections (3, 4).

*Xanthomonas* sp. strain Mitacek01 was isolated from an indoor environment vending machine surface and was sequenced on the Illumina MiSeq platform at the University of Georgia Genomics Facility using 2 × 300 paired-end chemistry. Generated reads were quality filtered with standard Illumina filtering settings resulting in 1,511,702 (453.5 Mb) quality sequences. All quality-filtered reads were assembled using the short read de Bruijn graph assembly (5) program Velvet (6). Velvet assembly run-time settings used were a k-mer value of 101 bp and a minimum contig coverage value of 7×. Gene models were created using the prokaryotic gene calling software package Prodigal (7). The Velvet assembly had a total size of 3,617,426 bp and an N₅₀ of 1,906,967 bp. The largest assembled contig was 1,906,967 bp, with a GC content of 68.5%. A total of 3,212 gene models were predicted. Translated protein sequences were functionally annotated using a combination of NCBI BLAST C + + homology search (8) and HMMER version 3.0 hmmscan (9) against the PFAM 26.0 database (10).

16S rRNA gene-based comparisons to *Xanthomonas* genomes publicly available in the GenBank database (n = 302,955,543, October 2015) revealed that strain Mitacek01 was closely related (97.0% similarity) to *Xanthomonas oryzae* pv. oryzae strain YM15, a causative agent of bacterial leaf streak in rice (11). *Xanthomonas campestris* strain 17, a phytopathogen capable of infecting a wide range of plants (12), and 21 different strains of *Xanthomonas citri*, causative agents of citrus canker (13). Despite close 16S rRNA gene sequence similarity to multiple *Xanthomonas* strains, BLAST analysis identified 393 genes (12.2%) within the Mitacek01 genome with no sequence homology (e value <10⁻⁰⁵) to any of the genes in the *Xanthomonas* pan genome (genomes = 32; protein sequences = 298,975; October 2015). The majority of these genes were hypothetical (147/393, 37.4%) and conserved hypothetical (186/395, 47.3%) proteins. In addition, genes encoding β-lactamase, glucolactone synthesis, and phenol metabolism were identified, further expanding the pan-metabolic repertoire of the genus *Xanthomonas*.

To identify putative virulence factors we used signalP (14) to identify secreted proteins and compared gene models to the pathogen-host interactions (PHI) (15) database to identify genes that had been previously established to affect *Xanthomonas* pathogenesis. A total of 497 secreted genes were identified, and 29 genes in the genome were present in the PHI database.

In conclusion, this initial genomic analysis of strain Mitacek01 highlights a high level of intergenomic diversity within the genus *Xanthomonas*, supporting previous findings for this genus (16, 17) and contributing to the available genomic resources for the study of an economically relevant group of phytopathogens.

**Nucleotide sequence accession number.** The draft genome of *Xanthomonas* sp. strain Mitacek01 has been deposited in GenBank under the accession number LK17000000.

**ACKNOWLEDGMENTS**

This is draft genome number 2 in the Student-Initiated Microbial Discovery (SIMD) project, which is supported in part by a grant from the Howard Hughes Medical Institute through the Science Education Program. W.D.H. is supported by NSF grants MCB-1051590, MRI-1338097, and CHE-1412500.

**REFERENCES**

1. Niño-liu DO, Ronald PC, Bogdanove AJ. 2006. *Xanthomonas oryzae* pathovars: model pathogens of a model crop. Mol Plant Pathol 7:303–324. http://dx.doi.org/10.1111/j.1364-3703.2006.00344.x.
2. Strange RN, Scott PR. 2005. Plant disease: a threat to global food security. Annu Rev Phytopathol 43:83–116. http://dx.doi.org/10.1146/annurev.phyto.43.113004.133839.
3. Da Silva ACR, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB, Monteiro-Vitorelo CB, Van Sluys MAV, Almeida NF, Alves LMC, do Amaral AM, Bertolini MC, Camargo LEA, Lamarotte G, Cannavan F, Cardozo J, Chamberbo F, Ciapina LP, Gicarelli RMB, Coutinho LL, Cursino-Santos JR, El-Dorry H, Faria JB, Ferreira AL, Ferreira RC, Ferro MI, Formighieri EF, Franco MC, Greggio CC, Gruber A, Katsuyama AM, Kishi LT, Leite RP, Lemos EG, Lemos MV, Locali EC, Machado MA, Madeira AM, Martinez-Rossi NM, Martins EC, Meidanis J, Menck CF, Miyaki CY, Moon DH, Moreira LM, Novo MT, Okura VK, Oliveira MC, Oliveira VR, Pereira HA. 2002. Comparison of the genomes of two Xanthomonas pathogens with differing host specificities. Nature 417:459–463. http://dx.doi.org/10.1038/417459a.

4. Qian W, Jia Y, Ren SX, He YQ, Feng JX, Lu LF, Sun Q, Ying G, Tang DJ, Tang H, Wu W, Hao P, Wang L, Jiang BL, Zeng S, Gu WY, Lu G, Rong L, Tian Y, Yao Z, Fu G, Chen B, Fang R, Qiang B, Chen Z, Zhao GP, Tang JL, He C. 2005. Comparative and functional genomic analyses of the pathogenicity of phytopathogen Xanthomonas campestris pv. campesiris. Genome Res 15:757–767. http://dx.doi.org/10.1101/gr.3378705.

5. Compeau PEC, Pevzner PA, Tesler G. 2011. How to apply de Bruijn graphs to genome assembly. Nat Biotechnol 29:987–991. http://dx.doi.org/10.1038/nbt.2023.

6. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829. http://dx.doi.org/10.1101/gr.074492.107.

7. Hyatt D, Chen G, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. http://dx.doi.org/10.1186/1471-2105-11-119.

8. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. http://dx.doi.org/10.1186/1471-2105-10-421.

9. Johnson LS, Eddy SR, Portugaly E. 2010. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics 11:431. http://dx.doi.org/10.1186/1471-2105-11-431.

10. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherton K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res 42:D222–D230. http://dx.doi.org/10.1093/nar/gkt1223.

11. Lee B-, Park YJ, Park DS, Kang HW, Kim JG, Song ES, Park IC, Yoon UB, Hahn JH, Koo BS, Lee GB, Kim H, Park HS, Yoon KO, Kim JH, Jung CH, Koh NH, Seo JS, Go SJ. 2005. The genome sequence of Xanthomonas oryzae pv. oryzae pv. oryzae KACC10331, the bacterial blight pathogen of rice. Nucleic Acids Res 33:577–586. http://dx.doi.org/10.1093/nar/gki206.

12. Leyns F, De Cleene M, Swings J, De Ley J. 1984. The host range of the genus Xanthomonas. Bot Rev 50:308–356. http://dx.doi.org/10.1007/BF02862635.

13. Jalan N, Kumar D, Yu F, Jones JB, Graham JH, Wang N. 2013. Complete genome sequence of Xanthomonas citri subsp. citri strain A-W12879, a restricted-host-range citrus canker-causing bacterium. Genome Announc 1(3):e00235–13. http://dx.doi.org/10.1128/genomeA.00235–13.

14. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786. http://dx.doi.org/10.1038/nmeth.1701.

15. Urban M, Pant R, Raghunath A, Jagadeesh AG, Pedro H, Hammond-Kosack KE. 2015. The pathogen-host interactions database (PHI-base): additions and future developments. Nucleic Acids Res 43:D645–D655. http://dx.doi.org/10.1093/nar/gku1165.

16. Triplett LR, Hamilton JP, Buell CR, Tisserat NA, Verdier V, Zink F, Winning I, Cottyn B, Detemmerman L, Dao S, Ouedraogo L, Sarra S, Teke C, Poussier S, Coellar R, Triplett L, Koita O, Koebnik R, Leach J, Szurek B, Maes M, Verdier V. 2014. Analysis of Xanthomonas oryzae pv. oryzae isolates from rice grown in the United States reveals substantial divergence from known X. oryzae pathovars. Appl Environ Microbiol 77:3930–3937. http://dx.doi.org/10.1128/AEM.00028-11.

17. Wonni I, Cottyn B, Detemmerman L, Dao S, Ouedraogo L, Sarra S, Teke C, Poussier S, Coellar R, Triplett L, Koita O, Koebnik R, Leach J, Szurek B, Maes M, Verdier V. 2014. Analysis of Xanthomonas oryzae pv. oryzae isolates from rice grown in the United States reveals substantial divergence from known X. oryzae pathovars. Appl Environ Microbiol 77:3930–3937. http://dx.doi.org/10.1128/AEM.00028-11.