AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Complete Genome Sequence of the African Strain AXO1947 of *Xanthomonas oryzae* pv. oryzae

J. C. Huguet-Tapia,[a] Z. Peng,[b] B. Yang,[c] Z. Yin,[d] S. Liu,[b] F. F. White[a]

Department of Plant Pathology, University of Florida, Gainesville, Florida, USA[a]; Department of Plant Pathology, Kansas State University, Manhattan, Kansas, USA[b]; Department of Genetics, Development and Cell Biology, Iowa State University, Ames, Iowa, USA[c]; Temasek Life Sciences Laboratory, National University of Singapore, Singapore, Republic of Singapore[d]

***Xanthomonas oryzae* pv. oryzae is the etiological agent of bacterial rice blight. Three distinct clades of *X. oryzae* pv. oryzae are known. We present the complete annotated genome of the African clade strain AXO194 using long-read single-molecule PacBio sequencing technology. The genome comprises a single chromosome of 4,674,975 bp and encodes for nine transcriptional activator-like (TAL) effectors. The approach and data presented in this announcement provide information for complex bacterial genome organization and the discovery of new virulence effectors, and they facilitate target characterization of TAL effectors.**

Received 21 December 2015 **Accepted** 22 December 2015 **Published** 11 February 2016

Citation Huguet-Tapia JC, Peng Z, Yang B, Yin Z, Liu S, White FF. 2016. Complete genome sequence of the African strain AXO1947 of *Xanthomonas oryzae* pv. oryzae. Genome Announc 4(1):e01730-15. doi:10.1128/genomeA.01730-15.

Address correspondence to F. F. White, ffwhite@ufl.edu.

Bacterial blight of rice, caused by the Gram-negative bacterium *Xanthomonas oryzae* pv. oryzae, is the most important bacterial disease of rice and, possibly, the most important bacterial disease in terms of agronomic impact. Three clades of related bacteria have been classified as *X. oryzae* pv. oryzae and complete genome sequences have been generated for Asian-related strains (1–3). Draft sequences are available in GenBank for several African linage strains, while one draft sequence is available for a North American strain (4). Strain AXO1947 (CFBP 1947) is of African origin (5)

Whole-genome sequencing of strain *X. oryzae* pv. oryzae AXO1947 was performed on the four single-molecule real-time cells of PacBio RS II. A total of 235,510 long circular reads (213-fold coverage) with an average length of 4,524 bp and 80% estimated accuracy were used as input for the self-correction and an assembly pipeline implemented in the PBcR package (6). The resulting contig was corrected by remapping reads with pbalign version 0.2, and a consensus was obtained with Quiver (7, 8). Final correction was performed for small insertions or deletions that were identified by mapping Illumina 2 × 300-bp MiSeq reads using Bowtie2 (9) with an average coverage of 88×. Error calling and generation of the final corrected assembly was conducted using the variant detection tool Pilon (10). Prediction and annotation of coding sequences were conducted with NCBI Prokaryotic Genome Annotation Pipeline (PGAAP) (11). Whole chromosome alignment and comparison was conducted with MUMmer3 (12).

The genome of *X. oryzae* pv. oryzae AXO1947 consists of a chromosome of 4,674,975 bp with 63.89% GC content. Preliminary draft annotation with PGAAP indicates that the genome is predicted to contain 3,706 genes encoding for proteins and 54 for tRNAs. Reciprocal average nucleotide identity (ANI index) (13) between AXO1947 and two other previously sequenced strains, POX99^{A} and MAFF311018, reveals that AXO1947 shares an average of 97% identity with both strains, while PXO99^{A} and MAFF311018 share 99% identity. Furthermore, AXO1947 chromosome alignments at the nucleotide level with genomes of strains PXO99^{A} and MAFF311018 reveal large chromosome inversions and rearrangements. Data mining indicates that AXO1947 contains nine TAL (transcriptional activator-like) effector genes. One TAL effector gene is identical to TalC, which induces the rice nodulin-3 *Os11N3* (*OsSWEET14*) (14).

Nucleotide sequence accession number. The final closed version of the *X. oryzae* pv. oryzae strain AXO1947 genome has been deposited at GenBank under the accession number CP013666.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation research award IOS-1238189 (S.L., F.F.W., B.Y.) and the Grand Challenges Explorations phase II grant OPP1068450 from the Bill & Melinda Gates Foundation (S.L., F.F.W., B.Y.). PacBio and Illumina sequencing services were provided by the Yale Center for Genome Analysis at Yale University and the Integrated Genomic Facility at Kansas State University, respectively.

FUNDING INFORMATION

This work was supported by National Science Foundation research award IOS-1238189 (S.L., F.F.W., B.Y.) and the Grand Challenges Explorations Phase II grant OPP1068450 from the Bill & Melinda Gates Foundation (S.L., F.F.W., B.Y.).

REFERENCES

1. **Ochiai H, Inoue Y, Takeya M, Sasaki A, Kaku H.** 2005. Genome sequence of *Xanthomonas oryzae* pv. *oryzae* suggests contribution of large numbers of effector genes and insertion sequences to its race diversity. JARQ **39:**275–287. http://dx.doi.org/10.6090/jarq.39.275.
2. **Lee B-M, Park Y-J, Park D-S, Kang H-W, Kim J-G, Song E-S, Park I-C, Yoon U-H, Hahn J-H, Koo B-S, Lee G-B, Kim H, Park H-S, Yoon K-O, Kim J-H, Jung C-H, Koh N-H, Seo J-S, Go S-J.** 2005. The genome

sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice. Nucleic Acids Res **33:**577–586. http://dx.doi.org/10.1093/nar/gki206.

3. **Salzberg SL, Sommer DD, Schatz MC, Phillippy AM, Rabinowicz PD, Tsuge S, Furutani A, Ochiai H, Delcher AL, Kelley D, Madupu R, Puiu D, Radune D, Shumway M, Trapnell C, Aparna G, Jha G, Pandey A, Patil PB, Ishihara H.** 2008. Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99A. BMC Genomics **9:**204. http://dx.doi.org/10.1186/1471-2164-9-204.
4. **Triplett LR, Hamilton JP, Buell CR, Tisserat NA, Verdier V, Zink F, Leach JE.** 2011. Genomic analysis of *Xanthomonas 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.
5. **Gu K, Tian D, Yang F, Wu L, Sreekala C, Wang D, Wang G-L, Yin Z.** 2004. High-resolution genetic mapping of *Xa27*(*t*), a new bacterial blight resistance gene in rice, *Oryza sativa* L. Theor Appl Genet **108:**800–807. http://dx.doi.org/10.1007/s00122-003-1491-x.
6. **Berlin K, Koren S, Chin C-S, Drake JP, Landolin JM, Phillippy AM.** 2015. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat Biotechnol **33:**623–630. http://dx.doi.org/10.1038/nbt.3238.
7. **Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J.** 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods **10:**563–569. http://dx.doi.org/10.1038/nmeth.2474.
8. **Chaisson MJ, Tesler G.** 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics **13:**238. http://dx.doi.org/10.1186/1471-2105-13-238.
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. **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. http://dx.doi.org/10.1371/journal.pone.0112963.
11. **Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, Kodira CD, Kyrpides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O.** 2008. Toward an online repository of Standard operating procedures (SOPs) for (meta)genomic annotation. Omics **12:**137–141. http://dx.doi.org/10.1089/omi.2008.0017.
12. **Delcher AL, Phillippy A, Carlton J, Salzberg SL.** 2002. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res **30:**2478–2483. http://dx.doi.org/10.1093/nar/30.11.2478.
13. **Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM.** 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol **57:**81–91. http://dx.doi.org/10.1099/ijs.0.64483-0.
14. **Yu Y, Streubel J, Balzergue S, Champion A, Boch J, Koebnik R, Feng J, Verdier V, Szurek B.** 2011. Colonization of rice leaf blades by an African strain of *Xanthomonas oryzae* pv. *oryzae* depends on a new TAL effector that induces the rice nodulin-3 Os11N3 gene. Mol Plant Microbe Interact **24:**1102–1113. http://dx.doi.org/10.1094/MPMI-11-10-0254.