Draft Genome Sequence of *Fusarium fujikuroi*, a Fungus Adapted to the Fuel Environment

Osman Radwan,a Thusitha S. Gunasekera,a,  b Oscar N. Ruizb

a Environmental Microbiology Group, University of Dayton Research Institute, Dayton, Ohio, USA
b Fuels and Energy Branch, Aerospace Systems Directorate, Air Force Research Laboratory, Wright-Patterson AFB, Ohio, USA

ABSTRACT  *Fusarium fujikuroi* isolate FUS01 is highly adapted to grow in jet fuel with predicted genes involved in hydrocarbon catabolism and carbon assimilation. The draft genome size is estimated at 49 Mb containing 18,578 proteins with high similarity to that of *F. fujikuroi* isolate B14.

Although the genomes of several fuel-adapted bacteria have been reported (1–3), far fewer reports have described the genomes of fuel-adapted fungi (4). The ability of the filamentous fungus *Fusarium fujikuroi* isolate FUS01 to adapt and grow in fuel-containing environments enthused us to sequence its genome.

A fungal isolate was recovered from a sample of Jet A fuel and identified as *F. fujikuroi* isolate FUS01 based on morphological characteristics and high similarity (99%) of the 18S rRNA gene to that of *F. fujikuroi*. Using a whole-genome shotgun approach, TruSeq paired-end libraries were generated and sequenced on a HiSeq 2000 platform, resulting in 49,249,613 paired-end reads with a read length of 100 bp (~4.92 Gb). The raw sequences were trimmed using Trimmomatic (5), and reads shorter than 40 bp were discarded. The sequence reads were *de novo* assembled with SPAdes software (6). The resulting 49,057,912-bp (100× sequence coverage) draft assembly comprises 881 contigs greater than 500 bp and has an N50 of 838,232 bp and an L50 of 16 contigs. The G+C content of the assembled genome is 47.34%, which is less than the 48.3% G+C content of *F. fujikuroi* isolate B14 (GenBank accession number ANFV00000000), the causal agent of the bakanae disease of rice (7). CEGMA (8) was then run on the draft assembly and identified 243 out of 248 ultraconserved eukaryotic genes in *F. fujikuroi* isolate FUS01 (97.98%). After masking repetitive sequences (2.80%) using the RepeatMasker program (9), the masked genome was used for gene prediction by AUGUSTUS version 2.5.5 (10) with an option set for *F. graminearum* parameters, resulting in 18,578 protein-coding genes.

Based on a BLASTp search against the UniRef90 and UniProt databases, significant matches (E value, 1E−5) were identified for 16,572 proteins; 15,427 hits were derived from *F. fujikuroi* isolate B14, while tRNAscan-SE (11) identified 299 tRNAs. A total of 967 fungal enzymes involved in carbohydrate metabolism and assimilation were identified (E value, 1E−4) using the Carbohydrate-Active Enzyme (CAZy) database (12). These enzymes include polysaccharide lyases (n = 25), glycosyl transferases (n = 118), glycoside hydrolases (n = 396), carbohydrate esterases (n = 193), carbohydrate-binding modules (n = 106), and auxiliary activities (n = 129).

The KEGG database and BLASTp research identified important proteins related to biofilm formation (agglutinin-like proteins, n = 17) and efflux of toxic substances (efflux pumps, n = 74), which are well studied in bacteria (2), suggesting a common mechanism of hydrocarbon adaptation between bacteria and fungi. Similarly, proteins involved in carbon metabolism (n = 81) and degradation of aromatic compounds (n = 9), such as benzoate 4-monoxygenase and cyclohexanone monoxygenase, are found. In agree-
ment with the capability of FUS01 to adapt and grow in fuel, we found genes involved in the oxidation of alkanes. These genes include the AlkB-related alkane hydroxylases CYP153 and CYP505 and the n-alkane-inducible cytochrome P-450 (13). Information from the current genome will facilitate an understanding of the mechanisms underlying fungal adaptation to, growth in, and degradation of hydrocarbon fuels.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number NCQQ00000000. The version described in this paper is the second version, NCQQ02000000.

ACKNOWLEDGMENTS

This material is based on research sponsored by AFRL/RQTF under agreement FA8650-16-2-2605. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFRL/RQTF or the U.S. Government.

REFERENCES

1. Brown LM, Gunasekera TS, Striebich RC, Ruiz ON. 2016. Draft genome sequence of Gordonia sihwensis strain 9, a branched alkane-degrading bacterium. Genome Announc 4(3):e00622-16. https://doi.org/10.1128/genomeA.00622-16.
2. Gunasekera TS, Bowen LL, Zhou CE, Howard-Byerly SC, Foley WS, Striebich RC, Dugan LC, Ruiz ON. 2017. Transcriptomic analyses elucidate adaptive differences of closely related strains of Pseudomonas aeruginosa in fuel. Appl Environ Microbiol 83:e03249-16. https://doi.org/10.1128/AEM.03249-16.
3. Ruiz ON, Brown LM, Striebich RC, Smart CE, Bowen LL, Lee JS, Little BJ, Mueller SS, Gunasekera TS. 2016. Effect of conventional and alternative fuels on a marine bacterial community and the significance to bioremediation. Energ Fuels 30:434–444. https://doi.org/10.1021/acs.energyfuels.5b02439.
4. Young D, Rice J, Martin R, Lindquist E, Lipzen A, Grigoriev I, Hibbett D. 2015. Degradation of Bunker C fuel oil by white-rot fungi in sawdust cultures suggests potential applications in bioremediation. PLoS One 10:e0130381. https://doi.org/10.1371/journal.pone.0130381.
5. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
6. 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. https://doi.org/10.1089/cmb.2012.0021.
7. Jeong H, Lee S, Choi GJ, Lee T, Yun SH. 2013. Draft genome sequence of Fusarium fujikuroi B14, the causal agent of the bakanae disease of rice. Genome Announc 1(1):e00035-13. https://doi.org/10.1128/genomeA.00035-13.
8. Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061–1067. https://doi.org/10.1093/bioinformatics/btm071.
9. Smit AFA, Hubley R, Green P. 2010. RepeatMasker open-3.0. http://www.repeatmasker.org.
10. Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a Web server for gene finding in eukaryotes. Nucleic Acids Res 32:W309–W312. https://doi.org/10.1093/nar/gkh379.
11. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 32:W309–W312. https://doi.org/10.1093/nar/gkh379.
12. Lombard V, Golaconda Ramulu HG, Drula E, Coutinho PM, Henrissat B. 2014. The Carbohydrate-Active Enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178.
13. Lida T, Sumita T, Ohta A, Takagi M. 2000. The cytochrome P450ALK multigene family of an n-alkane-assimilating yeast, Yarrowia lipolytica: cloning and characterization of genes coding for new CYP52 family members. Yeast 15:1077–1087.