AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of an Onion Basal Rot Isolate of *Fusarium proliferatum*

Andrew D. Armitage,[a] Andrew Taylor,[b] Michelle T. Hulin,[a] Alison C. Jackson,[b] Richard J. Harrison,[a] John P. Clarkson[b]

[a]NIAB EMR, Kent, United Kingdom

[b]Warwick Crop Centre, School of Life Sciences, University of Warwick, Wellesbourne, Warwick, United Kingdom

ABSTRACT *Fusarium proliferatum* is a component of the onion basal rot disease complex. We present an annotated *F. proliferatum* draft genome sequence, totaling 45.8 Mb in size, assembled into 597 contigs, with a predicted 15,418 genes. The genome contains 58 secondary metabolite clusters and homologs of the *Fusarium oxysporum* effector SIX2.

Fusarium proliferatum is a generalist pathogen of a range of crops as diverse as maize, pineapple, and asparagus (1). It is also a component of the onion basal rot complex alongside the more common pathogen *Fusarium oxysporum* f. sp. *cepae* and may cause discoloration of onion bulb scales (2–5). Genomic resources for *F. oxysporum* from onion, alongside those for *F. proliferatum* strains isolated from other crops, have recently become available (6–8). Expansion of these resources with a genome sequence of *F. proliferatum* from onion provides a basis for a study of host adaptation within *F. proliferatum* and for different strategies of onion infection within the *Fusarium* genus.

F. proliferatum isolate A8 was isolated from an onion bulb with basal rot symptoms obtained from a commercial grower in Bedfordshire, United Kingdom, in 2009. DNA extraction was performed on freeze-dried mycelium using a Macherey-Nagel NucleoSpin plant II kit (catalog no. 11912262; Fisher). Paired-end (PE) genomic libraries were then prepared using an Illumina TruSeq LT kit (catalog no. FC-121-2001). Libraries were sequenced using a 2 × 250-bp PE kit (catalog no. MS-102-2003) on an Illumina MiSeq version 2 instrument, generating 2,986,704 paired reads.

Reads were trimmed and adapters removed using fastq-mcf version 1.04.676 (9) before a 45.8-Mb genome assembly was generated using SPAdes version 3.5.0 (10) in 581 contigs (Table 1). Repeat masking was performed using RepeatMasker version 4.0.3 (http://www.repeatmasker.org) and TransposonPSI (http://transposonpsi.sourceforge.net, 2013-03-05 release). This masked 1.21 Mb of the genome. Genome completeness was assessed through the presence of conserved single-copy fungal genes using BUSCO version 3 (11). We used the Sordariomyceta *odb9* data set, identifying 3,694 of 3,725 (99%) of these genes as present in the assembly. Published RNA sequencing (RNA-seq) data (7) were aligned to the genome using Bowtie 2 version 2.2.4 and TopHat version 2.1.0 (12, 13), with mate-inner-dist set to −20 bp and mate-std-dev set to 70 bp. These alignments were used in the prediction of 15,418 genes encoding 15,448 proteins; 15,421 of these proteins were predicted using BRAKER1 version 2.0 (14), supplemented by an additional 37 proteins predicted by using CodingQuarry version 2.0 (15), located in intergenic regions of BRAKER1 gene models. BRAKER1 was run using the fungal option, and CodingQuarry was run in pathogen mode. Functional annotation was performed using InterProScan version 5.18-57.0 (16) and through BLASTp (E value $< 1 \times 10^{-100}$) searches against the July 2016 release of the Swiss-Prot database (17).

tBLASTx searches (E value $< 1 \times 10^{-10}$) for pathogenicity factors associated with

Citation Armitage AD, Taylor A, Hulin MT, Jackson AC, Harrison RJ, Clarkson JP. 2019. Draft genome sequence of an onion basal rot isolate of *Fusarium proliferatum*. Microbiol Resour Announc 8:e01385-18. https://doi.org/10.1128/MRA.01385-18.

Editor Antonis Rokas, Vanderbilt University

Address correspondence to Andrew D. Armitage, Andrew.Armitage@emr.ac.uk.

R.J.H. and J.P.C. contributed equally to this work.

Received 6 October 2018
Accepted 19 November 2018
Published 3 January 2019

TABLE 1 *F. proliferatum* isolate A8 genome statistics

Statistic	Value
Assembly statistics	
Assembly size (bp)	45,689,467
No. of contigs	581
Largest contig (bp)	1,926,525
GC content (%)	48.65
N_{50} (bp)	535,935
% repeat masked	2.65
Gene models	
Total no. of genes	15,418
Total no. of proteins	15,458
No. encoding secreted proteins	1,254
No. of genes encoding effector candidates:	
Secreted and effector-like structure	258
Secreted CAZYmes[a]	341
Secondary metabolites	
No. of gene clusters	58
mimp sequences	
No. of mimp sequences in genome	6
No. of genes in 2 kb of mimp sequences	11
No. of genes in 2 kb of mimp sequences encoding secreted proteins	1

[a] Carbohydrate-active enzymes.

Fusarium spp. identified two homologs of the SIX2 gene from *F. oxysporum* f. sp. *lycopersici* on contigs 12 and 246 (18, 19). Additional pathogenicity factors were identified following the same approaches used for our recent annotation of *F. oxysporum* f. sp. *cepae* genomes (7) (Table 1). Secondary metabolite gene clusters were predicted using antiSMASH (20), SignalP version 4.1 and TMHMM version 2 were used to predict genes encoding secreted proteins (21, 22), carbohydrate-active enzymes were predicted using dbCAN and the CAZY database classifications (23, 24), and proteins with an effector-like structure were predicted using EffectorP version 1.0 (25). Furthermore, miniature impala (mimp) sequences are located in the vicinity of *F. oxysporum* f. sp. genes that are important in pathogenicity (26) and as such were identified in the genome (Table 1) using previously described RegEx searches (7).

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number MRDB00000000 (BioProject number PRJNA338256). The version described in this paper is version MRDB01000000. Reads are available from GenBank under accession number SRR4408423.

ACKNOWLEDGMENTS

We thank representatives of Hazera Seeds Reinout de Heer, Wessel van Leeuwen, Ningwen Zhang, Tosca Ferber, and Hans van den Biggelaar for support and advice, as well as David Cole and Cathryn Lambourne for useful discussion. We also thank Viktória Vágány for initial isolation of the culture.

This research was funded by Hazera Seeds, BBSRC (grants BB/K020870/1 and BB/K020730/1) and AHDB (grant CP 116).

REFERENCES

1. Jurado M, Marín P, Callejas C, Moretti A, Vázquez C, González-Jaén MT. 2010. Genetic variability and fumonisin production by *Fusarium proliferatum*. Food Microbiol 27:50–57. https://doi.org/10.1016/j.fm.2009.08.001.
2. du Toit LJ, Inglis DA, Pelter GQ. 2003. *Fusarium proliferatum* pathogenic on onion bulbs in Washington. Plant Dis 87:750–750. https://doi.org/10.1094/PDIS.2003.87.6.750A.
3. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma L-J, Smirnov S, Purcell S, Rehman B, Elkins T, Engels R, Wang S, Nielsen CB, Butler J, Endrizzi M, Qui D, Ianakiev P, Bell-Pedersen D, Nelson MA, Werner-Washburne M, Selitrennikoff CP, Kinsey JA, Braun EL, Zelter A, Schulte U, Kothe GO, Jedd G, Mewes W, Staben C, Marcotte E, Greenberg D, Roy A, Foley K, Naylor J, Stange-Thomann N, Barrett R, Gnerre S, Kamal M, Kamvysselis M, Mauceli E, Bielke C, Rudd S, Frishman D, Krystofova S, Rasmussen C, Metzenberg RL, Perkins DD, Kroken S, et al. 2003. The genome sequence of the filamentous fungus Neurospora crassa. Nature 422:859–868. https://doi.org/10.1038/nature01554.
4. Haapalainen M, Latvala S, Kuivainen E, Qiu Y, Segerstedt M, Hannukkala

AO. 2016. Fusarium oxysporum, F. proliferatum and F. redolens associated with basal rot of onion in Finland. Plant Pathol 65:1310–1320. https://doi.org/10.1111/ppa.12521.
5. Ghanbarzadeh B, Mohammadi Goltapeh E, Safaie N. 2014. Identification of *Fusarium* species causing basal rot of onion in East Azarbaijan province, Iran and evaluation of their virulence on onion bulbs and seedlings. Arch Phytopathol Pflanzenschutz 47:1050–1062. https://doi.org/10.1080/03235408.2013.829628.
6. Almiman BF, Shittu TA, Muthumeenakshi S, Baroncelli R, Sreenivasaprasad S. 2018. Genome sequence of the mycotoxigenic crop pathogen Fusarium proliferatum strain ITEM 2341 from date palm. Microbiol Resour Announc 7:e00964-18. https://doi.org/10.1128/MRA.00964-18.
7. Armitage AD, Taylor A, Sobczyk MK, Baxter L, Greenfield BPJ, Bates HJ, Wilson F, Jackson AC, Ott S, Harrison RJ, Clarkson JP. 2018. Characterisation of pathogen-specific regions and novel effector candidates in *Fusarium oxysporum* f. sp. *cepae*. Sci Rep 8:13530. https://doi.org/10.1038/s41598-018-30335-7.
8. Niehaus E-M, Münsterkötter M, Proctor RH, Brown DW, Sharon A, Idan Y, Oren-Young L, Sieber CM, Novák O, Pěnčík A, Tarkowská D, Hromadová K, Freeman S, Maymon M, Elazar M, Youssef SA, El-Shabrawy ESM, Shalaby ABA, Houterman P, Brock NL, Burkhardt I, Tsavkelova EA, Dickschat JS, Galuszka P, Güldener U, Tudzynski B. 2016. Comparative "omics" of the *Fusarium fujikuroi* species complex highlights differences in genetic potential and metabolite synthesis. Genome Biol Evol 8:3574–3599. https://doi.org/10.1093/gbe/evw259.
9. Aronesty E. 2013. Comparison of sequencing utility programs. Open Bioinforma J 7:1–8. https://doi.org/10.2174/1875036201307010001.
10. 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.
11. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
12. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
13. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36. https://doi.org/10.1186/gb-2013-14-4-r36.
14. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32:767–769. https://doi.org/10.1093/bioinformatics/btv661.
15. Testa AC, Hane JK, Ellwood SR, Oliver RP. 2015. CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. BMC Genomics 16:170. https://doi.org/10.1186/s12864-015-1344-4.
16. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.
17. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. 2007. UniProtKB/Swiss-Prot. Methods Mol Biol 406:89–112.
18. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
19. Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, Houterman PM, Kang S, Shim W-B, Woloshuk C, Xie X, Xu J-R, Antoniw J, Baker SE, Bluhm BH, Breakspear A, Brown DW, Butchko RAE, Chapman S, Coulson R, Coutinho PM, Danchin EGJ, Diener A, Gale LR, Gardiner DM, Goff S, Hammond-Kosack KE, Hilburn K, Hua-Van A, Jonkers W, Kazan K, Kodira CD, Koehrsen M, Kumar L, Lee Y-H, Li L, Manners JM, Miranda-Saavedra D, Mukherjee M, Park G, Park J, Park S-Y, Proctor RH, Regev A, Ruiz-Roldan MC, Sain D, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature 464:367–373. https://doi.org/10.1038/nature08850.
20. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39:W339–W346. https://doi.org/10.1093/nar/gkr466.
21. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786. https://doi.org/10.1038/nmeth.1701.
22. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315.
23. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res 37:D233–D238. https://doi.org/10.1093/nar/gkn663.
24. Huang L, Zhang H, Wu P, Entwistle S, Li X, Yohe T, Yi H, Yang Z, Yin Y. 2018. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. Nucleic Acids Res 46:D516–D521. https://doi.org/10.1093/nar/gkx894.
25. Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, Manners JM, Taylor JM. 2016. EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytol 210:743–761. https://doi.org/10.1111/nph.13794.
26. Schmidt SM, Houterman PM, Schreiver I, Ma L, Amyotte S, Chellappan B, Boeren S, Takken FLW, Rep M. 2013. MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. BMC Genomics 14:119. https://doi.org/10.1186/1471-2164-14-119.