

Rothamsted Repository Download

A - Papers appearing in refereed journals

King, R., Urban, M. and Hammond-Kosack, K. E. 2017. Annotation of *Fusarium graminearum* (PH-1) version 5.0. *Genome Announcements*. 5 (2), pp. e01479-16.

The publisher's version can be accessed at:

- <https://dx.doi.org/10.1128/genomeA.01479-16>

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/8v2ww>.

© 12 January 2017, Rothamsted Research. Licensed under the Creative Commons CC BY.



Annotation of *Fusarium graminearum* (PH-1) Version 5.0

Robert King, Martin Urban, Kim E. Hammond-Kosack

Rothamsted Research, Harpenden, Herts, United Kingdom

ABSTRACT *Fusarium graminearum* floral infections are a major risk to the global supply of safe cereal grains. We report updates to the PH-1 reference genome and significant improvements to the annotation. Changes include introduction of legacy annotation identifiers, new gene models, secretome and effectorP predictions, and inclusion of extensive untranslated region (UTR) annotations.

The ascomycete fungus *Fusarium graminearum* is one of several commonly found pathogenic species identified on flowering wheat spikes that cause yield losses, grain quality reductions, and significant mycotoxin contamination in the United Kingdom, Europe, the Americas, Asia, and elsewhere. In 2014, the reference genome for isolate PH-1 was nearly completed with just two remaining gaps (1). In this study the remaining sequence gaps and misassemblies were corrected using new assembler software and the annotation updated. In the United Kingdom, a second pathogen of major concern is *Fusarium culmorum* for which a fragmented genomic reference sequence exists for the United Kingdom isolate UK99 (2). Both species are closely related and produce a range of mycotoxins including the type B trichothecene deoxynivalenol and its derivatives (3). Harvested grains contaminated with trichothecene mycotoxins negatively affect farmers profit margins and impact upon global food and feed security.

Here we report further updates for the reference genome PH-1 of *Fusarium graminearum*. The two remaining sequence gaps were closed by mapping previously not aligned PH-1 reads to the closed gaps in other closely related *F. graminearum* isolates (unpublished genomes). An extensive update of the annotation (14,145 genes) has been included to adopt a new gene naming schema, include legacy annotation gene ID's, and to re-annotate genes with a focus upon those lacking an ATG as a start codon. A new prediction of the secretome has resulted in a reduced cohort of gene predictions ($n = 870$), much more in line with those recently identified in other fungal pathogens species, of which 97 are predicted to be GPI-anchored. We also used effectorP to predict a sub-set of secreted effector proteins ($n = 183$), and updated the gene models with 5,694 5' and 5,036 3' untranslated region (UTR) annotations.

The *F. graminearum* strain PH-1 annotation was transferred to the new assembly via masker 2 (4) followed by a re-annotation using AUGUSTUS (5), and GeneMark (6) to update gene models. Curation was performed on the pre-identified genes lacking an ATG start codon using Geneious (version 9.1 created by Biomatters) as the browser with the new gene models. Secretome predictions were done by the identification of proteins with a signal peptide, but lacking a transmembrane domain. Loci coding for extracellular proteins were predicted using WoLF PSORT (7), followed by Big-PI to predict GPI-anchors (8). The subset of fungal effector proteins was further characterized using EffectorP (9).

Accession number(s). Raw data and the assembled sequences have been updated in the European Nucleotide Archive (ENA). The study accession number is PRJEB5475.

Received 3 November 2016 Accepted 8 November 2016 Published 12 January 2017

Citation King R, Urban M, Hammond-Kosack KE. 2017. Annotation of *Fusarium graminearum* (PH-1) version 5.0. Genome Announc 5:e01479-16. <https://doi.org/10.1128/genomeA.01479-16>.

Copyright © 2017 King et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Robert King, Robert.king@rothamsted.ac.uk.

Accession numbers for the assembled chromosomes are [HG970330](#) to [HG970335](#). Secretome and effector predictions can be found at <https://goo.gl/PKlnwu>.

ACKNOWLEDGMENTS

This research was supported by the Biotechnology and Biological Sciences Research Council of the United Kingdom (BBSRC) through the Institute Strategic Programme Grant 20:20 Wheat (BB/J/00426X/1) and the BBSRC Embrapa Joint Wheat projects (BB/N004493/1 and BB/N018095/1).

We thank the staff at the European nucleotide Archive (ENA), Cambridge, United Kingdom, in particular, Ana M. Cerdeño-Tárraga for her assistance in updating *F. graminearum* PH-1 to the public repository.

REFERENCES

1. King R, Urban M, Hammond-Kosack MCU, Hassani-Pak K, Hammond-Kosack KE. 2015. The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics* 16:544. <https://doi.org/10.1186/s12864-015-1756-1>.
2. Urban M, King R, Andongabo A, Maheswari U, Pedro H, Kersey P, Hammond-Kosack KE. 2016. First draft genome sequence of a UK strain (UK99) of *Fusarium culmorum*. *Genome Announc* 4(5):e00771-16. <https://doi.org/10.1128/genomeA.00771-16>.
3. Scherm B, Orrù M, Balmas V, Spanu F, Azara E, Delogu G, Hammond TM, Keller NP, Migheli Q. 2011. Altered trichothecene biosynthesis in *TRI6*-silenced transformants of *Fusarium culmorum* influences the severity of crown and foot rot on durum wheat seedlings. *Mol Plant Pathol* 12: 759–771. <https://doi.org/10.1111/j.1364-3703.2011.00709.x>.
4. Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. <https://doi.org/10.1186/1471-2105-12-491>.
5. Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7:62. <https://doi.org/10.1186/1471-2105-7-62>.
6. Lukashin AV, Borodovsky M. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* 26:1107–1115. <https://doi.org/10.1093/nar/26.4.1107>.
7. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Res* 35:W585–W587. <https://doi.org/10.1093/nar/gkm259>.
8. Eisenhaber B, Wildpaner M, Schultz CJ, Borner GH, Dupree P, Eisenhaber F. 2003. Glycosylphosphatidylinositol lipid anchoring of plant proteins. Sensitive prediction from sequence- and genome-wide studies for Arabidopsis and rice. *Plant Physiol* 133:1691–1701. <https://doi.org/10.1104/pp.103.023580>.
9. 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>.