Genome Sequences of Six Wheat-Infecting *Fusarium* Species Isolates

Paula M. Moolhuijzen,[a] John M. Manners,[b] Stephen A. Wilcox,[c] Matthew I. Bellgard,[a] Donald M. Gardiner[d]

Centre for Comparative Genomics, Murdoch University, Perth, Australia[a]; Commonwealth Scientific and Industrial Research Organisation (CSIRO) Plant Industry, Black Mountain, Canberra, Australian Capital Territory, Australia[b]; Queensland Bioscience Precinct, Brisbane, Queensland, Australia[d]; Australian Genome Research Facility, the Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia[c]

Fusarium **pathogens represent a major constraint to wheat and barley production worldwide. To facilitate future comparative studies of** *Fusarium* **species that are pathogenic to wheat, the genome sequences of four** *Fusarium pseudograminearum* **isolates, a single** *Fusarium acuminatum* **isolate, and an organism from the** *Fusarium incarnatum-F. equiseti* **species complex are reported.**

Address correspondence to Donald M. Gardiner, Donald.Gardiner@csiro.au.

Diseases incited by *Fusarium* pathogens are among the major biotic constraints on wheat and barley crop production (1). Furthermore, many cereal-infecting fusaria produce mycotoxins that are hazardous to humans and animals. *Fusarium* pathogens can cause three diseases on cereals (head blight, crown rot, and root rot), depending on the host developmental stage. Head blight disease causes enormous losses in northern hemisphere production areas, while crown rot is prevalent in arid regions. In Australia, crown rot is responsible for annual losses of ~$100 million in the wheat and barley industries (2, 3). In drier wheat-growing regions, *Fusarium pseudograminearum*, causing crown rot disease, predominates (4–6) and is the most commonly identified species in Australian wheat fields. The *Fusarium incarnatum-F. equiseti* species complex (FIESC) and *Fusarium acuminatum* isolates, which are less often associated with major disease epidemics, are regularly identified as co-occurring with the other pathogens in field surveys (4, 7). To undertake comparative analyses, six separate genomes covering three *Fusarium* species were sequenced.

Isolates from the CSIRO Plant Industry *Fusarium* collection (Brisbane, Australia) were selected for sequencing after confirming species identifications by sequencing the elongation factor 1 alpha gene (8). Four *F. pseudograminearum* isolates were chosen (Table 1) from different provenances to augment the already existing *F. pseudograminearum* genome (9). Single isolates of *F. acuminatum* and a FIESC member (10) confirmed as species 5 (K. O'Donnell, personal communication) were also sequenced (Table 1). All isolates were tested for virulence on wheat using a laboratory assay for crown rot infection (9).

Illumina TruSeq DNA sequencing libraries were constructed from *in vitro*-grown mycelium-derived DNA. The average insert size range was 383 to 494 bp. Bar-coded samples were sequenced with HiSeq 2000 100-bp paired-end reads. In Yabi (11), the read sequences were assessed for quality with FastQC v0.10.0 (12).

TABLE 1 Isolate provenance, genome assembly, and annotation statistics

Species	Isolate	Locus	Isolate provenance (Australia)	No. of scaffolds	No. of N_{50}[a]	N_{50}[a] (Mbp)	Max[a] (Mbp)	Sum (Mbp)	No. of coding genes	Accession no.
Fusarium sp. FIESC5	CS3069	BN850	Wilga Downs, Queensland	3,555	321	0.03	0.25	38.02	13,743	CBMI010000000
F. pseudograminearum	CS3220	BN846	Liverpool plains, New South Wales	191	12	1.07	2.7	37.16	12,615	CBMC010000000
F. pseudograminearum	CS3427	BN847	Wilga Downs, Queensland	182	12	0.93	2.02	37.07	12,577	CBMD010000000
F. pseudograminearum	CS3487	BN848	Tamworth, New South Wales	364	20	0.62	1.49	37.01	12,749	CBME010000000
F. pseudograminearum	CS5834	BN849	Tammin, Western Australia	228	12	1.04	2.51	37.48	12,633	CBMF010000000
F. acuminatum	CS5907	BN851	Stockdale, Western Australia	716	79	0.17	0.61	43.89	15,353	CBMG01000000

[a] N_{50} and Max statistics refer to scaffold sequences.

Poor quality sequence, adaptors, and redundant reads were removed with ConDeTri v2.0 (13) and rmDupPCR.pl (13), respectively. Any contaminating foreign DNA was filtered out using BWA v0.6.1 (14), samTools v0.1.18 (15), NCBI bacterial genomes, and International Wheat Genetics Symposium (IWGS) wheat survey sequences (16). Mitochondrial sequence was filtered from chromosomal data using BWA alignment to NCBI mitochondrial genomes. The genome and mitochondria were assembled with Velvet v1.2.03 (17) at an optimal hash length of 53. Assembly statistics are given in Table 1. Scaffolds were mapped at a minimum 70 percent identity to the *Fusarium graminearum* PH-1 genome (accession no. CM000575.1) (18) using BLAT v34 (19) into a pseudochromosomal sequence. Low complexity and repeat elements were masked with the RepeatMasker v3.3.0 (20) fungi repeat library and Censor v4.2.28 (21) *Fusarium* library. Protein-coding gene predictions were made with GeneMark-ES v2.3e (22) and Augustus v2.5.5 (23), and *F. pseudograminearum* CS3096 Illumina RNA-Seq unstranded reads were assembled with TopHat v1.4.0 (24) and CuffLinks v1.3.0 (25). Noncoding RNAs were predicted using Aragorn v1.2.33 (26) and RNAmmer v1.2 (27). Protein-coding genes were annotated with AutoFACT v3.4 (28). While most of the genomes encoded around 13,000 predicted proteins, the *F. acuminatum* genome was predicted to encode 15,353 proteins. This genome was also the largest, at 44 Mbp. Sequence data, including scaffolds and pseudochromosomal scaffolds, have been deposited in EMBL.

Nucleotide sequence accession numbers. The six whole-genome shotgun projects have been deposited at DDBJ/EMBL/GenBank under umbrella BioProject no. PRJEB1746. The whole-genome sequence (WGS) accession numbers can be found in Table 1. The versions described in this paper are the first versions.

ACKNOWLEDGMENTS

REFERENCES

1. **Parry DW, Jenkinson P, McLeod L.** 1995. *Fusarium* ear blight (scab) in small grain cereals—a review. Plant Pathol. **44:**207–238.
2. **Murray GM, Brennan JP.** 2009. Estimating disease losses to the Australian wheat industry. Australas. Plant Pathol. **38:**558–570.
3. **Murray GM, Brennan JP.** 2009. The current and potential costs from diseases of barley in Australia. Grains Research and Development Corporation, Australia Government. http://www.grdc.com.au/Resources/Publications/2009/11/The-current-and-potential-costs-from-diseases-of-barley-in-Australia.
4. **Bottalico A, Perrone G.** 2002. Toxigenic *Fusarium* species and mycotoxins associated with head blight in small-grain cereals in Europe. Eur. J. Plant Pathol. **108:**611–624.
5. **Obanor F, Erginbas-Orakci G, Tunali B, Nicol JM, Chakraborty S.** 2010. *Fusarium culmorum* is a single phylogenetic species based on multilocus sequence analysis. Fungal Biol. **114:**753–765.
6. **Smiley RW, Gourlie JA, Easley SA, Patterson LM.** 2005. Pathogenicity of fungi associated with the wheat crown rot complex in Oregon and Washington. Plant Dis. **89:**949–957.
7. **Akinsanmi OA, Mitter V, Simpfendorfer S, Backhouse D, Chakraborty S.** 2004. Identity and pathogenicity of *Fusarium* spp. isolated from wheat fields in Queensland and northern New South Wales. Aust. J. Agric. Res. **55:**97–107.
8. **O'Donnell K, Kistler HC, Cigelnik E, Ploetz RC.** 1998. Multiple evolutionary origins of the fungus causing Panama disease of banana: concordant evidence from nuclear and mitochondrial gene genealogies. Proc. Natl. Acad. Sci. U. S. A. **95:**2044–2049.
9. **Gardiner DM, McDonald MC, Covarelli L, Solomon PS, Rusu AG, Marshall M, Kazan K, Chakraborty S, McDonald BA, Manners JM.** 2012. Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts. PLOS Pathog. **8:**e1002952. doi:10.1371/journal.ppat.1002952.
10. **O'Donnell K, Sutton DA, Rinaldi MG, Gueidan C, Crous PW, Geiser DM.** 2009. Novel multilocus sequence typing scheme reveals high genetic diversity of human pathogenic members of the *Fusarium incarnatum-F. equiseti* and *F. chlamydosporum* species complexes within the United States. J. Clin. Microbiol. **47:**3851–3861.
11. **Hunter A, Macgregor A, Szabo T, Wellington C, Bellgard M.** 2012. Yabi: an online research environment for grid, high performance and cloud computing. Source Code Biol. Med. **7:**10/1186/1751-0473-7-11.
12. **Andrews S.** 2011. FastQC. Babraham Institute, Cambridge, United Kingdom. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
13. **Smeds L, Kunstner A.** ConDeTri—a content dependent read trimmer for Illumina data. PLOS One **6:**e26314. doi:10.1371/journal.pone.0026314.
14. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:**1754–1760.
15. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R.** 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:**2078–2079.
16. **Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo MC, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KF, Edwards KJ, Bevan MW, Hall N.** Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature **491:**705–710.
17. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18:**821–829.
18. **Cuomo CA, Güldener U, Xu JR, Trail F, Turgeon BG, Di PA, Walton JD, Ma LJ, Baker SE, Rep M, Adam G, Antoniw J, Baldwin T, Calvo S, Chang YL, DeCaprio D, Gale LR, Gnerre S, Goswami RS, Hammond-Kosack K, Harris LJ, Hilburn K, Kennell JC, Kroken S, Magnuson JK, Mannhaupt G, Mauceli E, Mewes H, Mitterbauer R, Muehlbauer G, Munsterkotter M, Nelson D, O'Donnell K, Ouellet T, Qi W, Quesneville H, Roncero MIG, Seong KY, Tetko IV, Urban M, Waalwijk C, Ward TJ, Yao J, Birren BW, Kistler HC.** 2007. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science **317:**1400–1402.
19. **Kent WJ.** 2002. BLAT—the BLAST-like alignment tool. Genome Res. **12:**656–664.
20. **Chen N.** 2004. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics **4:**4.4.10. doi:10.1002/0471250953.bi0410s05.
21. **Kohany O, Gentles AJ, Hankus L, Jurka J.** 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics **7:**474. doi:10.1186/1471-2105-7-474.
22. **Borodovsky M, Lomsadze A.** Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. Curr. Protoc. Bioinformatics **4:**4.6.1–4.6.10. doi:10.1002/0471250953.bi0406s35.
23. **Stanke M, Waack S.** 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics **19**(Suppl 2):ii215–ii225. doi:10.1093/bioinformatics/btg1080.
24. **Trapnell C, Pachter L, Salzberg SL.** 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25:**1105–1111.
25. **Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L.** Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. **28:**511–515.
26. **Laslett D, Canback B.** 2004. Aragorn, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. **32:**11–16.
27. **Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35:**3100–3108.
28. **Koski LB, Gray MW, Lang BF, Burger G.** 2005. AutoFACT: an automatic functional annotation and classification tool. BMC Bioinformatics **6:**151. doi:10.1186/1471-2105-6-151.