Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium


DOI
10.1038/nature08850

Publication date
2010

Document Version
Final published version

Published in
Nature

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).
Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*


*Fusarium* species are among the most important mycopathogenic and toxigenic fungi. To understand the molecular underpinnings of pathogenicity in the genus *Fusarium*, we compared the genomes of three phenotypically diverse species: *Fusarium graminearum*, *F. verticillioides* and *Fusarium oxysporum f. sp. lycopersici*. Our analysis revealed lineage-specific (LS) genomic regions in *F. oxysporum* that include four entire chromosomes and account for more than one-quarter of the genome. LS regions are rich in transposons and genes with distinct evolutionary profiles but related to pathogenicity, indicative of horizontal acquisition. Experimentally, we demonstrate the transfer of two LS chromosomes between strains of *F. oxysporum*, converting a non-pathogenic strain into a pathogen. Transfer of LS chromosomes between otherwise genetically isolated strains explains the polyphyletic origin of host specificity and the emergence of new pathogenic lineages in *F. oxysporum*. These findings put the evolution of fungal pathogenicity into a new perspective.

*Fusarium* species are among the most diverse and widely dispersed plant-pathogenic fungi, causing economically important blights, root rots or wilts. Some species, such as *F. graminearum* (Fg) and *F. verticillioides* (Fv), have a narrow host range, infecting predominantly the cereals (Fig. 1A). By contrast, *F. oxysporum* (Fo), has a remarkably broad host range, infecting both monocotyledonous and dicotyledonous plants and is an emerging pathogen of immunocompromised human s and other mammals. Aside from their differences in host adaptation and specificity, *Fusarium* species also vary in reproductive strategy. Some, such as *F. oxysporum*, are asexual, whereas others are both asexual and sexual with either self-fertility (homothallism) or obligate out-crossing (heterothallism) (Fig. 1B).

Previously, the genome of the cereal pathogen *F. graminearum* was sequenced and shown to encode a larger number of proteins in pathogenicity related protein families compared to non-pathogenic fungi, including predicted transcription factors, hydrolytic enzymes, and transmembrane transporters. We sequenced two additional *Fusarium* species, *Fv* and *Fo*, and found that these fungi produce fumonisin mycotoxins that can contaminate grain, and *F. oxysporum* f. sp. *lycopersici* (Fol), a tomato pathogen. Here we present the comparative analysis of the genomes of these three species.

**Results**

**Genome organization and gene clusters.** We sequenced *Fv* strain 7600 and *Fol* strain 4287 (Methods, Supplementary Table 1) using a whole-genome shotgun approach and assembled the sequence using 454 Life Sciences, Branford, Connecticut 06405, USA (C.D.K.); University of Texas Southwestern Medical Center at Dallas, Dallas, Texas 75390, USA (L.L.); University of California, Los Angeles, California 90095, USA (J.G.D.); CSIRO Plant Industry, Queensland Bioscience Precinct, St Lucia, Brisbane, Queensland, 4067 Australia (D.M.G.); University of California, Irvine, California 92697, USA (A.C.); Texas A&M University, College Station, Texas 77843, USA (D.L.); University of California, Irvine, California 92697, USA (J.M.); Centre for Sustainable Pest and Disease Management, Rothamsted Research, Harpenden AL5 2JQ, UK (G.S.R.); Pacific Northwest National Laboratory, Richland, Washington 99352, USA (M.D.G.); USDA ARS, University of Minnesota, St. Paul, Minnesota 55108, USA (A.D.A.); USDA ARS-NCAR, Peoria, Illinois 61604, USA (M.J.-O.D.); European Bioinformatics Institute, Cambridge CB10 1SD, UK (A.R.A.); University of California, Los Angeles, California 90095, USA (D.W.B.); CSIRO Plant Industry, Queensland Bioscience Precinct, St Lucia, Brisbane, Queensland, 4067 Australia (D.M.G.); BIRD Institute, University of Arizona, Tucson, Arizona 85721, USA (C.B.); Seoul National University, Seoul 151-742, Korea (S.-Y.P.); University of Wisconsin-Madison, Madison, Wisconsin 53706 USA (J.G.D.); University of California, Los Angeles, California 90095, USA (D.W.B.); University of Texas Southwestern Medical Center at Dallas, Dallas, Texas 75390, USA (L.L.); INRA, Institut National de la Recherche Agronomique, 06903 Sophia-Antipolis, France (E.G.J.D.); Seoul National University, Seoul 151-742, Korea (S.-Y.P.).

These authors contributed equally to this work.
The Fol genome (60 megabases) is about 44% larger than that of its most closely related species, Fv (42 Mb), and 65% larger than that of Fg (36 Mb), resulting in a greater number of protein-encoding genes in Fol (Table 1).

The relatedness of the three Fusarium genomes enabled the generation of large-scale unambiguous alignments (Supplementary Figs 1–3) and the determination of orthologous gene sets with high confidence (Methods, Supplementary Information C). On average, Fol and Fv orthologues display 91% nucleotide sequence identity, and both have 85% identity with Fg counterparts (Supplementary Fig. 4). Over 9,000 conserved syntenic orthologues were identified among the three genomes. Compared to other ascomycete genomes, these three species orthologues are enriched for predicted transcription factors (P = 2.6 × 10⁻¹⁰), lytic enzymes (P = 0.001), and transmembrane transporters (P = 7 × 10⁻⁹) (Supplementary Information C and Supplementary Tables 3–8), in agreement with results reported for the Fg genome5.

Fusarium species produce diverse secondary metabolites, including mycotoxins that exhibit toxicity to humans and other mammals8. In the three genomes, we identified a total of 46 secondary metabolite biosynthesis (SMB) gene clusters. Microarray analyses confirmed the co-expression of genes in 14 of 18 Fg and 10 of 16 Fv SMB gene clusters. Ten out of the 14 Fg and eight out of the 10 Fv co-expressed SMB gene clusters are novel (Supplementary Information D, Supplementary Fig. 5 and Supplementary Table 9, and online materials), emphasizing the potential impact of uncharacterized secondary metabolites on fungal biology.

Lineage-specific chromosomes and pathogenicity. The genome assembly of Fol has 15 chromosomes, the Fv assembly 11 and the Fg assembly only four (Table 1). The smaller number of chromosomes in Fg is the result of chromosome fusion relative to Fv and Fo, and fusion sites in Fg match previously described high diversity regions (Supplementary Fig. 3, ref. 5). Global comparison among the three Fusarium genomes shows that the increased genomic territory in Fol is due to additional, unique sequences that reside mostly in extra chromosomes. Syntenic regions in Fol cover approximately 80% of the Fg and more than 90% of the Fv genome (Supplementary Information E and Supplementary Table 10), referred to as the ‘core’ of the genomes. Except for telomere-proximal regions, all 11 mapped chromosomes in the Fv assembly (41.1 Mb) correspond to 11 of the 15 chromosomes in Fol (41.8 Mb). The co-linear order of genes between Fol and Fv has been maintained within these chromosomes, except for one chromosomal translocation event and a few local rearrangements (Fig. 2a).

The unique sequences of Fol are a substantial fraction (40%) of the Fol assembly, designated as Fol lineage-specific (Fol LS) regions, to distinguish them from the conserved core genome. The Fol LS regions include four entire chromosomes (chromosomes 3, 6, 14 and 15), parts of chromosome 1 and 2 (scaffold 27 and scaffold 31, respectively), and most of the small scaffolds not anchored to the optical map (Fig. 2b). In total, the Fol LS regions encompass 19 Mb, accounting for nearly all of the larger genome size of Fol.

Notably, the LS regions contain more than 74% of the identifiable transposable elements (TEs) in the Fol genome, including 95% of all DNA transposons (Fig. 2b, Supplementary Fig. 6 and Supplementary Table 11). In contrast to the low content of repetitive sequence and...
minimal amount of TEs in the Fv and Fg genomes (Table 1 and Supplementary Table 11), about 28% of the entire Fol genome was identified as repetitive sequence (Methods), including many retro-elements (copia-like and gypsy-like LTR retrotransposons, LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements) and DNA transposons (Tc1-mariner, hAT-like, Mutator-like, and MITEs) (Supplementary Information E.3), as well as several large segmental duplications. Many of the TEs are full-length and present as highly similar copies. Particularly well represented DNA transposon classes in Fol are pogo, hAT-like elements and MITEs (in total approximately 550, 200 and 350 copies, respectively). In addition, there are one intra-chromosomal and two inter-chromosomal segmental duplications, totalling approximately 7 Mb and resulting in three- or even fourfold duplications of some regions (Fig. 2c). Overall, these regions share 99% sequence identity (Supplementary Fig. 7), indicating recent duplication events.

Only 20% of the predicted genes in the Fol LS regions could be functionally classified on the basis of homology to known proteins. These genes are significantly enriched ($P < 0.0001$) for the functional categories 'secreted effectors and virulence factors', 'transcription factors', and 'proteins involved in signal transduction', but are deficient in genes for house-keeping functions (Supplementary Information E and Supplementary Tables 12–18). Among the genes with a predicted function related to pathogenicity were known effector proteins (see below) as well as necrosis and ethylene-inducing peptides and a variety of secreted enzymes predicted to degrade or modify plant or fungal cell walls (Supplementary information E and Supplementary Tables 14, 15). Notably, many of these enzymes are expressed during early stages of tomato root infection (Supplementary Tables 15, 16 and Supplementary Fig. 8). The expansion of genes for lipid metabolism and lipid-derived secondary messengers in Fol LS regions indicates an important role for lipid signalling in fungal pathogenicity (Supplementary Fig. 9 and Supplementary Tables 13, 17). A family of transcription factor sequences related to FTF1, a gene transcribed specifically during early stages of infection of F. oxysporum f. sp. phaseoli (Supplementary Information E and Supplementary Table 4; ref. 10) is also expanded.

The recently published genome of F. solani, a more diverged species, enabled us to extend comparative analysis to a larger evolutionary framework (Fig. 1). Whereas the 'core' genomes are well conserved among all four sequenced Fusarium species, the Fol LS regions are also absent in Fs (Supplementary Fig. 2). Additionally, Fs has three LS chromosomes distinct from the genome core and the Fol LS regions. In conclusion, each of the four Fusarium species
carries a core genome with a high level of synteny whereas *Fol* and *Fs* each have LS chromosomes that are distinct with regard to repetitive sequences and genes related to host–pathogen interactions.

### Origin of LS regions.

Three possible explanations for the origin of LS regions in the *Fol* genome were considered: (1) *Fol* LS regions were present in the last common ancestor of the four *Fusarium* species but were then selectively and independently lost in *Fv*, *Fg* and *Fs* lineages during vertical transmission; (2) LS regions arose from the core genome by duplication and divergence within the *Fol* lineage; and (3) LS regions were acquired by horizontal transfer. To distinguish among these hypotheses, we compared the sequence characteristics of the genes in the *Fol* LS regions to those of genes in *Fusarium* core regions and genes in other filamentous fungi. If *Fol* LS genes have clear orthologues in the other *Fusarium* species, or paralogues in the core region of *Fol*, this would favour the vertical transmission or duplication with divergence hypotheses, respectively. We found that, whereas 90% of the *Fol* genes in the core regions have homologues in the other two *Fusarium* genomes, about 50% of the genes on *Fol* LS regions lack homologues in either *Fv* or *Fg* (1 × 10⁻²⁰). Furthermore, there is less sequence divergence between *Fol* and *Fv* orthologues in core regions compared to *Fol* and *Fg* orthologues (Fig. 3a), consistent with the species phylogeny. In contrast, the LS genes that have homologues in the other *Fusarium* species are roughly equally distant from both *Fv* and *Fg* genes (Fig. 3b), indicating that the phylogenetic history of the LS genes differs from genes in the core region of the genome.

Both codon usage tables and codon adaptation index (CAI) analysis indicate that the LS-encoding genes exhibit distinct codon usage (Supplementary Information E.5, Supplementary Fig. 10 and Supplementary Table 19) compared to the conserved genes and the *Fusarium* orthologues in *Fol* genomes, further supporting their distinct evolutionary origins. The most significant differences were observed for amino acids Gln, Cys, Ala, Gly, Val, Glu and Thr, with a preference for G and C over A and T among the *Fol* LS genes (Supplementary Table 20). Such GC bias is also reflected in the slightly higher GC-content in their third codon positions (Supplementary Fig. 11).

Of the 1,285 LS-encoded proteins that have homologues in the NCBI protein set, nearly all (93%) have their best BLAST hit to other ascomycete fungi (Supplementary Fig. 12), indicating that *Fol* LS regions are of fungal origin. Phylogenetic analysis based on concatenated sampling of the 362 proteins that share homologues in seven selected ascomycete genomes — including the four sequenced *Fusarium* genomes, *Magnaporthe grisea*², *Neurospora crassa*³ and *Aspergillus nidulans*⁴ — places their origin within the genus *Fusarium* but basal to the three most closely related *Fusarium* species *Fg*, *Fv* and *Fol* (Fig. 3c, Supplementary Table 21). Taken together, we conclude that horizontal acquisition from another *Fusarium* species is the most parsimonious explanation for the origin of *Fol* LS regions.

### LS regions and host specificity.

*F. oxysporum* is considered a species complex, composed of many different asexual lineages that can be pathogenic towards different hosts or non-pathogenic. The *Fol* LS regions differ considerably in sequence among *Fo* strains with different host specificities, as determined by Illumina sequencing of *Fo* strain *Fo*5176, a pathogen of *Arabidopsis*⁵ and EST (expressed sequence tag) sequences from *Fo* f. sp. *vasinfectum*⁶, a pathogen of cotton (Supplementary Information E.2). Despite less than 2% overall sequence divergence between shared sequences of *Fo* and *Fo*5176 (Supplementary Fig. 13A), for most of the sequences in the *Fols* LS regions there is no sequence similarity between strains (Supplementary Fig. 14)⁷. Previously, small, polymorphic and conditionally dispensable chromosomes conferring host-specific virulence have been reported in the fungi *Nectria haematococca*⁸ and *Alternaria alternata*⁹. Small (<2.3 Mb) and variable chromosomes are absent in non-pathogenic *F. oxysporum* isolates (Supplementary Fig. 14), indicating that *Fol* LS chromosomes may also be specifically involved in pathogenic adaptation.

### Transfer of Fo pathogenicity chromosomes.

It is well documented that small proteins are secreted during *Fo* colonizing the tomato xylem system¹⁰,¹¹ and at least two of these, Six1 (*Avr3*) and Six3 (*Avr2*), are involved in virulence functions¹²,¹³. Interestingly, the genes for these proteins, as well as a gene for an *in planta*-secreted oxidoreductase (*ORX1*)¹⁴, are located on chromosome 14, one of the *Fol* LS chromosomes. These genes are all conserved in strains causing tomato wilt, but are generally not present in other strains¹⁵. The genome data enabled the identification of the genes for three additional small *in planta*-secreted proteins on chromosome 14, named...
SIX5, SIX6 and SIX7 (Supplementary Table 22) based on mass spectrometry data obtained previously. Together these seven genes can be used as markers to identify each of the three superfamilies (SC 22, 36 and 51) localized to chromosome 14 (Supplementary Fig. 23 and Supplementary Fig. 15).

In view of the combined experimental findings and computational evidence, we proposed that LS chromosome 14 could be responsible for pathogenicity of Fol towards tomato, and that its mobility between strains could explain its presence in tomato wilt pathogens, comprising several clonal lineages polyphyletic within the Fo species complex, but absent in other lineages. To test these hypotheses, we investigated whether chromosome 14 could be transferred and whether the transfer would shift pathogenicity between different strains of Fo, using the genes for in planta-secreted proteins on chromosome 14 as markers. Fo007, a strain that is able to cause tomato wilt, was co-incubated with a non-pathogenic isolate (Fom) or banana (Foc), respectively. A gene conferring resistance against zeocin (BLE) was inserted close to SIX1 as a marker to select for transfer of chromosome 14 from the donor strain into Fo-47, Fom or Foc. The receiving strains were transformed with a hygromycin resistance gene (HYG), inserted randomly into the genome; three independent hygromycin resistant transfectants per recipient strain were selected. Microconidia of the different strains were isolated and mixed in a 1:1 ratio on agar plates. Spores emerging on these plates after 6–8 days of incubation were selected for resistance to both zeocin and hygromycin. Double drug-resistant colonies were recovered with Fom and Fo-47, but not using Foc as the recipient, at a frequency of roughly 0.1 to 10 per million spores (Supplementary Table 24).

Pathogenicity assays demonstrated that double drug-resistant strains derived from co-incubating Fo007 with Fo-47, referred to as Fo-47, had gained the ability to infect tomato to various degrees (Fig. 4a, b). In contrast, none of the double drug-resistant strains derived from co-incubating Fo007 with Fom were able to infect tomato. All Fo-47 strains contained large portions of Fol chromosome 14 as demonstrated by PCR amplification of the seven gene markers (Fig. 4c, Supplementary Fig. 15 and Supplementary Information F). The parental strains, as well as the sequenced strain Fok2487, each have distinct karyotypes. This enabled us to determine with chromosome electrophoresis whether the entire chromosome 14 of Fo007 was transferred into Fo-47 strains. All Fo-47 strains had the same karyotype as Fo-47, except for the presence of one or two additional small chromosomes (Fig. 4d). The chromosome present in all Fo-47 strains (Fig. 4d, arrow number 1) was confirmed to be chromosome 14 from Fo007 based on its size and a Southern hybridization using a SIX6 probe (Fig. 4e). Interestingly, two double drug-resistant strains (Fo-47 1C and Fo-47 2A in Fig. 4a), which caused the highest level of disease (Fig. 4a, b), have a second extra chromosome, corresponding in size to the smallest chromosome in the donor strain Fo007 (Fig. 4d, arrow number 2).

To rigorously assess whether additional genetic material other than chromosome 14 may have been transferred from Fo007 into Fo-47 strains, we developed PCR primers for amplification of 29 chromosome-specific markers from Fo007 but not Fo-47. These markers (on average two for each chromosome) were used to screen Fo-47 strains for the presence of Fo007-derived genomic regions (Supplementary information F.4 and Supplementary Fig. 16). All Fo-47 strains were shown to have the chromosome 14 markers (Supplementary Fig. 17), but not Fo007 markers located on any core chromosome, confirming that core chromosomes were not transferred. Interestingly, the two Fo-47 strains (1C and 2A) that have the second small chromosome and caused more disease symptoms were also positive for an additional Fo007 marker (Supplementary Fig. 17), associated with a large duplicated LS region in Fok2487: scaffold 18 (1.3 Mb on chromosome 3) and scaffold 21 (1.0 Mb on chromosome 6) (Fig. 2c). The presence of most or all of the sequence of scaffold 18/21 in strains 1C and 2A was confirmed with an additional nine primer pairs for genetic markers scattered over this region (data not shown, see Supplementary Tables 25a, b for primer sequences) (Fig. 4d).

Taken together, we conclude that pathogenicity of Fo-47 strains towards tomato can be specifically attributed to the acquisition of Fol chromosome 14.

Figure 4 | Transfer of a pathogenicity chromosome. a, Tomato plants infected with Fo007, Fo-47 or double drug resistant Fo-47 strains (1A through 3C) derived from this parental combination, two weeks after inoculation as described for b, b, Eight of nine Fo-47 strains derived from pairing Fo007 and Fo-47 show pathogenicity towards tomato. Average disease severity in tomato seedlings was measured 3 weeks after inoculation in arbitrary units (a.u.). The overall phenotype and the extent of browning of vessels was scored on a scale of 0–4: 0, no symptoms; 1, slightly swollen and/or bent hypocotyl; 2, one or two brown vascular bundles in hypocotyl; 3, at least two brown vascular bundles and growth distortion (strong bending of the stem and asymmetric development); 4, all vascular bundles are brown, plant either dead or stunted and wilted. c, The presence of SIX genes and ORX1 in Fom, Fo-47 and Fol isolates and in double drug-resistant strains derived from co-incubation of Fol/Fom and Fol/Fo-47, assessed by PCR on genomic DNA. Co-incubations were performed with the isolates shown in bold. Three independent transformants of Fom and Fo-47 with a randomly inserted hygromycin resistance gene (H1, H2, H3) were investigated. d, Fo-47 strains derived from a Fo007/Fo-47 co-incubation have the same karyotype as Fo-47, plus one or two chromosomes from Fo007. Protoplasts from Fok2487, Fo007 (with BLE on chromosome 14), three independent HYG transformants of Fo-47 (lane Fo-47 H1, H2 and H3) and nine Fo-47 strains (lane 1A to 3C, the number 1, 2 or 3 referring to the HYG resistant transformant from which they were derived) were loaded on a CHEF (contour-clamped homogeneous electric field) gel. Chromosomes of S. pombe were used as a molecular size marker. Arrows 1 and 2 point to additional chromosomes in the Fo-47 strains relative to Fo-47. e, Southern blot of the CHEF gel shown in d, hybridized with a SIX6 probe, showing that chromosome 14 (arrow 1 in d) is present in all strains except Fo-47 (H1, H2 and H3).
chromosome 14, which contains all known genes for small in planta-secreted proteins. In addition, genes on other LS chromosomes may further enhance virulence as demonstrated by the two strains containing the additional LS chromosome from Fol907. We did not find a double drug-resistant strain with a tagged chromosome of Fo-47 in the Fol907 background. Also, a randomly tagged transformant of Fol907 did not render any double drug-resistant colonies when co-incubated with Fo-47 (data not shown). This indicates that transfer between strains may be restricted to certain chromosomes, perhaps determined by various factors, including size and TE content of the chromosome. Their propensity for transfer is supported by the fact that the smallest LS chromosome in Fol907 moved to Fo-47 without being selected for drug resistance in two out of nine cases.

Discussion

Comparison of Fusarium genomes revealed a remarkable genome organization and dynamics of the assexual species Fol. This tomato pathogen contains four unique chromosomes making up more than one-quarter of its genome. Sequence characteristics of the genes in the LS regions indicate a distinct evolutionary origin of these regions. Experimentally, we have demonstrated the transfer of entire LS chromosome regions in both co- and non-co-incubation between two otherwise genetically isolated members of Fol. The relative ease by which new tomato pathogenic genotypes are generated supports the hypothesis that such transfer between Fo strains may have occurred in nature and has a direct impact on our understanding of the evolving nature of fungal pathogens. Although rare, horizontal gene transfer has been documented in other eukaryotes, including metazoans. However, spontaneous horizontal transfer of such a large portion of a genome and the direct demonstration of associated transfer of host-specific pathogenicity has not been previously reported.

Horizontal transfer of host specificity factors between otherwise distant and genetically isolated lineages of Fo may explain the apparent polyphyletic origins of host specialization and the rapid emergence of new pathogenic lineages in otherwise distinct and incompatible genetic backgrounds. Fol LS regions are enriched for genes related to host–pathogen interactions. The mobilization of these chromosomes could, in a single event, transfer an entire suite of genes required for host–pathogen interactions. If the recipient lineage has an environmental adaptation different from the donor, transfer could increase the overall incidence of disease in the host by introducing pathogenicity in a genetic background pre-adapted to a local environment. Such knowledge of the mechanisms underpinning rapid pathogenesis will affect the development of strategies for disease management in agricultural settings.

METHODS SUMMARY

Generation of genome sequencing and assembly. The whole genome shotgun (WGS) assemblies of Fo (8x coverage) and Fol (6.8x coverage) were generated using Sanger sequencing technology and assembled using Arachne. Physical maps were created by anchoring the assemblies to the Fo genetic linkage map and to the Fol optical map, respectively. Defining hierarchical synteny. Local-orientation anchors were detected using PatternHunter (1 X 10^5) (ref. 29). Contiguous sets of anchors with conserved order and orientation were chained together within 10 kb distance and filtered to ensure that no block overlaps another block by more than 90% of its length. Identification of repetitive sequences. Repeats were detected by searching the genome sequence against itself using CrossMatch (≥ 200 bp and ≥ 60% sequence similarity). Full-length TEs were annotated using a combination of computational predictions and manual inspection. Large segmental duplications were identified using Map Align. Characterization of proteomes. Orthologous genes were determined based on BLASTP and pair-wise syntenic alignments (SL). The blast score ratios tests were used to compare relatedness of proteins among three genomes. The EMBL-PPS tool ‘cusp’ (http://emboss.sourceforge.net/) was used to calculate codon usage frequencies. Gene Ontology terms were assigned using Blast2GO software (BLASTP 1 X 10^30) and tested for enrichment using Fisher’s exact test, corrected for multiple testing. A combination of homology search and manual inspection was used to characterize gene families. Potentially secreted proteins were identified using SignalP (http://www.cbs.dtu.dk/services/SignalP/) after removing trans-membrane/mitochondrial proteins based with TMHMM (http://www.cbs.dtu.dk/services/TMHMM/), Phobius (except in the first 50 amino acids), and TargetP (RC score 1 or 2) predictions. Small cysteine-rich secreted proteins were defined as secreted proteins that are less than 200 amino acids in length and contain at least 4% cysteine residues. GPI (glycosyl phosphatidylinositol)-anchor proteins were identified by the GPI-anchor attachment signal among the predicted secreted proteins using a custom PERL script.

Received 26 June 2009; accepted 20 January 2010.


**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** The 4X sequence of *F. verticillioides* was provided by Syngenta Biotechnology Inc. Generation of the other 4X sequence of *F. verticillioides* and 6.8X sequence of *F. oxysporum* f. sp. *lycopersici* was funded by the National Research Initiative of USDA’s National Institute of Food and Agriculture through the Microbial Genome Sequencing Program (2005-35600-16405) and conducted by the Broad Institute Sequencing Platform. Wayne Xu and the Minnesota Supercomputing Institute for Advanced Computational Research are also acknowledged for their support. The authors thank Leslie Gaffney at the Broad Institute for graphic design and editing and Tracy E. Anderson of the University of Minnesota, College of Biological Sciences Imaging Center for spore micrographs.

**Author Contributions** L.-J.M., H.C.D., M.R. and H.C.K. coordinated genome annotation, data analyses, experimental validation and manuscript preparation. L.-J.M. and H.C.D. made equivalent contributions and should be considered joint first authors. H.C.K. and M.R. contributed equally as corresponding authors. K.A.B., C.A.C., J.J.C., M.-J.D., A.D.P., M.D., M.F., J.G., M.G., B.H., P.M.H., S.K., W.-B.S., C.W., X.X. and J.-R.X. made major contributions to genome sequencing, assembly, analyses and production of complementary data and resources. All other authors are members of the genome sequencing consortium and contributed annotation, analyses or data throughout the project.

**Author Information** All sequence reads can be downloaded from the NCBI trace repository. The assemblies of Fv and Fol have been deposited at GenBank under the project accessions AAIM02000000 and AAXH01000000. Detailed information can be accessed through the Broad Fusarium comparative website: http://www.broad.mit.edu/annotation/genome/fusarium_group.3/MultiHome.html. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.C.K. (hckist@umn.edu) or M.R. (m.rep@uva.nl).