ARTICLES

Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*

Li-Jun Ma[1]*, H. Charlotte van der Does[2]*, Katherine A. Borkovich[3], Jeffrey J. Coleman[4], Marie-Josée Daboussi[5], Antonio Di Pietro[6], Marie Dufresne[5], Michael Freitag[7], Manfred Grabherr[1], Bernard Henrissat[8], Petra M. Houterman[2], Seogchan Kang[9], Won-Bo Shim[10], Charles Woloshuk[11], Xiaohui Xie[12], Jin-Rong Xu[11], John Antoniw[13], Scott E. Baker[14], Burton H. Bluhm[11], Andrew Breakspear[15], Daren W. Brown[16], Robert A. E. Butchko[16], Sinead Chapman[1], Richard Coulson[17], Pedro M. Coutinho[8], Etienne G. J. Danchin[8]†, Andrew Diener[18], Liane R. Gale[15], Donald M. Gardiner[19], Stephen Goff[20], Kim E. Hammond-Kosack[13], Karen Hilburn[15], Aurélie Hua-Van[5], Wilfried Jonkers[2], Kemal Kazan[19], Chinnappa D. Kodira[1]†, Michael Koehrsen[1], Lokesh Kumar[1], Yong-Hwan Lee[21], Liande Li[3], John M. Manners[19], Diego Miranda-Saavedra[22], Mala Mukherjee[10], Gyungsoon Park[3], Jongsun Park[21], Sook-Young Park[9]†, Robert H. Proctor[16], Aviv Regev[1], M. Carmen Ruiz-Roldan[6], Divya Sain[3], Sharadha Sakthikumar[1], Sean Sykes[1], David C. Schwartz[23], B. Gillian Turgeon[24], Ilan Wapinski[1], Olen Yoder[25], Sarah Young[1], Qiandong Zeng[1], Shiguo Zhou[23], James Galagan[1], Christina A. Cuomo[1], H. Corby Kistler[15] & Martijn Rep[2]

***Fusarium* species are among the most important phytopathogenic 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*, *Fusarium 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[1]. 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[2] and is an emerging pathogen of immunocompromised humans[3] and other mammals[4]. Aside from their differences in host adaptation and specificity, *Fusarium* species also vary in reproductive strategy. Some, such as *Fo*, 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 *Fg* 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[5]. We sequenced two additional *Fusarium* species, *Fv*, a maize pathogen that produces 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 Arachne (Table 1, ref. 6). Chromosome level ordering of the scaffolds was achieved by anchoring the assemblies either to a genetic map for *Fv* (ref. 7), or an optical map for *Fol* (Supplementary Information A and Supplementary Table 2). We predicted *Fol* and *Fv* genes and re-annotated a new assembly of the *Fg* genome using a combination of manual and automated annotation (Supplementary Information B).

[1]The Broad Institute, Cambridge, Massachusetts 02141, USA. [2]University of Amsterdam, Amsterdam 1098XH, The Netherlands. [3]University of California Riverside, California 92521, USA. [4]University of Arizona, Tucson, Arizona 85721, USA. [5]Université Paris-Sud, 91405 Paris, France. [6]Universidad de Cordoba, Cordoba 14071, Spain. [7]Oregon State University, Corvallis, Oregon 97331, USA. [8]CNRS, Universités Aix-Marseille, 13628 Aix-en-Provence, France. [9]Penn State University, University Park, Pennsylvania 16802, USA. [10]Texas A&M University, College Station, Texas 77843, USA. [11]Purdue University, West Lafayette, Indiana 47907, USA. [12]University of California, Irvine, California 92697, USA. [13]Centre for Sustainable Pest and Disease Management, Rothamsted Research, Harpenden AL5 2JQ, UK. [14]Pacific Northwest National Laboratory, Richland, Washington 99352, USA. [15]USDA ARS, University of Minnesota, St. Paul, Minnesota 55108, USA. [16]USDA-ARS-NCAUR, Peoria, Illinois 61604, USA. [17]European Bioinformatics Institute, Cambridge CB10 1SD, UK. [18]University of California, Los Angeles, California 90095, USA. [19]CSIRO Plant Industry, Queensland Bioscience Precinct, St Lucia, Brisbane, Queensland, 4067 Australia. [20]BIO5 Institute, University of Arizona, Tucson, Arizona 85721, USA. [21]Seoul National University, Seoul 151-742, Korea. [22]Cambridge Institute for Medical Research, Cambridge CB2 0XY, UK. [23]University of Wisconsin-Madison, Madison, Wisconsin 53706 USA. [24]Cornell University, Ithaca, New York 14853, USA. [25]17885 Camino Del Roca, Ramona, California 92065, USA. †Present addresses: 454 Life Sciences, Branford, Connecticut 06405, USA (C.D.K.); 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.

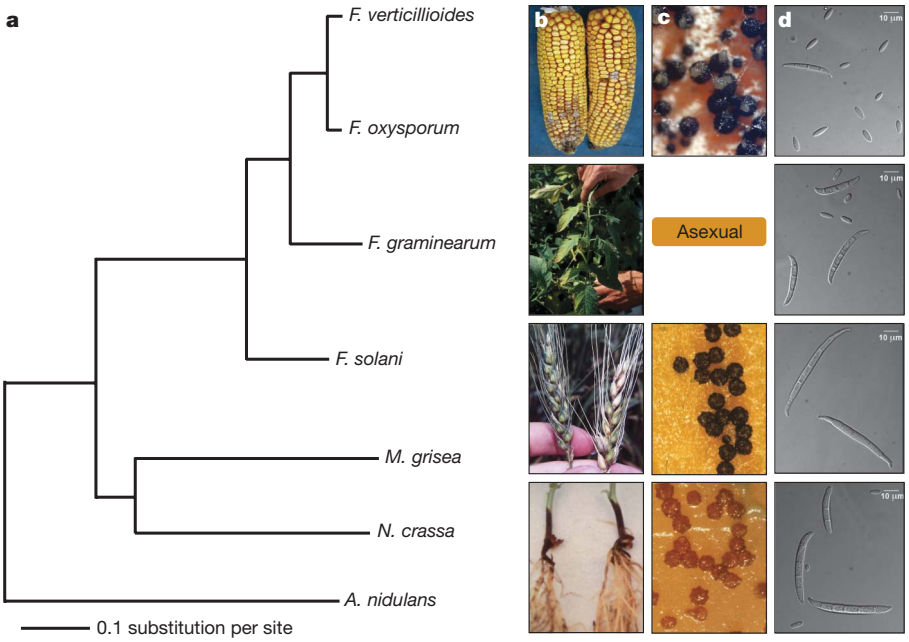

Figure 1 | Phylogenetic relationship of four *Fusarium* species in relation to other ascomycete fungi and phenotypic variation among the four *Fusarium* species. a, Maximum-likelihood tree using concatenated protein sequences of 100 genes randomly selected from 4,694 *Fusarium* orthologous genes that have clear 1:1:1:1 correlation among the *Fusarium* genomes and have unique matches in *Magnaporthe grisea*, *Neurospora crassa* and *Aspergillus nidulans*. The tree was constructed with PHYML[35] (WAG model of evolution[36]). Branches are labelled with the percentage of 10,000 bootstrap replicates. **b–d**, Phenotypic variation within the genus *Fusarium*: **b**, disease symptoms of (top to bottom) kernel rot of maize (*Fv*), wilt of tomato (*Fol*), head blight of wheat (*Fg*) and root rot of pea (*Fs*); **c**, the perithecial states of *Fv* (*Gibberella moniliformis*), *Fol* (no sexual state), *Fg* (*G. zeae*) and *Fs* (*Nectria haematococca*); and **d**, micro- and macroconidia of *Fv*, *Fol*, *Fg* and *Fs*. Scale bars, 10 μm. *Fg* produces only macroconidia.

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 ($P = 2.6 \times 10^{-6}$), lytic enzymes ($P = 0.001$), and transmembrane transporters ($P = 7 \times 10^{-9}$) (Supplementary Information C and Supplementary Tables 3–8), in agreement with results reported for the *Fg* genome[5].

Fusarium species produce diverse secondary metabolites, including mycotoxins that exhibit toxicity to humans and other mammals[8]. 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.

Table 1 | Genome statistics.

Species	*F. oxysporum*	*F. verticillioides*	*F. graminearum*
Strain	4287	7600	PH-1
Sequence coverage (fold)	6	8	10
Genome size (Mb)	59.9	41.7	36.2
Number of chromosomes	15	11*	4
Total scaffolds	114	31	36
N_{50} scaffold length (Mb)	1.98	1.96	5.35
Coding genes	17,735	14,179	13,332
Median gene length (bp)	1,292	1,397	1,355
Repetitive sequence (Mb)	16.83	0.36	0.24
Transposable elements (%)	3.98	0.14	0.03
NCBI accession	AAXH01000000	AAIM02000000	AACM00000000

* *Fv* was reported to contain 12 chromosomes[7], 11 chromosomes were mapped to the assembled genome, and no genetic markers from the smallest chromosome (600 kb or less) were found in the sequence data. N_{50} represents the size N such that 50% of the nucleotides is contained in scaffolds of size *N* or greater.

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

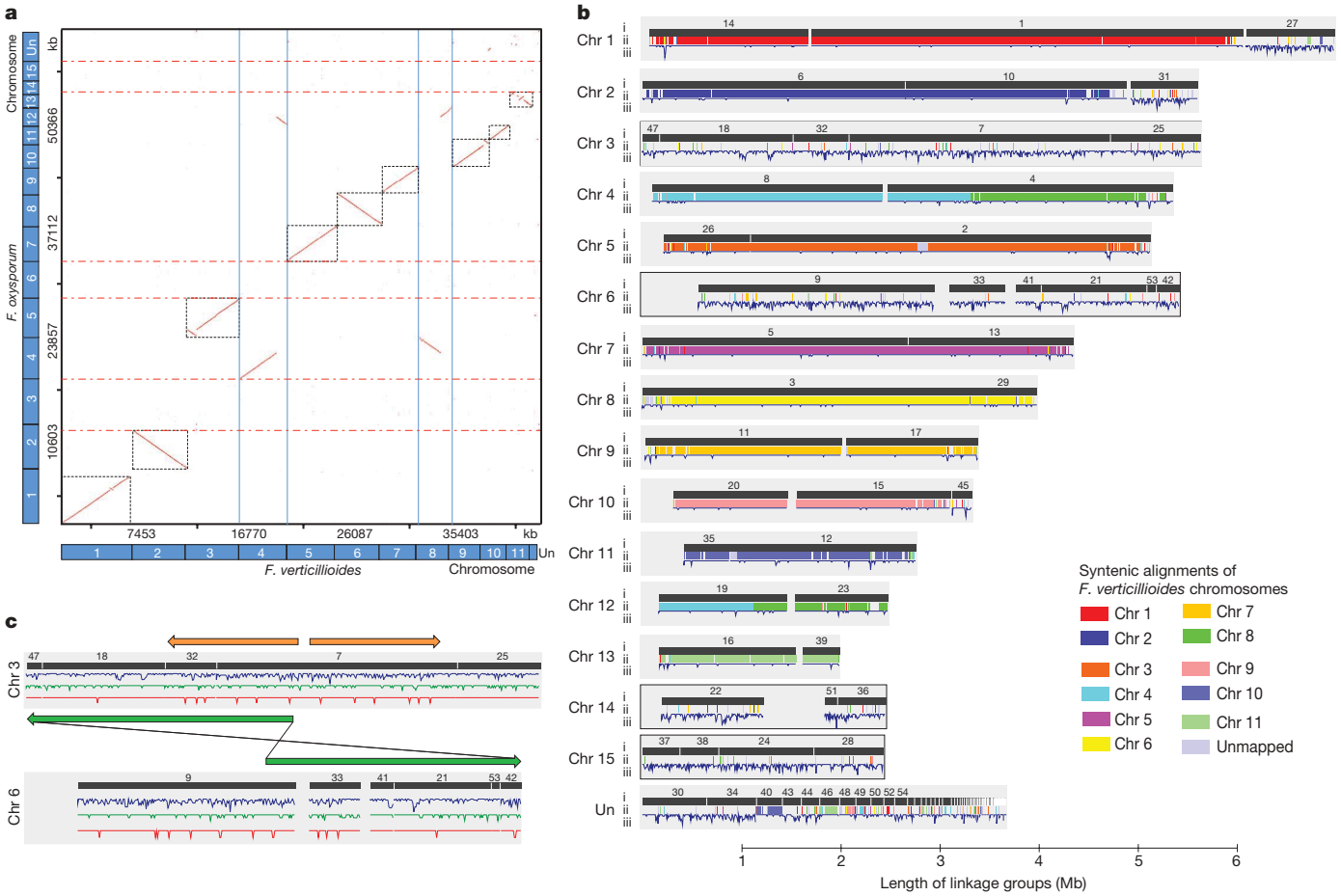

Figure 2 | Whole genome comparison between *Fv* and *Fol*. **a**, Argo[37] dotplot of pair-wise MEGABLAST alignment (1×10^{-10}) between *Fv* and *Fol* showing chromosome correspondences between the two genomes in the black dashed boxes. The vertical blue lines illustrate the chromosomal translocations, and the red dashed horizontal boxes highlight the *Fol* LS chromosomes. **b**, Global view of syntenic alignments between *Fol* and *Fv* and the distribution of transposable elements. *Fol* linkage groups are shown as the reference, and the length of the light grey background for each linkage group is defined by the *Fol* optical map. For each chromosome, row i represents the genomic scaffolds positioned on the optical linkage groups separated by scaffold breaks. Scaffold numbers for *Fol* are given above the blocks; row ii displays the syntenic mapping of *Fv* chromosomes, with one major translocation between chr 4/chr 12 in *Fol* and chr 4/chr 8 in *Fv*; row iii represents the density of transposable elements calculated with a 10 kb window. LS chromosomes include four entire chromosomes (chr 3, chr 6, chr 14 and chr 15) and parts of chromosome 1 and 2 (scaffold 27, scaffold 31), which lack similarity to syntenic chromosomes in *Fv* but are enriched for TEs. **c**, Two of the four *Fol* LS chromosomes showing the inter- (green) and intra- (yellow) chromosomal segmental duplications. The three traces below are density distribution of TEs (blue lines), secreted protein genes (green lines) and lipid metabolism related genes (red line). Chr, chromosome; Un, unmapped.

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[9] 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*[11], 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[11] 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 genes in the *Fv* genome, 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*[12], *Neurospora crassa*[13] and *Aspergillus nidulans*[14] — 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*[15] and EST (expressed sequence tag) sequences from *Fo* f. sp. *vasinfectum*[16], a pathogen of cotton (Supplementary Information E.2). Despite less than 2% overall sequence divergence between shared sequences of *Fol* and *Fo*5176 (Supplementary Fig. 13A), for most of the sequences in the *Fol* LS regions there is no counterpart in *Fo*5176. (Supplementary Fig. 13B). Also *Fov* EST sequences[16] have very high nucleotide sequence identity to the *Fol* genome (average 99%), but only match the core regions of *Fol* (Supplementary Information E.2). Large-scale genome polymorphism within *Fo* is also evident by differences in karyotype between strains (Supplementary Fig. 14)[17]. Previously, small, polymorphic and conditionally dispensable chromosomes conferring host-specific virulence have been reported in the fungi *Nectria haematococca*[18] and *Alternaria alternata*[19]. Small (<2.3 Mb) and variable chromosomes are absent in

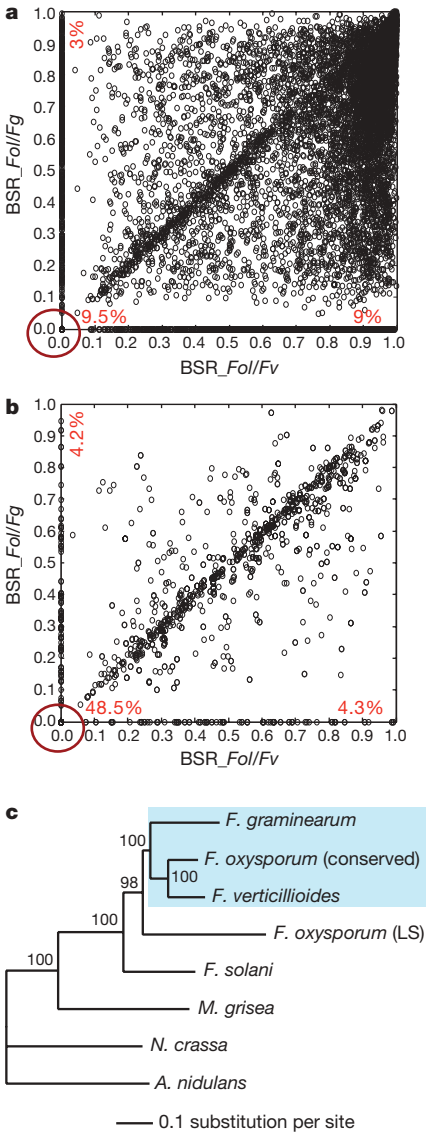

Figure 3 | Evolutionary origin of genes on the *Fol* LS chromosomes. The scatter plots of BLAST score ratio (BSR)[30] based on three-way comparisons of proteins encoded in core regions (**a**) and the *Fol* LS chromosomes (**b**). The numbers indicate the percentage of genes that lack homologous sequences in *Fv* and *Fg* (lower left corner), present in *Fv* but not *Fg* (*x*-axis) and present in *Fg* but not in *Fv* (*y*-axis). **c**, Discordant phylogenetic relationship of proteins encoded in the LS regions. The maximum-likelihood tree was constructed using the concatenated protein sequences of 100 genes randomly selected from 362 genes that share homologues in seven selected ascomycetes genomes including the four *Fusarium* genomes, *M. grisea*, *N. crassa* and *A. nidulans*. The trees were constructed with PHYML[35] (WAG model of evolution[36]). The percentages for the branches represent the value based on a 10,000 bootstrapping data set.

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 *Fol* colonizing the tomato xylem system[20,21] and at least two of these, Six1 (Avr3) and Six3 (Avr2), are involved in virulence functions[22,23]. Interestingly, the genes for these proteins, as well as a gene for an *in planta*-secreted oxidoreductase (*ORX1*)[20], 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 strains[24]. 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[20]. Together these seven genes can be used as markers to identify each of the three supercontigs (SC 22, 36 and 51) localized to chromosome 14 (Supplementary Table 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 absence in other lineages[24]. 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. *Fol*007, a strain that is able to cause tomato wilt, was co-incubated with a non-pathogenic isolate (*Fo*-47) and two other strains that are pathogenic towards melon (*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 transformants 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 *Fol*007 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 *Fol*007 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 *Fol*4287, each have distinct karyotypes. This enabled us to determine with chromosome electrophoresis whether the entire chromosome 14 of *Fol*007 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 *Fol*007 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 *Fol*007 (Fig. 4d, arrow number 2).

To rigorously assess whether additional genetic material other than chromosome 14 may have been transferred from *Fol*007 into *Fo*-47$^+$ strains, we developed PCR primers for amplification of 29 chromosome-specific markers from *Fol007* but not *Fo-47*. These markers (on average two for each chromosome) were used to screen *Fo*-47$^+$ strains for the presence of *Fol*007-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 *Fol*007 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 *Fol*007 marker (Supplementary Fig. 17), associated with a large duplicated LS region in *Fol*4287: 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*

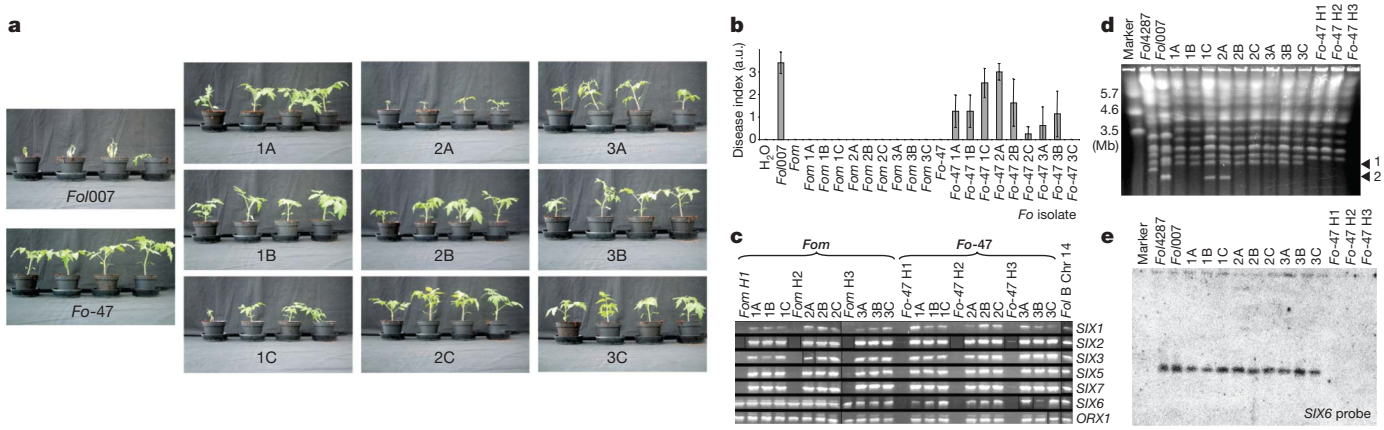

Figure 4 | Transfer of a pathogenicity chromosome. a, Tomato plants infected with *Fol*007, *Fo*-47 or double drug resistant *Fo*-47$^+$ strains (1A through 3C) derived from this parental combination, two weeks after inoculation as described for **b**. **b**, Eight of nine *Fo*-47$^+$ strains derived from pairing *Fol*007 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 *Fol*007/*Fo*-47 co-incubation have the same karyotype as *Fo*-47, plus one or two chromosomes from *Fol*007. Protoplasts from *Fol*4287, *Fol*007 (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 *Fol*007. We did not find a double drug-resistant strain with a tagged chromosome of *Fo*-47 in the *Fol*007 background. Also, a randomly tagged transformant of *Fol*007 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 *Fol*007 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 asexual 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 chromosomes through simple co-incubation between two otherwise genetically isolated members of *Fo*. 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[24] 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[26]. 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[27] and the rapid emergence of new pathogenic lineages in otherwise distinct and incompatible genetic backgrounds[28]. *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 compatibility to a new genetic lineage. If the recipient lineage had 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 pathogen adaptation 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 *Fv* (8× coverage) and *Fol* (6.8× coverage) were generated using Sanger sequencing technology and assembled using Arachne[6]. Physical maps were created by anchoring the assemblies to the *Fv* genetic linkage map[7] and to the *Fol* optical map, respectively.

Defining hierarchical synteny. Local-alignment anchors were detected using PatternHunter (1×10^{10}) (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 $\geq 60\%$ sequence similarity). Full-length TEs were annotated using a combination of computational predictions and manual inspection. Large segmental duplications were identified using Map Aligner[30].

Characterization of proteomes. Orthologous genes were determined based on BLASTP and pair-wise syntenic alignments (SI). The blast score ratio tests[31] were used to compare relatedness of proteins among three genomes. The EMBOSS tool 'cusp' (http://emboss.sourceforge.net/) was used to calculate codon usage frequencies. Gene Ontology terms were assigned using Blast2GO[32] software (BLASTP 1×10^{20}) and tested for enrichment using Fisher's exact test, corrected for multiple testing[33]. A combination of homology search and manual inspection was used to characterize gene families[34,35]. 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 phosphatidyl inositol)-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.

1. Agrios, G. N. *Plant Pathology* 5th edn (Academic Press, 2005).
2. Armstrong, G. M. & Armstrong, J. K. in *Fusarium: Diseases, Biology and Taxonomy* (eds Nelson, P. E., Toussoun, T. A. & Cook R.) 391–399 (Penn State University Press, 1981).
3. O'Donnell, K. *et al.* Genetic diversity of human pathogenic members of the *Fusarium oxysporum* complex inferred from multilocus DNA sequence data and amplified fragment length polymorphism analyses: evidence for the recent dispersion of a geographically widespread clonal lineage and nosocomial origin. *J. Clin. Microbiol.* **42,** 5109–5120 (2004).
4. Ortoneda, M. *et al. Fusarium oxysporum* as a multihost model for the genetic dissection of fungal virulence in plants and mammals. *Infect. Immun.* **72,** 1760–1766 (2004).
5. Cuomo, C. A. *et al.* The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* **317,** 1400–1402 (2007).
6. Jaffe, D. B. *et al.* Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Res.* **13,** 91–96 (2003).
7. Xu, J. R. & Leslie, J. F. A genetic map of *Gibberella fujikuroi* mating population A (*Fusarium moniliforme*). *Genetics* **143,** 175–189 (1996).
8. Desjardins, A. E. & Proctor, R. H. Molecular biology of *Fusarium* mycotoxins. *Int. J. Food Microbiol.* **119,** 47–50 (2007).
9. Qutob, D. *et al.* Phytotoxicity and innate immune responses induced by Nep1-like proteins. *Plant Cell* **18,** 3721–3744 (2006).
10. Ramos, B. *et al.* The gene coding for a new transcription factor (*ftf1*) of *Fusarium oxysporum* is only expressed during infection of common bean. *Fungal Genet. Biol.* **44,** 864–876 (2007).
11. Coleman, J. J. *et al.* The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet.* **5,** e1000618 (2009).
12. Dean, R. A. *et al.* The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434,** 980–986 (2005).
13. Galagan, J. E. *et al.* The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422,** 859–868 (2003).
14. Galagan, J. E. *et al.* Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438,** 1105–1115 (2005).
15. Thatcher, L. F., Manners, J. M. & Kazan, K. *Fusarium oxysporum* hijacks COI1-mediated jasmonate signaling to promote disease development in *Arabidopsis*. *Plant J.* **58,** 927–939 (2009).
16. Dowd, C., Wilson, I. W. & McFadden, H. Gene expression profile changes in cotton root and hypocotyl tissues in response to infection with *Fusarium oxysporum* f. sp. *vasinfectum*. *Mol. Plant Microbe Interact.* **17,** 654–667 (2004).
17. Teunissen, H. A. *et al.* Construction of a mitotic linkage map of *Fusarium oxysporum* based on Foxy-AFLPs. *Mol. Genet. Genomics* **269,** 215–226 (2003).
18. Miao, V. P., Covert, S. F. & VanEtten, H. D. A fungal gene for antibiotic resistance on a dispensable ("B") chromosome. *Science* **254,** 1773–1776 (1991).
19. Harimoto, Y. *et al.* Expression profiles of genes encoded by the supernumerary chromosome controlling AM-toxin biosynthesis and pathogenicity in the apple pathotype of *Alternaria alternata*. *Mol. Plant Microbe Interact.* **20,** 1463–1476 (2007).
20. Houterman, P. M. *et al.* The mixed xylem sap proteome of *Fusarium oxysporum*-infected tomato plants. *Mol. Plant Pathol.* **8,** 215–221 (2007).
21. van der Does, H. C. *et al.* Expression of effector gene *SIX1* of *Fusarium oxysporum* requires living plant cells. *Fungal Genet. Biol.* **45,** 1257–1264 (2008).
22. Houterman, P. M. *et al.* The effector protein Avr2 of the xylem colonizing fungus *Fusarium oxysporum* activates the tomato resistance protein I-2 intracellularly. *Plant J.* **58,** 970–978 (2009).
23. Rep, M. *et al.* A small, cysteine-rich protein secreted by *Fusarium oxysporum* during colonization of xylem vessels is required for I-3-mediated resistance in tomato. *Mol. Microbiol.* **53,** 1373–1383 (2004).
24. van der Does, H. C. *et al.* The presence of a virulence locus discriminates *Fusarium oxysporum* isolates causing tomato wilt from other isolates. *Environ. Microbiol.* **10,** 1475–1485 (2008).
25. Gladyshev, E. A., Meselson, M. & Arkhipova, I. R. Massive horizontal gene transfer in bdelloid rotifers. *Science* **320,** 1210–1213 (2008).
26. O'Donnell, K., Kistler, H. C., Cigelnik, E. & Ploetz, R. C. Multiple evolutionary origins of the fungus causing Panama disease of banana: concordant evidence from nuclear and mitochondrial gene genealogies. *Proc. Natl Acad. Sci. USA* **95,** 2044–2049 (1998).
27. Gale, L. R., Katan, T. & Kistler, H. C. The probable center of origin of *Fusarium oxysporum* f. sp. *lycopersici* VCG 0033. *Plant Dis.* **87,** 1433–1438 (2003).

28. Li, M., Ma, B., Kisman, D. & Tromp, J. Patternhunter II: highly sensitive and fast homology search. *J. Bioinform. Comput. Biol.* **2,** 417–439 (2004).
29. Zhou, S. *et al.* Single-molecule approach to bacterial genomic comparisons via optical mapping. *J. Bacteriol.* **186,** 7773–7782 (2004).
30. Rasko, D. A., Myers, G. S. & Ravel, J. Visualization of comparative genomic analyses by BLAST score ratio. *BMC Bioinformatics* **6,** 2 (2005).
31. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21,** 3674 (2005).
32. Blüthgen, N. *et al.* Biological profiling of gene groups utilizing Gene Ontology. *Genome Inform* **16,** 106–115 (2005).
33. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37** (Database issue), D233–D238 (2009).
34. Miranda-Saavedra, D. & Barton, G. J. Classification and functional annotation of eukaryotic protein kinases. *Proteins* **68,** 893–914 (2007).
35. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).
36. Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18,** 691–699 (2001).
37. Engels, R. *et al.* Combo: a whole genome comparative browser. *Bioinformatics* **22,** 1782–1783 (2006).

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

Acknowledgements The 4× sequence of *F. verticillioides* was provided by Syngenta Biotechnology Inc. Generation of the other 4× sequence of *F. verticillioides* and 6.8× 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.
 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).

Supplementary Information: ***Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium oxysporum***

Outline:

A. Genome sequencing, assembling and mapping.

An 8X whole genome shotgun (WGS) assembly of *Fv* strain 7600 was generated combining genomic sequences from the Broad Institute and Syngenta Biotechnology Inc. (Syngenta). A 6.28X *Fo* assembly was generated at the Broad Institute also using WGS approach (Table S1). Both genomes were sequenced using Sanger technology and assembled using Arachne [1].

A total of 142 RFLP markers comprising the *F. verticillioides* (*Fv*) genetic linkage map [2] were sequenced in both directions at the Broad. Both forward and reverse sequences were treated as independent markers, resulting in 284 marker sequences. More than 99% of the assembly was anchored to the genetic map by locating the position of the sequence of 131 marker sequences within the current assembly using BlastN. The majority of unplaced markers on the map were caused by incomplete sequence data.

An optical map of *F. oxysporum f. sp. lycopersici* from the sequenced strain was constructed using the restriction enzyme *Bsi*WI [3], consisting of 15 single optical map contigs (~63 Mb) with an average of 55X coverage. Alignments were made between optical maps and the *in silico* maps of the sequence scaffolds based on the restriction recognition sequence and the lengths of the restriction fragments using map aligner software developed at the Broad Institute. The optical maps and sequence assemblies were highly congruent with more than 96% of the assembled scaffolds aligned to the optical map. The assembled sequence scaffolds were also ordered and oriented, and the gaps were estimated based on the alignment of the assembled sequence scaffolds to the optical maps (Table S2). Overall, more than 91% of the optical maps were covered by assembled sequence scaffolds.

The genetic linkage group maps for *Fv* and the optical linkage group maps for *Fol* can be accessed at:http://www.broad.mit.edu/annotation/genome/fusarium_group.3/maps/Index.html

B. Annotation.

B.1. Repetitive sequences and transposable elements.

Repeat sequences were detected by searching the genome sequence against itself using CrossMatch, filtering for alignments longer than 200 bp with greater than 60% sequence similarity. Full-length transposable elements were annotated using a combination of computational predictions based on BLAST analysis for transposase genes and manual

inspection. Their genome distribution was characterized using the sensitive mode of RepeatMasker version open-3.0.8, with cross_match version 0.990329, RepBase Update 9.04, RM database version 20040702.

B.2. Predictions of protein coding genes.

To improve annotation accuracy, the gene structures of the *Fol*, *Fv,* and *F. graminearum (Fg)* genomes were predicted with a combination of manual annotation, FGENESH, GENEID, and FindORFs. The manually curated gene sets were generated at the Broad Institute and used to train the *ab initio* gene prediction programs FGENESH and GENEID. The genome sequences of the three *Fusarium* species were then analyzed with both FGENESH and GENEID to provide an initial set of predicted genes. ESTs from *Fg* (68,000) and *Fv* (87,000) were analyzed with FindORFs, an EST-based gene finding program developed at the Broad Institute, to identify mis-annotated gene loci and improve gene structure predictions. In addition, the final gene set of *Fg* was mapped onto the *Fv* and *Fol* genomes by FindORFs to provide additional lines of evidence to improve the automated gene prediction in these two fungi. For *Fg*, the gene set predicted at MIPS [5] also was included to improve gene annotation. Furthermore, targeted manual annotation was carried out for loci in cases where automated gene predictions were inconsistent or where there was conflict with EST evidence or BlastX analysis. The re-annotation of *Fg* greatly improved the gene structures and greater than 90% accuracy was reached when comparing this annotation against a set of 4000 high confidence genes.

***B.3.* Characterization of protein families.**

Transcription factors were identified using the data analysis pipeline in the Fungal Transcription Factor Database (FTFD; http://ftfd.snu.ac.kr/) [4]. **Carbohydrate-active enzymes** were identified by searching against a library of catalytic and carbohydrate-binding modules of carbohydrate-active enzymes (see also www.cazy.org) [5] using blastp. T**ransporters** were identified by searching for the PFAM domains representing ABC transporters (PF00005, PF00664, PF01061) and assisted with manual inspection. The potential **secreted proteins** were predicted using SignalP 3.0 D score (http://www.cbs.dtu.dk/services/SignalP/). The trans-membrane proteins were removed using the TMHMM software (http://www.cbs.dtu.dk/services/TMHMM/) and the Phobius predicted transmembrane proteins (except in the first 50 amino acids). We further removed putative mitochondrial proteins using TargetP mitochondrial target peptide prediction RC score 1 or 2. **Small cysteine-rich proteins**

were identified among predicted secreted proteins that are less than 200 aa in length and contain at least 4% cysteine residues. **GPI-anchored proteins** were identified by the GPI-anchor attachment signal with a PERL script among the predicted secreted proteins. **Kinomes** were characterized using a multi-level hidden Markov model (HMM) library of the protein kinase superfamily under the HMMER software suite (v. 2.3.2, http://hmmer.janelia.org), correcting for database size with the '-Z' option. The automatically-retrieved sequences were individually inspected and protein kinase homologies were determined by building kinase group-specific phylogenetic trees with the annotated kinomes of *S. cerevisiae*, *S. pombe* and *Encephalitozoon cuniculi* [6]. **G proteins** were identified from the best hits using the corresponding *N. crassa* protein, using a cut-off score of 1e-12 during BLAST searches. For **G protein coupled receptors**, the proteins listed are the best hits (E-value cutoff: 1e-09) using the *N. crassa, M. grisea* or *F. graminearum* proteins as queries in BLASTp searches. For GprK homologues *A. nidulans* AN7795, *N. crassa* NCU09883 and *F. graminearum* FGSG_04628 were used for searches (E-value cut-off 1e-09) with BLASTp. All GPCR hits have 7-transmembrane (7-TM) helices, with the N-terminus outside and the C-terminus on the cytoplasmic side, as verified by HMTMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/), TMPRED (http://www.ch.embnet.org/software/TMPRED_form.html), PHOBIUS (http://phobius.binf.ku.dk/) or TOPPRED (http://bioweb.pasteur.fr/ seqanal/interfaces/toppred.html), except as noted in parentheses (supplemental results). Deviations from the 7-TM structure may be due to mis-annotation or sequencing errors and many of the e values were much lower than 1e-09. **Pathogenicity and virulence factors** were identified using the pathogen-host interactions database [8] (http://phi-base.org) and through additional searches of the peer-reviewed literature.

C. Features common to *Fusarium* genomes.

Orthologous gene sets were determined based on pair-wise syntenic alignments that projected genes from one genome onto the other. All the data are available through the website: http://www.broad.mit.edu/annotation/genome/fusarium_group.3/OrthologSearch.html

The protein sequences for *Fusarium solani* [9] (telomorph: *Nectria haematococca*), used as outgroup for phylogenetic relationships among the *Fg, Fv,* and *Fol* genomes, was downloaded from the Joint Genome Institute (http://genome.jgi-psf.org/Necha2/Necha2.home.html). For a full list of annotated genes for each category, see web site: http://www.broad.mit.edu/annotation/genome/fusarium_group/SupplementaryPage.html.

Compared to other Ascomycete genomes, these three-species orthologs are enriched for predicted

transcription factors (p=2.6e-6), lytic enzymes (p=0.001), and transmembrane transporters (p=7e-9) (Tables S3-8).

C.1 Transcription factors.

We identified 2,280 transcription factors (TFs) in the three *Fusarium* genomes, which were classified into 46 TF families (Table S4). For a full list of annotated genes for each category, see: http://www.broad.mit.edu/annotation/genome/fusarium_group/SupplementaryPage.html. The most abundant TF family in *Fusarium* genomes is Zn2Cys6, a family of fungal specific transcriptional regulatory proteins that contain an N-terminal Cys-rich motif and play essential roles in both primary and secondary metabolism, drug resistance, and meiotic development [10]. The numbers of TFs in most families were similar across the three species. However, the numbers of TF families of bHLH, Zn2Cys6 and bZIP were significantly higher in *Fol* than in the other species. For example, in the bHLH TF family, *Fol* has 46 TFs, while the other species have only 16 (*Fg*) or 15 (*Fv*). Similarly, the number of TFs in the bZIP family was 55 in *Fol*, whereas those in *Fg* and *Fv* are 21, and 19, respectively. The higher numbers in these classes as well as higher numbers of homeodomain-like transcription factors can be explained by the addition of genes in *Fol* LS regions.

C.2. CAZY enzymes.

We conducted a comprehensive analysis of various carbohydrate-active enzymes (glycosidases, glycosyltransferases and polysaccharide lyases) encoded by the three *Fusarium* (Table S3) and other fungal genomes in a standardized manner to allow objective comparisons. For a full list of annotated genes for each category, see web site: http://www.broad.mit.edu/annotation/genome/fusarium_group/SupplementaryPage.html.

Among the 13,332 predicted genes for *Fg*, 14,179 for *Fv* and 17,735 for *Fol*, we detected, respectively, 243, 300, and 364 genes encoding glycoside hydrolases (GHs) and 102, 106, and 123 genes encoding glycosyltransferases (GTs). Overall, the three fungi have a complete set of plant cell wall degrading enzymes (including enzymes for the digestion of three major categories: cellulose, hemicelluloses and pectins). Notably, there are more pectin and pectate lyases from families PL1 and PL3 in *Fusarium* genomes than in other fungi. Several GH families have greatly

expanded in *Fol*. In comparison to the highly variable number of GHs, the number of GTs and the population of each GT family is much more constant. This observation is in line with the fact that while most GHs are secreted to the environment, and therefore co-evolved with the ecological niche of the fungi, GTs are dedicated to tasks that are less dependent on the outside medium, such as synthesis of the fungal cell wall polysaccharides, N- and O-glycosylation of proteins, and metabolism of storage carbohydrates.

The over-representation of these lytic enzymes is the combination of the increase of copy numbers of some gene family as well as the unique presence of some gene families. For instance, all three *Fusarium* genomes contain the highest number of pectate lyases (21,23, 24) comparing to less than 10 in other genomes. The PL family uniquely present in *Fusarium* genomes includes PL9, while the expansion is observed for families PL1, PL3, and PL4 (Table S5). The detailed comparison can be found at online supplementary data at: http://www.broadinstitute.org/annotation/genome/fusarium_group/SupplementaryPage.html

C.3 Transporters:

All three *Fusarium* species have a considerably higher number of ABC transporters than other fungi (Table S3). Typically, most fungi have 30–45 ABC transporters encoded in their genomes, regardless of whether they exist as a yeast, saprophyte, or plant/animal pathogen. However, the number of *Fusarium* ABC transporters ranges from 54 in *Fg* to 67 in *Fol* (Table S3). All five families of ABC transporters are represented in the *Fusarium* (Table S6). *Fol* has an 18% increase in the number of ABC transporters when compared to the number of transporters in *Fv* and an even larger increase when compared to *Fg*. For example, *Fol* contains four copies of a partial ABC transporter in the *Fol* LS regions (FOXG_16340.2, FOXG_07314.2, FOXG_06754.2, and FOXG_06510.2). Two of these copies are on the same chromosome (chromosome 3), while one is on chromosome 6, and the fourth is on the unpositioned supercontig 34. The larger number of ABC transporters may explain the increase in host range of *Fo* as a species as they may provide tolerance to antimicrobials generated by different hosts.

There are five ABC transporters present in all the sequenced *Fusarium* isolates, which are not present in any other sequenced fungal genome (Table S6, colored in red). The genes surrounding some of the ABC transporter genes are conserved and so may bring insight into their function. Since the genes are orthologous, only the *Fg* name will be used for this discussion. FGSG_11028.3 is similar to YOR1 transporters, residing in a cluster of genes, which are co-

expressed in plants [11] and contain the gene *NPS1* which encodes a siderophore that has yet to be chemically characterized [12.] The juxtaposition of the two genes suggests a scenario in which the ABC transporter may be partially responsible for efflux of the siderophore. Similarly, both FGSG_11028.3 and another Fusarium specific ABC transporter, FGSG_06141.3, are adjacent to a gene for a major facilitator transporter in all three *Fusarium* sp. The other two *Fusarium*-specific transporters (FGSG_03805.3 and FGSG_03032.3) are similar to heavy metal tolerance transporters (HMT). FGSG_10577.3 is a Pleiotropic Drug Resistance (PDR) protein transporter unique to *Fusarium*.

The family of PDR transporters (also referred to as the ABCG family) is largely responsible for tolerance to otherwise toxic compounds, including azoles, cyclohexamide, and plant secondary metabolites [13]. Moreover, the only pathogenicity related *Fusarium* ABC transporters, *F. sambucinum* GpABC1, which is responsible for tolerance to the potato antimicrobial compounds rishitin and lubimin [14], and *F. culmorum* FcABC1 [15] belong to this family. All three sequenced isolates have orthologs of this protein (FGSG_04580.3, FVEG_11089.3, and FOXG_13653.2). This family has more members in each of the *Fusarium* sp. than most other fungi (Table S3), such as *Neurospora crassa* (7 members), *Magnaporthe grisea* (8), *Aspergillus fumigatus* (15), and *A. nidulans* (16). In addition to the overall large number of PDR genes in all three *Fusarium* genomes, there is a 26% increase in these transporters in the broad host range *Fol* when compared to the *Fv* and *Fg*. The increased number of these proteins in all *Fusarium* species may explain why they are not typically inhibited *in vitro* with miconazole, itraconazole, or fluorocytosine [16].

C.4 Signaling.

Previous studies have demonstrated high conservation of G protein subunits and predicted G-protein coupled receptors (GPCRs) in filamentous fungi [17]. The three *Fusarium* species contain one protein corresponding to each of the three classes of Gα subunits of heterotrimeric G proteins found in other filamentous fungi (Table S7). However, *Fol* possesses an additional protein (FOXG_08807) with weak homology to the Group II Gα subunit. This protein is not similar to Gpα4 from *Ustilago maydis* [18], raising the possibility that certain filamentous fungi have evolved diverse Gα proteins. In *N. crassa*, the Group II Gα *GNA*-2 is required for mass accumulation on poor carbon sources and also appears to play a compensatory role in the absence of the Group I or Group III subunit [19]. Therefore, the additional Group II subunit in *Fol* may be required for growth on alternative carbon sources and may also provide additional genetic buffering, should

the Group I or Group III subunit genes become nonfunctional. *Fg*, *Fol* and *Fv* have two Gβ subunits; one homologous to canonical Gβ proteins identified in other filamentous fungi and another similar to *N. crassa CPC*-2, a RACK-1 homologue (Table S7). Like other filamentous fungi, the three *Fusarium* species also contain one Gg subunit. These three species do not possess homologues of the Gβ (*Krh1*/*Krh2*) or Gg (*Gpg1*) mimics identified in *S. cerevisiae*. RGS (Regulator of G protein signaling) proteins are negative regulators of Gα proteins in eukaryotes. *A. nidulans*, *N. crassa*, and *M. grisea* carry one copy of the RGS protein *FlbA*. In contrast, two copies of *FlbA* were identified in *Fg* and *Fv* genomes, and, notably *F. oxysporum* possesses eight *FlbA* -like genes.

The three *Fusarium* genomes all possess the nine classes of seven transmembrane-helix (7-TM), predicted G protein coupled receptors (GPCRs) present in filamentous Ascomycetes [17] (Table S8). The number of pheromone, carbon-sensing, *Stm*-1 related and rat growth hormone releasing factor-related GPCRs are similar in the three *Fusarium* species and other sequenced filamentous fungi. However, the *Fusarium* genomes contain more members of several other classes of predicted GPCRs. The *Fusarium* species possess a greater number of cAMP receptor-like GPCRs; *Fg* and *Fv* have four and *Fol* has five homologous proteins (Table S8). While *Fg* only has two homologues of *Homo sapiens* membrane progestin receptor (mPR)-like GPCRs, *Fv* has three and *Fol* has seven. The mPR receptors couple to G proteins on the plasma membrane and down-regulate adenylyl cyclase in response to progesterone exposure in reproductive tissues of animals [20]. The GprK/AtRGS1-like GPCR, which has 7-TMs and a RGS domain, has not been demonstrated to function as a GPCR in fungi [7,17]. Nevertheless, *Fg* has one GprK homologue and *Fol* and *Fv* each have two proteins. The *Fusarium* species possess a greater number of microbial opsins/opsin-related proteins [21] than other filamentous fungi: three in *Fg* and *Fv* and four in *Fol*. Finally, proteins homologous to the pathogenesis-related *PTH11* GPCR are abundantly present in all three *Fusarium* species, with 99 in *Fg*, 98 in *Fol* and 90 in *Fv*, far more than those in *Magnaporthe grisea* (60), *Aspergillus nidulans* (70) and *N. crassa* (25).

D. Secondary metabolism gene clusters

Fusarium produces numerous secondary metabolites that are diverse in biosynthetic origin, structure and biological activity. Some exhibit toxicity to mammals and as a result are considered mycotoxins. The two *Fusarium* mycotoxins of greatest concern are trichothecenes and fumonisins. Trichothecenes inhibit eukaryotic protein synthesis and induce hemorrhaging and

feed refusal in livestock [22]. Fumonisins disrupt sphingolipid metabolism, cause multiple animal diseases [23], and are epidemiologically associated with esophageal cancer and neural tube defects in humans [24,25]. Thus, the presence of these mycotoxins in cereal grains and grain-based food and feed poses a potential hazard to human and animal health. *Fg* and *Fv* are among the most important mycotoxin-producing *Fusarium* species because of their worldwide occurrence and because they produce trichothecenes and fumonisins, respectively.

In fungi, genes responsible for secondary metabolite biosynthesis are often clustered. The clusters typically include a terpene synthase (TS), polyketide synthase (PKS), or nonribosomal peptide synthetase (NRPS) gene that is responsible for a fundamental step in metabolite biosynthesis. Clusters also typically include the core genes responsible for structural modifications of initial metabolite, for metabolite transport, and for coordinated transcriptional regulation of cluster genes.

Few secondary metabolite biosynthetic gene clusters have been identified and functionally characterized in *Fusarium*. Given the structural diversity of secondary metabolites produced by the genus as a whole, many more *Fusarium* clusters are likely to be present in the genomes of these fungi. Here, genome sequences of *Fg*, *Fo*, and *Fv* were used to identify potential genes clusters associated with PKS and TS genes. Such analyses may provide insight into conservation of secondary metabolic pathways among species, into enzyme and cluster evolution, and into the functions of secondary metabolites in the ecology of producing species.

To identify secondary metabolite biosynthetic (SMB) gene clusters in *Fusarium*, an automated prediction program was developed based on the presence of pfam domains described for PKSs and other proteins involved in fungal secondary metabolism. A significant initial effort was made to identify all PKS genes present in each of the four *Fusarium* genomes. This effort used a combination of BLAST and text search analyses of preliminary annotations at the Broad and JGI websites. The BLAST query sequences were predicted keto synthase domains from reducing and non-reducing (NR) PKSs, the two major classes of fungal PKSs. These analyses provided a preliminary list of putative KS-encoding DNA from the four species. Nucleotide sequences flanking the putative KS-encoding DNA were subjected to BLASTX analysis and manually annotated to identify the full-length PKSs. Since most secondary gene clusters contain genes with common domains related to Transporter; Cytochrome P450; oxidoreductase; Zn(2)-Cys(6) transcription factor; Oxidoreductases; FAD binding domain; Transporter; O-methyltransferase; Transferase family, we searched the putative SMB gene clusters by pfam domains analysis of 20

kb flanking either side of each PKS genes. PKS gene-flanking regions that included minimum of four pfam domains were considered putative SMB clusters and subjected to microarray analysis using available microarray data from both *Fg* and *Fv*. Conservation of the putative clusters among the four sequenced *Fusarium* genomes was used to further define the cluster boundaries. A similar approach was also taken to identify terpenoid biosynthetic gene clusters. Briefly, BLAST analysis was used to identify putative terpene synthetase (TS) genes, and the regions flanking these genes were then subjected to the pfam analysis described above.

Phylogenetic relationships of *Fusarium* PKSs within and among species were examined by maximum parsimony analysis of deduced amino acid sequences of the ketosynthase (KS) and acyl transferase (AT) domains of all the PKS genes identified in the *Fusarium* genomes. Because of a lack of conservation in 95 to160-amino acid region between the KS and AT domains, this inter-domain region was removed from the alignment. The aligned sequences were then used to construct a gene genealogy using parsimony in PAUP* 4.0b10 [26]. Bootstrap values were based on 1000 pseudoreplications. Phylogenetic trees generated from deduced amino acid sequences of other functional domains in the PKSs were very similar to each other and to the tree based on the combined KS and AT domains.

A total of 39 PKS and 7 TS related SMB gene clusters were identified (Table S9), including 18 from *Fg*, 16 from *Fv* and 12 from *Fol*. The average cluster contains eight genes and spans over 24 kb. Despite its increased genome size and large protein encoding gene set, *Fol* has the least SMB gene clusters identified through this approach.

Microarray analysis confirmed co-regulation of genes in 14 of 18 *Fg*, and 10 of 16 *Fv* clusters (online materials). Differential expression of *Fg* clusters was examined under multiple conditions, including during sexual reproduction, infection of multiple plant hosts, and growth in complete medium. *Fg* SMB clusters expressed *in planta* include the previously described trichothecene (FG3_21) and fusarin C clusters (FG3_19). Six SMB clusters (FG3_15, FG3_20, FG3_28, FG3_35, FG3_40 and FG3_45) were also preferentially expressed during perithecial formation and sexual reproduction. Cluster FG3_33 is responsible for biosynthesis of aurofusarin, a red pigment commonly observed in laboratory cultures of *Fg*. Indeed FG3_33 cluster genes were expressed exclusively in culture, and closely related orthologs of cluster genes were absent in the genomes of the other fusaria examined. Two clusters (FG3_12, FG3_13) are constitutively expressed and four, including the zearalenone cluster (FG3_34) were poorly expressed under all

conditions tested (online materials). The lack of expression of this known cluster is attributable to the fact that zearalenone inducing conditions were not included among microarray tests.

Among the identified PKS SMB gene clusters, only 2 (FG3_15, and FG3_28) appear to be conserved in all species examined (Table S9). This finding is consistent with the limited similarity shared by the PKSs (Figure S5). Surprisingly, both of the orthologous PKS clusters (FG3_15, and FG3_28) are preferentially expressed during perithecial formation. Cluster FG3_15 includes the PKS gene *PGL1* responsible for biosynthesis of the dark pigment in the perithecial wall in *Fg* and *Fv* [27]. It is interesting to note that *F. solani* has orthologs of *PGL1* (JGI_101778) and of cluster FG3_15. However, another non-reducing PKS gene, *PKSN* (JGI33672), is responsible for the red perithecial pigmentation in this species [28]. The conservation of the *PGL1* cluster in the asexual species *Fol* places in doubt the direct functional role of this cluster in the sexual reproduction processes. However, the *pgl1* mutants form un-pigmented perithecia with viable ascospores demonstrating that the dark pigments produced through such pathway is not fundamentally essential for sexual reproduction/perithecial formation, and instead may protect the organism from UV light, stress, and oxidative damage. We also noticed that one centrally located gene unique to the FG3_15 cluster, FGSG_09185 (adenylosuccinate lyase), is constitutively expressed and not regulated with the other genes of the cluster. This suggests that the cluster may have acquired this gene relatively recently, subsequent to the evolution of co-regulation of the rest of the cluster.

Interestingly, the non-reduced PKS encoding SMB gene clusters are uniquely expanded in the *Fg* genome (5, 2, 2 in *Fg*, *Fv*, and *Fol* respectively). This group of PKS gene clusters includes the previously characterized biosynthetic clusters for the mycotoxin zearalenone (FG3_34) and the red pigment aurofusarin (FG3_33) [27,29]. One *Fg* specific expanded gene cluster (FG3_26) include 17 genes, are expressed specific *in planta* (Figure S5 and online materials). In contrast, clade III of the reducing PKS encoding clusters including the known fumonisin biosynthetic gene cluster (FV3_21) is uniquely expanded in *Fv*, (1:5:3 in *Fg, Fv* and *Fol* respectively). The two *Fv* specific clusters have 8 and 13 genes identified and span 24 and 33 kb. The only *Fg* clade III cluster is expressed during sexual development [30]. The second clade of the reduced PKS gene clusters has expansion in the *Fv* and *Fol* lineages over *Fg* (2:4:4), including three orthologous clusters conserved between *Fv* and *Fol*. Orthologs of the putative fusarin C biosynthetic cluster [31] is present in *Fg*, *Fv*, and *Fs*, but absent in *Fol*.

There is a clear expansion of terpenoid related SMB gene clusters in *Fg* (5, 1, 1 in *Fg*, *Fv*, and *Fo* respectively), including the well-characterized trichothecene biosynthetic cluster FG3_21. The co-regulation of several of the *Fg* TS clusters are confirmed by expression analysis (Table S9 and online materials).

Species within genus *Fusarium* have evolved extraordinary chemical diversity and produce highly complex and biologically active secondary metabolites. The function of most SMB clusters described here are unknown. These novel SMB clusters can be used to discover new secondary metabolites and their potential impacts on the biological properties of the fungi. The co-expression of most of such clusters, especially the novel clusters, supports their potential functionality and will promote the study of these novel secondary metabolites. Only a very small fraction of the identified SMB clusters are shared among species, reflecting great genetic diversity within the *Fusarium*.

E. Features of *Fusarium oxysporum* LS chromosomes

E.1 Synteny and Fol LS chromosomes.

We performed multiple genome alignments to detect synteny and large-scale genomic re-arrangements in order to understand genome evolution among the three *Fusarium* genomes. PatternHunter [32] (1e-10) was used to compute local-alignment anchors. Pairs of anchors separated by fewer than 10,000 nucleotides on both genomes were merged progressively to form co-linear blocks, which were further filtered to ensure that no block overlaps another block by more than 90% of its length. The average sequence similarity between each pair of genomes was calculated based on the syntenic alignments in 100 base pair windows (Table S10).

Over 90% of the *Fv* genome can be unambiguously aligned to the orthologous regions in *Fol* with an average 90% sequence identity (Table S10, Figure 2). The pair-wise genome comparison reveals that all eleven chromosomes in *Fv* have corresponding chromosomes in *Fol* with minimal local rearrangements and only one major chromosomal translocation event (Figure 2). In contrast, four of the *Fol* chromosomes (chr3, chr6, chr14, and chr15), parts of chromosome 1 and 2 (scaffold 27, scaffold 31), and most of the unmapped scaffolds, lack significant orthologous sequence in the closely related genomes (Figure 2), defining the *Fol* LS regions.

The smaller number of chromosomes in *F. graminearum* appears to be a result of chromosome fusion. Whole genome alignments of *F. graminearum* to *F. verticillioides* reveals that the vast majority (86% overall) of each *F. verticillioides* chromosome maps to a single *F. graminearum*

chromosome (Figure S3). The alignments display end to end synteny in large blocks, with the exception of *F. verticillioides* chromosome ends. The previously described highly polymorphic regions of *F. graminearum* [33] correspond to *F. verticillioides* chromosome ends, including internal fusion sites. This suggests that the fused subtelomeric regions in *F.graminearum* continue to evolve similarly to conserved subtelomeres, and that a sequence or epigenetic modification continues to mark the fused sites similarly to subtelomeres.

E.2 the variation of Fol LS regions between strains within the F. oxysporum species complex.

We have used both Illumina sequence of *F. oxysporum* strain Fo5176, a pathogen of *Arabidopsis* [34], and EST sequences from *F. oxysporum* f. sp. *vasinfectum* [35], a strain of *F. oxysporum* that infects cotton, to access the variation of the *Fol* LS regions within the *F. oxysporum* species complex. We have generated one run of Illumina reads for *F. oxysporum* strain Fo5176, a pathogen of *Arabidopsis*, using the Applied Biosystems and Illumina high-throughput sequencing platform Genome Analyzer. The raw data was processed by removing low quality reads using quality filter, resulting a total of 26,743,945 reads of 51 bp in length (1 Gb). The software MAQ [36] was used to map a total of 10,688,499 reads to the reference assembly of *Fol*, covering 46 Mb (60%) of the *Fol* assembly. In the aligned region the median SNP density was 14 SNPs per 1Kb, reflecting a short evolutionary distance between these two strains (Figure S13 A). Despite the low sequence divergence indicated from shared sequences, 40% of the reference genome was not represented in strain Fo5176. Measuring by the coverage of the Fo5176 reads in 10 Kb windows, we observed that the reference genome was not equally covered by the Illumina reads (Figure S13 B). By contrast, the majority of the genome (70% of the 10 Kb windows) has a coverage centering around 85% sequence alignment (ranging from 60%-100%). We found a clear enrichment of genomic regions that have a very low coverage (0-10%), most of which corresponded to *Fol*-LS regions in the reference assembly (Figure S13 B). While the genomic regions conserved between *Fol* and *Fv* had an greater than 80% coverage when compared to Fo5176, with a drop of coverage near telomeric regions of each conserved chromosomes, the *Fol*-LS regions had much lower sequence coverage (average 34.5%), indicating considerable sequence variation in the *Fol* LS regions within the *F. oxysporum* species complex. This provides direct support, analyzed at a genome wide level, for the existence of large LS-regions in genomes of members of the *F. oxysporum* species complex, as predicted by the *Fol-Fv* comparison and is consistent with studies where strain-specific sequences have been isolated in the *F. oxysporum* complex [37].

A total of 3,034 expressed sequence tags (ESTs) isolated from cotton root and hypocotyl tissues infected with *Fusarium oxysporum* f. sp. *vasinfectum* [35] were downloaded from NCBI and searched against *Fol* transcripts (blastn 1e-100). A total 198 fungal ESTs were uniquely matched to the *Fol* transcripts with high sequence similarity (an average of 99%). All these transcripts are located on orthologous chromosomes conserved between *Fol* and *Fv*. For a full list of annotated genes in each category see: http://www.broad.mit.edu/annotation/genome/fusarium_group/SupplementaryPage.html.

***E. 3 Transposable elements*.**

Many types of transposable elements [38], including both retroelements and DNA transposons, have been identified in the genome of *Fol*. Interestingly, DNA transposons were the most represented elements among the manually annotated TEs (65%), a unique situation as generally, in fungal genomes, the most abundant class of TEs are retroelements [39]. DNA transposons can be assigned to five distinct superfamilies: *Tc1-mariner* with the *pogo* and the *impala* families, *hAT* with 13 distinct families, *Mutator*-like with 5 families, MITEs with 6 distinct families and a *Helitron*-like family.

The *pogo* family, which is generally well-represented in fungal genomes [39,40] is uniquely amplified in the *Fol* genome with more than 500 copies belonging to eight different subfamilies, and representing approximately half of the DNA transposons. The copy number varies greatly from one *pogo* subfamily to another, from less than ten copies to half of the overall *Fot* copies. In the most numerous subfamilies, potentially active elements could be identified. Interestingly, a comparison of the representation of each *Fot* family in two different *F. oxysporum* strains, strain FOM24 (f. sp. *melonis*) and *Fol* suggested that amplification is strain-dependent. This is exemplified by *Fot1*, present in more than 100 copies in FOM24 [39] but only five copies in *Fol* (this work). Finally, on the basis of their small and homogeneous size, criteria commonly accepted for MITEs, we also identified 140 copies of a novel MITE, sharing TIRs with *Fot5*.

The *impala* family is characterized by a small number of copies diversified in three different subfamilies, none of which contain active elements. A family of MITEs, *mimp*, has been extensively characterized in the *Fol* genome [41] and demonstrated to be transactivated by the *impala* transposase both in *Fo* and heterologous genetic backgrounds.

The *hAT* superfamily is also well-represented in *Fol*: thirteen distinct families have been recognized, including the *Folyt* element which has already been demonstrated to be active in *Fol* [42]. The number of copies varies from four to 60 depending of the family. Four distinct families of small elements, three of which could clearly be associated with an identified full length *hAT*-like element were detected and classified as MITEs. *Hop* elements, previously characterized as *Mutator*-like TEs [43] were also identified in *Fol*. Although present in a much lower number of copies, they were classified into five distinct subfamilies.

A *Helitron* (Strider) of 6129 bp was identified in *Fol*. The ORF contains one intron and encodes an 1899 amino acid replicase/helicase/endonuclease. Four copies are disrupted by a *hAT* transposon and there are at least 15 nearly identical full length copies and another nine copies perhaps full length but truncated due to sequencing gaps or contig ends. Their high sequence similarities suggest a recent expansion of this element in the *Fol* genome.

Among retroelements, at least four different families of LTR elements were recognized or newly identified: the *Yaret2* family, a *gypsy*-like element (*Skippy*) and two *copia*-like retrotransposons (*Han*, *Gollum*), three of these composed of both full length and solo LTR elements. Non-LTR elements were also found in *Fol*: one family of LINE (*Yaret1*) and one family of SINE (*Foxy*). The SINE-like retrotransposon *Foxy* is the most abundant retroelement in *Fol*, which is in agreement with earlier AFLP analysis [44].

Transposable elements are over-represented in the LS regions of *Fol* (Table S11). More than 74% of all the identifiable TEs, including 95% of all the DNA transposable elements are present in *Fol* LS regions. This biased distribution of TEs is more significant for DNA transposons than for retro-elements, which are the main TEs present in *Fv* and *Fg* genomes. Our analysis suggests there has been a recent lineage-specific expansion of the DNA elements in association with the formation of *Fol* LS regions. A significant part of this expansion (approximately 50%) results from the presence of segmental duplications, some large genomic regions being found up to four times. For a full list of annotated genes for each category, see web site: http://www.broad.mit.edu/annotation/genome/fusarium_group/SupplementaryPage.html.

E. 4. *Segmental duplication.*

We aligned each of the 15 *Fol* optical contigs with all other optical contigs using Map Aligner [3] and discovered four large segmental duplications among *Fol* LS chromosomes, covering about 7 Mb (Figure 2C). The map alignment confidence score rewards matching restriction sites and

fragment length, and at the same time penalizes discrepancies including missing or extra restriction sites and differing length of the restriction fragments. A map alignment score of six with at least five matched restriction fragments is considered to be significant. The local map alignment confidence scores for all the four duplicated regions range from 49.99 to 269.34, and the matched restriction fragment numbers range from 58 to 210, reflecting high confidence for the aligned duplicated regions.

All of the segmental duplications detected by the optical mapping process are located in the *Fol* LS regions with high nucleotide sequence identity among the aligned regions (>99%), indicating that the observed duplications occurred very recently (Figure S7). To detect all potential segmental duplications, especially potential duplications across the conserved versus the *Fol* LS regions that may have accumulated more nucleotide changes, we searched for two genomic regions that contain a minimum of three genes that are in the same order and orientation. We identified three small duplications occurring between *FOL* LS regions (chromosome 3, scaffold 25) and telomere proximal regions of conserved chromosomes 4, 5 and 9, accounting for a total of 14 genes. These results clearly indicate that such limited exchange of genetic material between the conserved and *Fol* LS regions is not the major process involved in formation of the *Fol* LS regions.

E.5 ***Genes and gene categories expanded on LS chromosomes.***

Blast score ratio tests

Blast Score Ratio (BSR) analysis is designed for the rapid comparison of complete proteomes of three genomes, and enables a visual evaluation of the overall degree of similarity of these proteomes and their genomic structure [45]. The BLAST raw score for each *Fol* protein against itself is stored as the Reference score. Each *Fol* protein is then compared to each protein in *Fg* and *Fv* proteomes with each best BLAST raw score recorded as Query score respectively. The BSR for each reference protein is computed (Query score divided by the Reference score) with a normalization from 0 to 1. A score of 1 indicates a perfect match of the Reference to a Query sequence and score of 0 indicates no BLAST match of the Reference in the Query proteome. For each protein in the *Fol* genome, two numbers are generated from the best matches in *Fg* and *Fv*, respectively. The normalized pair of numbers were plotted using Matlab 7.0 software package. The *Fol* proteins encoded in the conserved regions versus the *Fol* LS regions exhibited very different profiles under this test (Figure 3).

Codon usage tests

Codon usage tables for each chromosome were computed using the EMBOSS tool 'cusp' (http://oryx.ulb.ac.be/embosshelp/cusp.html). The Pearson correlation coefficient (r) was calculated to determine the linear dependence of codon usage between each pair of chromosomes. While the r is close to 1 for both comparisons among the conserved chromosomes (0.999 ± 0.00167) and among the *Fol* LS chromosomes (0.998 ± 0.002), the r between conserved and *Fol* LS chromosomes is lower (0.990 ± 0.002), showing a weaker linear relationship between these two regions (Table S19).

To quantify this, we calculated the Pearson correlation coefficients of codon usage preferences between different chromosomes, and grouped them into three classes: A) between chromosomes within the conserved regions (top 10 rows in Table S19), B) between chromosomes within the LS regions (last 3 columns in Table S19), and C) between chromosomes in the conserved regions and the chromosomes in the LS regions (highlighted red in Table S19). The correlation coefficients in group A showed no difference from those in group B (p=0.30, two-sample t-test), indicating that the codon usage preferences are all highly similar within each group. However, the correlation coefficients between group A and C are significantly different (p=4.3e-38), and so are the correlation coefficients between B and C (p=1.9e-10), suggesting that the genes in the LS regions utilize a significantly different codon sets from those in the conserved regions (Table S19). The most significant differences were observed for amino acid Q, C, A, G, V, E, and T, with a preference for G and C over A and T among the genes encoded in the *Fol* LS regions (Table S20).

In addition, we computed the codon adaptation index (CAI) using the the EMBOSS tool 'cai' (http://oryx.ulb.ac.be/embosshelp/cai.html) as the geometric mean of the relative synonymous codon usage (RSCU). The CAI distribution for the genes encoded in the conserved region of *Fol* is almost identical to the CAI distribution of genes encoded in the *Fv* genome, however, the CAI distribution of the genes encoded in the *Fol* LS regions is shifted toward the lower end (Supplementary Figure S10).

Cross-kingdom proteome comparison suggests that the *Fol* LS regions are of fungal origin. Among the predicted proteins sharing homologous sequences to the protein set at the NCBI database (excluding *Fusarium* proteins), less than 2.6% of the LS-encoded proteins have bacterial proteins as their best hit, comparable to proteins encoded in the conserved regions (2.3%). Significantly, 93% have their best BLAST hit to other Ascomycete fungi (Figure S12). Using present/absent as a measure of closeness among these LS chromosomes and other fungal

genomes within Ascomycetes, the closest hit is *Fs*, where HT is also reported [9]. A distance matrix to show the distance between the genes encoded in LS regions and other selected fungal genomes are shown (Table S21)

E.5. Functional enrichment of genes encoded in Fol LS regions

Gene ontology (GO) terms [46] were assigned to 7,709 out of the total 17,735 genes predicted for *Fol* (43.5%) based on BLASTP (1e-20) using the program Blast2GO [47]. A total of 595 genes out of the 3,261 (18.3%) encoded on the *Fol* LS chromosomes were assigned with 318 GO terms. Among these, only 24 GO terms are enriched (FDR <0.01) (Table S12), while 98 of them are under-represented (also see detailed results from the GO-enrichment test at http://www.broad.mit.edu/annotation/genome/fusarium_group/SupplementaryPage.html). The *Fol* LS enriched functional categories include the proteins secreted to the extracellular space, genes involved in the fungal cell wall degrading process, signal transduction (PLC, *FlbA*, and kinases), and gluconeogenesis (Figure S9). Detailed annotation of the genes encoded on chromosome 14, a *Fol* LS chromosome capable of mobilization between vegetatively incompatible strains and encoding host range determinants, also reveals enrichment for secreted proteins (9.5% of total proteins), TFs (10% of proteins with functional assignment) and function related with DNA modifications (20%). Conversely, house-keeping genes are nearly absent from this chromosome (See detailed results from chromosome 14 at http://www.broad.mit.edu/annotation/genome/fusarium_group/SupplementaryPage.html).

Secreted proteins

Among the potentially secreted proteins (See all the secreted proteins at http://www.broad.mit.edu/annotation/genome/fusarium_group/SupplementaryPage.html), we identified 253 that are encoded on *Fol* LS regions (Table S13) that comprise small cysteine-rich proteins (potential plant effectors), necrosis and ethylene-inducing proteins (NEPs) and carbohydrate active enzymes (CAZYs), all considered likely to play a role in pathogenesis. In contrast, the *Fol* LS regions contain none of the 137 predicted to be wall-anchored proteins. Small, cysteine-rich proteins produced by fungi may be of ecological relevance by functioning as toxins, inhibitors of hydrolytic enzymes secreted by other organisms (e.g. proteases and chitinases) or by suppressing defense responses of host plants. Importantly, peptides secreted by the fungus within a host may be recognized by the plant innate immune system, triggering disease resistance. Indeed, almost all avirulence factors of fungi identified to date are small, cysteine-rich

proteins [48,49]. In fact, the three avirulence factors of *Fol*, corresponding to the three known resistance genes in tomato, are small proteins secreted in xylem ("SIX" proteins), each having between two and eight cysteine residues [50]. At least two AVR genes (probably all three) are encoded on LS chromosomes [37,50]. Seven additional SIX proteins also have between two and eight cysteines and are encoded by LS chromosomes; several also have roles in virulence and/or enhancement or suppression of disease resistance response.

Other secreted proteins uniquely encoded on the LS chromosomes include plant necrosis–inducing proteins (NEPs) that can act as virulence factors during infection of plants [51]. NEPs activate host defense-associated responses, including accumulation of pathogenesis-related (PR) proteins and reactive oxygen species [49], that result in cell death. Remarkably, three NEP genes from *Fol* located on LS regions (FOXG_14409.2, FOXG_15072.2 and FOXG_17014.2) are closely related to "wilt-specific" gene lineage containing only genes from the vascular pathogens *F. oxysporum* and *Verticillium dahliae* (AAS45247.1). Possibly, the *F. oxysporum*- and *Verticillium* specific NEP genes contribute to the development of the characteristic wilt symptoms of infected plants.

Secreted carbohydrate active enzymes (CAZYs) have multiple functions in virulence, including host penetration, nutrient acquisition and pathogen cell wall remodeling. The *Fg* LS chromosomes are significantly enriched for glycoside hydrolase (GH) genes, resulting from expansion of several families of GHs that are important for degrading plant cell wall components (Table S3 and S14. Also see detailed annotation of all CAZY at http://www.broad.mit.edu/annotation/genome/fusarium_group/SupplementaryPage.html). The expansion of secreted GHs and their enrichment on the LS chromosomes suggests that the acquisition of LS chromosomes by *Fol* has increased the diversity of its emzyme arsenal against major plant cell wall components, such as pectins or arabinogalactans. This may well be related to the broad host range of *Fol*.

To test whether these GH genes are expressed during early stages of plant infection, reverse transcriptase PCR was performed on total RNA isolated from infected tomato roots at three and seven days after inoculation (Figure S8 and Table S15). The analysis included GH families previously implicated in fungal virulence such as β-glucuronidase (GH79), β-1,3-glucanase (GH64), polygalacturonase (GH28) or chitinase (GH18), among others (Table S15). For most enzyme families, two classes of members were included: 1) *Fol* specific genes located on LS chromosomes, most of them present in multiple copies (e.g. FOXG_06568, FOXG_12407,

FOXG_12535, FOXG_15491), and 2) single orthologous copy genes present in all three *Fusarium* species (e.g. FOXG_02657, FOXG_09503, FOXG_13051, FOXG_00921). We detected transcripts of all the orthologous GH genes and approximately half of the *Fol* LS genes tested. By contrast, we failed to detect transcripts of any of the four GT genes included in the analysis (Figure S8 and Table S15).

Among the transcripts detected *in planta* are two *Fol* LS specific chitinases (FOXG_14329, FOXG_15151) and two *Fol* LS specific genes encoding a chitin-binding protein (CBM12) that shares homology only with bacterial genes (FOXG_15172, FOXG_14308). Interestingly, the LS chromosomes contain four copies (FOXG_14840, FOXG_16128, FOXG_06883, FOXG_16251) of an endochitinase gene orthologous to *ech-42* from the fungal mycoparasite *Trichoderma harzianum* [52]. Expansion of chitinases and other fungal wall degrading enzymes on LS chromosomes may reflect selection for competitiveness of *Fg* with other cohabitant fungi in its main habitat, the plant rhizosphere.

Expression of genes for secreted CAZYs in planta. Two week-old tomato plants were inoculated with microconidia of *F. oxysporum* f.sp. *lycopersici* strain 4287 as described [52]. Total RNA was isolated from roots of non-infected and infected plants three days after inoculation, using the RNeasy Plant Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. Quality and quantity of extracted RNA were verified by running aliquots in ethidium bromide stained agarose gels and by spectrophotometric analysis in a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, USA), respectively. The isolated RNA was treated with deoxyribonuclease I (DNase I, Fermentas, USA) and reverse transcribed into first strand cDNA with ribonuclease inhibitor RNasin Plus RNase inhibitor (Promega, USA) and M-MLV reverse transcriptase (Invitrogen S.A., Spain) using a poly-dT antisense primer. Gene-specific primers were used for PCR amplification (Table S16). The *F. oxysporum* actin 1 gene was used as a positive control for gene expression, and *F. oxysporum* genomic DNA was used as a control template for size comparison with bands amplified from cDNA of intron-containing genes. PCR products were electrophoretically separated in 2% agarose gels, stained with ethidium bromide and photographed.

Peroxidases.

Three genes on LS-chromosomes encode catalase-peroxidases (EC 1.11.1.6) (FOXG_14234, FGXG_17106, and FOXG_17130), a subgroup of heme-dependent peroxidases. This group comprises secretory peroxidases that exhibit both peroxidase and catalase activities and provide

protection to cells under oxidative stress [53]. Plant cells in contact or close proximity to the pathogen usually respond defensively with an oxidative burst. This may lead to direct damage to the pathogen, plant cell wall reinforcement and/or programmed cell death. The catalase-peroxidases of *Fol* may have important roles in reducing the effects of oxidative stress. The three copies of these genes on LS chromosomes in *Fol* are paralogs of a syntenic copy (FOXG_17180) found in all species indicating a specific expansion within *Fol*.

Lipid metabolism and signaling.

Species-specific expansions are observed in *Fol* among gene families that are important in lipid metabolism and generation of lipid-derived second messengers. Particularly, we observed a large expansion of genes for the perilipin-like Cap20 protein that regulates the release of non-esterified free fatty acids (NEFA), as well as phospholipase C (PLC) genes that generate the signaling molecules inositol 1,4,5-triphosphate (IP3) and diacyl glycerol (Table S17). These lipid-derived molecules play crucial roles in intra- and inter-cellular signaling [54] and regulate pathogenic traits required for the development of fungal diseases [55].

There is an increasing body of knowledge about how perilipin-like proteins regulate lipid homeostasis in mammals [56,57], however, little is known about the factors that mediate lipid storage and control lipid homeostasis in fungi. In mammalian systems, Perilipin A [25] is a lipid droplet-associated phosphoprotein that acts dually as a suppressor of basal (constitutive) lipolysis and as an enhancer of cyclic AMP-dependent protein kinase (PKA)-stimulated lipolysis through controlling the accessibility of lipase to the hydrophobic triglyceride core [57]. Such control is mediated through a G protein-coupled receptor (GPCR) that activates adenylyl cyclase and, consequently, protein kinase A (PKA) [58]. Adipose triglyceride lipase and Hormone-sensitive lipase are the major enzymes involved in adipose tissue triacylglycerol catabolism [59] and promote the release of non-esterified free fatty acid (NEFA). The resulting NEFA can be used for energy generation in beta-oxidation, membrane phospholipid synthesis, signaling, and in regulation of transcription factors such as the peroxisome proliferator-activated receptors (PPARs) [56,57].

As pathogenicity determinants, perilipin-like *CAP*20 (first identified in the phytopathogenic fungus *Colletotrichum gloeosporioides)* and its homologous gene *MPL1* (*Metarhizium anisopliae* perilipin-like *MPL1*) were uniquely induced by a host signal during appressorium formation and localized at the infection front and within infected host tissue [60]. A single copy of this gene is detected in pezizomycotinal fungi of the phylum Ascomycota, but is absent in yeasts and other

fungi. As anticipated, three *Fusarium* genomes share an orthologous *CAP20* gene. However, the *Fol* genome contains eight additional copies of *CAP20* homologs in the LS genomic regions. In addition to the expansion of the *CAP20* homologues in *Fol* LS genomic regions, we also observed the expansions of lipases that act on carboxylic esters, a family of GPCRs, and a group of Regulator of G protein Signaling (RGS) proteins among these regions as well (Table S17). The expanded GPCRs share sequence similarity to human membrane Progestin Receptor (mPR). In addition to two ancestral copies conserved among other fungi, *Fol* encodes another four mPR-like GPCRs in the LS regions. The mPR-GPCRs transfer signals to G proteins at the plasma membrane and down-regulate adenylyl cyclase [20]. Even though homologues of human mPR-like proteins have not been functionally characterized in any fungus [17], based on mammalian systems, we predict that they could regulate lipid homeostasis through *Cap20*. In connection with regulating Gα proteins and presumably acting upstream of PKA, we observed the expansion of one of the RGS, a *FlbA*-like protein. *FlbA*-like gene products may act as GTPase-activating proteins to negatively control Gα protein signaling [61]. While most sequenced fungal genomes contain a single *FlbA* gene, two orthologous copies were identified in the three *Fusarium* genomes, while the *Fol* LS regions encode six additional *FlbA*-like genes. Another key element of lipolysis, the lipase gene family, is also expanded in the LS regions, parti cularly the genes for lipases that act on carboxylic esters.

The PLC gene (YPL268W in yeast) encodes a calcium-dependent phospholipase that hydrolyzes phosphatidylinositol 4,5-biphosphate (PIP2) to generate the signaling molecules diacylglycerol DAG and inositol 1,4,5-triphosphate (IP3). Both DAG and IP3 serve as secondary messengers that regulate cellular processes such as transcription, mRNA export, DNA repair, chromatin remodeling, glucose-induced calcium signaling, telomere elongation, vacuolar biogenesis, stress responses and cell wall synthesis [62]. DAG also been implicated as a second message in the plant pathogenic fungus *M. grisea* where it promotes formation of appressoria, structures required for infection, on normally non-conducive hydrophilic surfaces [63]. In addition to one orthologous PLC1 gene encoded in each *Fusarium* genome, there are six copies in the *Fol* LS regions. It is noteworthy that four Fol PLC genes, FOXG_07102, FOXG_07611, FOXG_15032 and FOXG_14448, have a Pleckstrin homology (PH)-like domain at the amino terminus, which is lacking from other fungal PLC genes, such as those in *Neurospora* and yeast. PH-like sequences, thought to target PLC to the plasma membrane through binding to PIP3, the phospholipid product of PI-3 kinase, are also found at the amino terminus of human b and d isozymes. The presence of

this domain might be important for targeting the *Fusarium* PLCs to the plasma membrane in certain cell types.

E.6. RIP and other silencing pathways.

One clear difference among the three genomes is Repeat Induced Point mutation (RIP), an Ascomycete-specific defense mechanism that recognizes and inactivates duplicated segments during the sexual cycle by C:G to T:A mutations [62,64]. RIP was recently shown to operate on introduced repeated sequences in the homothallic *Fg* [33], and a severe form of RIP was observed in a group of skippy-like elements in *Fv* (Figure S18). Four almost full-length SLREs are found on chromosomes 8, 9, 11 and on an unmapped contig. None of these elements are predicted to be functional, but based on the mutational profile, the *Fol* and *Fv* SLRE elements shared a common ancestor. The *Fol* SLRE, with nine full-length or almost full-length copies, shows no evidence of RIP. The predicted proteins are relatives of "Maggy" from *M. grisea* [65]. No C:G to T:A mutations are found in *Fol* within nine individual elements that are located on chromosomes 1, 3, 4, 6, 10, 14, 15, and in unmapped contigs (Figure S19a). The ~240-bp 5'- and 3'-LTRs are retained in all nine copies (Figure S19 b, c). Because this element is found on both LS and non-LS *Fol* chromosomes, these elements either were introduced to *Fol* with the LS chromosomes or invaded even more recently.

If not all *Fusarium* species can make use of RIP to inactivate transposable elements, how do they control the movement of such "selfish DNA" or at least dampen their effects? Two additional well-studied silencing phenomena, DNA methylation and RNAi-related mechanisms akin to "quelling" in *Neurospora* [66] may help to inactivate TEs during asexual reproduction and vegetative growth (Table S18). All *Fusarium* species have the potential to express a *Neurospora* DIM-2 homologue but to date ESTs for this gene has only been found in *Fv*. Two additional genes that encode proteins required for DNA methylation in *Neurospora*, the histone H3 K9 methyltransferase DIM-5 [67] and the heterochromatin proteins HP1 [68] are present in all three *Fusarium* species (Table S18). ESTs from five of the six genes have been found. The one gene without ESTs, *Fg Hpo*, has been expressed as a fusion protein with both GFP and RFP and its localization is in heterochromatic foci as observed for *Neurospora* HP1 (L. Connolly, P. Phatale and M. Freitag, in preparation). While the sequenced *Fg* strain and another frequently used strain (Z-3639) [69] have very little DNA methylation, both *Fol* and *Fv* strains show different banding patterns when examined with a pair of isoschizomers that are insensitive (*Dpn* II) or sensitive (*Sau3A* I) to cytosine DNA methylation (K. Pomraning, J. Whalen and M. Freitag, in

preparation). Thus methylation may be one way to control expression of TEs in *Fol* and *Fv* but further experiments to evaluate the distribution of methylated DNA in *Fusarium* species are required.

While the DNA methylation and small RNA silencing machineries appear not to be connected in *Neurospora* [68], RNA interference or quelling can act on its own to inhibit expression of aberrant RNAs or TEs [70]. This type of silencing occurs in many organisms, including those that do not methylate their DNA. All three *Fusarium* genomes contain homologues for genes previously shown to be important for quelling or meiotic silencing in *Neurospora* [71], i.e., two dicer homologues, two argonaute homologues and three homologues of RNA-dependent RNA polymerases (Table S18). Interestingly, three additional *QDE*-2 homologues are found in *Fol*, all of them on LS chromosomes (Figure S20 and Table S18). As with heterochromatin silencing pathways that rely on histone or DNA methylation, little is known about RNAi or meiotic silencing in *Fusarium* species.

F. Transfer of LS chromosomes between strains.

F.1 Transformation of Fusarium strains

To select for transfer of chromosome 14 from into Fom, Foc, Fo-47 or Fol, the receiving strains were transformed with a hygromycin (*HYG*) resistance gene, inserted randomly into the genome - three independent hygromycin resistant transformants were selected for the experiments. The hygromycin resistance gene was present in the vector pPK2-HPH-GFP which has the *GFP* gene cloned in frame behind *HPH* [72] and was transformed to Fom (NRRL 26406), Fo-47 and Foc (NRRL 25603) using *Agrobacterium* mediated transformation [73]. For transformation of Fol (Fol007) with HYG the same strategy was applied, except that a vector was used with *CFP* cloned behind *HPH* (pPK2-HPH-CFP). The donor strain (Fol007) has a phleomycin resistance gene (*BLE*, confers resistance against zeocin) inserted in chromosome 14 (B chr. 14) 180 bp downstream of the *SIX1* ORF and was obtained through homologous recombination at the *SIX1* locus, as described in [74]. All media containing zeocin were buffered with 100mM Tris pH 8.0.

F.2. Co-incubation of spores

To select for transfer of chromosome 14 from the donor strain (Fol007) into Fom, Foc, Fo-47 or Fol007, one donor and one receiving strain were co-incubated. All *F. oxysporum* strains are grown at 25 °C in the dark. For co-incubation, microconidia of the different strains were isolated

after three days of growth (175 rpm) in synthetic medium (1% KNO_3, 3% sucrose, 0.17% Yeast Nitrogen Base w/o amino acids and ammonium), mixed in a 1:1 ratio (10^5 conidia of each strain) and plated out on rich medium (potato dextrose agar-PDA). After six days of growth new microconidia had formed. Five milliliter sterile MQ water was pipetted on the co-incubated plates and microconidia were scraped off. For selection of double resistant colonies 20, 100 and 500 µl of the microconidia suspension was plated out on agar plates (PDA) containing both zeocin and hygromycin (buffered with 100 mM Tris pH 8.0). To determine the total number of plated conidia 20, 100 and 500 µl of a 1:1,000,000 dilution was plated on PDA plates without antibiotic. After two days, the number of double drug resistant colonies resistant were scored.

F.3. Bioassays

To test pathogenicity the root dip method was used [75]. Briefly, microconidia were collected from 5-day-old cultures in Czapek Dox broth (CDB; Difco) and used for root inoculation of 10-day-old tomato plants (Moneymaker C32) at a spore density of 10^7 spores per ml. The seedlings were then potted individually and grown at 25 °C in the greenhouse. Three weeks after inoculation, 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. Reisolation of isolates from diseased plants was done by placing slices of hypocotyls on agar plates (PDA) containing both zeocin and hygromycin. From mycelium growing out of the slices, single-spore colonies were derived and used for infection of tomato to satisfy Koch's postulates.

F.4. PCR amplifications

To test whether just the *BLE* resistance gene or the entire chromosome 14 was present in double drug resistant strains derived from the co-incubation experiments, three double drug resistant colonies of each successful combination (*Fol*007 with three independent *HYG*-transformants of *Fom* and Fo-47) were tested for the presence of genes located on different parts of chromosome 14, including *SIX1*, *SIX2, SIX3*, *SIX5, SIX6*, *SIX7* and *ORX1* (Figure S15). All seven genes were present in each of the nine double drug resistant strains (18 strains in total) indicating the presence of most or all of chromosome 14 in every case (Figure 4A – *SIX6* and *ORX1* are also present in the *Fom* strain used and could therefore not be used to distinguish between *Fol* and *Fom*-derived material). PCRs were performed on genomic DNA as described in [37]. For *SIX1-7* and *ORX1* the following primers were used:

SIX1-F1:	CCCCGAATTGAGGTGAAG
SIX1-R1:	AATAGAGCCTGCAAAGCATG
SIX2-F2:	TCTATCCGCTTTCTTCTCTC
SIX2-R2:	CAACGCCGTTTGAATAAGCA
SIX3-F1:	CCAGCCAGAAGGCCAGTTT
SIX3-R2:	GGCAATTAACCACTCTGCC
SIX5-F1:	ACACGCTCTACTACTCTTCA
SIX5-R1:	GAAAACCTCAACGCGGCAAA
SIX6-F1:	CTCTCCTGAACCATCAACTT
SIX6-R1:	CAAGACCAGGTGTAGGCATT
SIX7-F1:	CATCTTTTCGCCGACTTGGT
SIX7-R1:	CTTAGCACCCTTGAGTAACT
ORX1-F1:	CCAGGCCATCAAGTTACTC
ORX1-R1:	TCTCCAATATGGCAGATTGTG

To track individual chromosomes, a marker system based on differences in locations of the Foxy transposon was set up. The screen uses specific primers based on sequences ca. 500 bp upstream of 48 Foxy insertions based on the Fol genome sequence as released by the Broad Institute (isolate Fol4287). Each specific primer was used in combination with a reverse primer in Foxy and together the primer combinations covered all 15 chromosomes. (Table S25). Each was used with a reverse primer in Foxy (GAGAGAATTCTGGTCGGTG). First the primers were tested on Fol007 and Fo-47. From the primers that were positive on Fol007 and negative on Fo-47, twenty-nine primers were selected for testing the double drug-resistant strains for the presence of Fol007-derived chromosomes.

The additional Foxy insertion in strain 1C and 2A is located in scaffold 18 (1.3 Mb) on chromosome 3 and scaffold 21 (1.0 Mb) on chromosome 6. The presence of most or all of the sequence of scaffold 18/21 in strains 1C and 2A was confirmed with an additional 9 primers for Foxy insertions scattered over this region (Table S25). Given the differences in karyotype between *Fol*4287 and *Fol*007, it is highly likely that this region (scaffold 18/21 in *Fol*4287) is present in the smallest chromosome in *Fol*007 (Figure 4D).

F.5. CHEF & Southern blotting

Protoplasts (10^8/ml) from Fol4287, Fol007, three independent HYG transformants of Fo-47 and nine double drug-resistant strains were loaded on a CHEF gel as described in [44]. The gel (1% Seakem Gold agarose [FMC] in 0.5x TBE) ran for 255 hours, using switch times between 1200-4800 s at 1.5 V/cm. Chromosomes of *S. pombe* were used as a molecular size marker (BioRad). Chromosomes separated in the gel were blotted to Hybond-N+ (Amersham Pharmacia). PCR with primers SIX6-F1 and SIX6-R1 was used to generate a ~800 bp fragment containing the entire *SIX6* open reading frame. This fragment was radioactively labelled with γ^{32}P dATP using the DecaLabel™ DNA labeling kit from MBI Fermentas (Vilnius, Lithuania). Hybridization was done overnight at 65°C in 0.5M phosphate buffer pH 7.2 containing 7% SDS and 1 mM EDTA. Blots were washed at 65°C with 0.5 X SSC, 0.1% SDS. The position of chromosomes hybridizing to the *SIX6* sequence was visualized by phosphoimaging (Molecular Dynamics).

Reference cited

1 Jaffe, D. B. et al., Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Res* **13** (1), 91 (2003).

2 Xu, J. R. and Leslie, J. F., A genetic map of *Gibberella fujikuroi* mating population A (*Fusarium moniliforme*). *Genetics* **143** (1), 175 (1996).

3 Zhou, S. et al., Single-molecule approach to bacterial genomic comparisons via optical mapping. *J Bacteriol* **186** (22), 7773 (2004).

4 Park, J. et al., FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. *Bioinformatics* **24** (7), 1024 (2008).

5 Cantarel, B. L. et al., The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37** (Database issue), D233 (2009).

6 Miranda-Saavedra, D. and Barton, G. J., Classification and functional annotation of eukaryotic protein kinases. *Proteins* **68** (4), 893 (2007).

7 Lafon, A. et al., G-protein and cAMP-mediated signaling in aspergilli: a genomic perspective. *Fungal Genet Biol* **43** (7), 490 (2006).

8 Winnenburg, R. et al., PHI-base update: additions to the pathogen host interaction database. *Nucleic Acids Res* **36** (Database issue), D572 (2008).

9 Coleman, J. J. et al., The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet* **5** (8), e1000618 (2009).

10 Todd, R. B. and Andrianopoulos, A., Evolution of a fungal regulatory gene family: the Zn(II)2Cys6 binuclear cluster DNA binding motif. *Fungal Genet Biol* **21** (3), 388 (1997).

11 Guldener, U. et al., Development of a *Fusarium graminearum* Affymetrix GeneChip for profiling fungal gene expression *in vitro* and *in planta*. *Fungal Genet Biol* **43** (5), 316 (2006).

12 Haas, H., Eisendle, M., and Turgeon, B. G., Siderophores in fungal physiology and virulence. *Annual Review of Phytopathology* **46**, 149 (2008).

13 Stergiopoulous, I., Zwiers, L.-H. , and De Waard, M.A., Secretion of natural and synthetic toxic compounds from filamentous fungi by membrane transports of the ATP-binding cassette and major facilitator subfamily. *European Journal of Plant Pathology* **108**, 719 (2002).

14 Fleissner, A., Sopalla, C., and Weltring, K. M., An ATP-binding cassette multidrug-resistance transporter is necessary for tolerance of *Gibberella pulicaris* to phytoalexins and virulence on potato tubers. *Mol Plant Microbe Interact* **15** (2), 102 (2002).

15 Skov, J. , Lemmens, M. , and Giese, H. , Role of a *Fusarium culmorum* ABC transporter (FcABC1) during infection of wheat and barley. *Physiol Molecul Plant Pathol* **64**, 245 (2004).

16 Reuben, A. et al., Antifungal susceptibility of 44 clinical isolates of *Fusarium* species determined by using a broth microdilution method. *Antimicrob Agents Chemother* **33** (9), 1647 (1989).

17 Li, L. et al., Heterotrimeric G protein signaling in filamentous fungi. *Annual Review of Microbiology* **61**, 423 (2007).

18 Regenfelder, E. et al., G proteins in *Ustilago maydis*: transmission of multiple signals? *The EMBO Journal* **16** (8), 1934 (1997).

19 Kays, A. M. and Borkovich, K. A., Severe impairment of growth and differentiation in a *Neurospora crassa* mutant lacking all heterotrimeric G alpha proteins. *Genetics* **166** (3), 1229 (2004).

20 Thomas, P., Characteristics of membrane progestin receptor alpha (mPRalpha) and progesterone membrane receptor component 1 (PGMRC1) and their roles in mediating rapid progestin actions. *Frontiers in Neuroendocrinology* **29** (2), 292 (2008).

21 Baasiri, R. A. et al., Overlapping functions for two G protein alpha subunits in *Neurospora crassa*. *Genetics* **147** (1), 137 (1997).

22 Murphy, M. and Armstrong, D., Fusariosis in patients with neoplastic disease. . *Infect.Med* **12**, 66 (1995).

23 Seefelder, W. et al., Induction of apoptosis in cultured human proximal tubule cells by fumonisins and fumonisin metabolites. *Toxicol Appl Pharmacol* **192** (2), 146 (2003).

24 Hendricks, K., Fumonisins and neural tube defects in South Texas. *Epidemiology* **10** (2), 198 (1999).

25 Marasas, W. F. et al., Fumonisins disrupt sphingolipid metabolism, folate transport, and neural tube development in embryo culture and in vivo: a potential risk factor for human neural tube defects among populations consuming fumonisin-contaminated maize. *J Nutr* **134** (4), 711 (2004).

26 Swofford, D. L., *PAUP. Phylogenetic Analysis Using Parsimony*. (Sinauer Associates, Sunderland, Massachusetts., 2003).

27 Gaffoor, I. and Trail, F., Characterization of two polyketide synthase genes involved in zearalenone biosynthesis in *Gibberella zeae*. *Applied and environmental microbiology* **72** (3), 1793 (2006).

28 Gaffoor, I. et al., Functional analysis of the polyketide synthase genes in the filamentous fungus *Gibberella zeae* (anamorph *Fusarium graminearum*). *Eukaryot Cell* **4** (11), 1926 (2005).

29 Proctor, R. H., Butchko, R. A., Brown, D. W., and Moretti, A., Functional characterization, sequence comparisons and distribution of a polyketide synthase gene required for perithecial pigmentation in some Fusarium species. *Food Addit Contam* **24** (10), 1076 (2007).

30 Frandsen, R. J. et al., The biosynthetic pathway for aurofusarin in *Fusarium graminearum* reveals a close link between the naphthoquinones and naphthopyrones. *Molecular Microbiology* **61** (4), 1069 (2006).

31 Hallen, H. E. and Trail, F., The L-type calcium ion channel cch1 affects ascospore discharge and mycelial growth in the filamentous fungus *Gibberella zeae* (anamorph *Fusarium graminearum*). *Eukaryot Cell* **7** (2), 415 (2008).

32 Li, M., Ma, B., Kisman, D., and Tromp, J., Patternhunter II: highly sensitive and fast homology search. *J Bioinform Comput Biol* **2** (3), 417 (2004).

33 Cuomo, C. A. et al., The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* **317** (5843), 1400 (2007).

34 Thatcher, L. F., Manners, J. M., and Kazan, K., *Fusarium oxysporum* hijacks COI1-mediated jasmonate signaling to promote disease development in *Arabidopsis*. *Plant J* **58** (6), 927 (2009).

35 Dowd, C., Wilson, I. W., and McFadden, H., Gene expression profile changes in cotton root and hypocotyl tissues in response to infection with *Fusarium oxysporum* f. sp. *vasinfectum*. *Mol Plant Microbe Interact* **17** (6), 654 (2004).

36 Li, H., Ruan, J., and Durbin, R., Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* **18** (11), 1851 (2008).

37 van der Does, H. C. et al., The presence of a virulence locus discriminates *Fusarium oxysporum* isolates causing tomato wilt from other isolates. *Environmental Microbiology* (2008).

38 Daboussi, M. J. and Capy, P., Transposable elements in filamentous fungi. *Annual Review of Microbiology* **57**, 275 (2003).

39 Daboussi, M. J., Daviere, J. M., Graziani, S., and Langin, T., Evolution of the *Fot1* transposons in the genus *Fusarium*: discontinuous distribution and epigenetic inactivation. *Mol Biol Evol* **19** (4), 510 (2002).

40 Daboussi, M. J., Langin, T., and Brygoo, Y., *Fot1*, a new family of fungal transposable elements. *Mol Gen Genet* **232** (1), 12 (1992).

41 Bergemann, M. et al., Genome-wide analysis of the *Fusarium oxysporum* mimp family of MITEs and mobilization of both native and *de novo* created mimps. *J Mol Evol* (2008).

42 Gomez-Gomez, E., Anaya, N., Roncero, M. I., and Hera, C., *Folyt1*, a new member of the hAT family, is active in the genome of the plant pathogen *Fusarium oxysporum*. *Fungal Genet Biol* **27** (1), 67 (1999).

43 Chalvet, F. et al., *Hop*, an active Mutator-like element in the genome of the fungus *Fusarium oxysporum*. *Mol Biol Evol* **20** (8), 1362 (2003).

44 Teunissen, H. A. et al., Construction of a mitotic linkage map of *Fusarium oxysporum* based on Foxy-AFLPs. *Mol Genet Genomics* **269** (2), 215 (2003).

45 Rasko, D. A., Myers, G. S., and Ravel, J., Visualization of comparative genomic analyses by BLAST score ratio. *BMC Bioinformatics* **6**, 2 (2005).

46 Ashburner, M. et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* **25** (1), 25 (2000).

47 Conesa, A. et al., Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21** (18), 3674 (2005).

48 Catanzariti, A. M., Dodds, P. N., and Ellis, J. G., Avirulence proteins from haustoria-forming pathogens. *FEMS Microbiology Letters* **269** (2), 181 (2007).

49 Fudal, I. et al., Heterochromatin-like regions as ecological niches for avirulence genes in the *Leptosphaeria maculans* genome: map-based cloning of *AvrLm6*. *Mol Plant Microbe Interact* **20** (4), 459 (2007).

50 Houterman, P. M. et al., The effector protein Avr2 of the xylem colonizing fungus *Fusarium oxysporum* activates the tomato resistance protein I-2 intracellularly. *Plant J* (2009).

51 Qutob, D. et al., Phytotoxicity and innate immune responses induced by *Nep1*-like proteins. *The Plant Cell* **18** (12), 3721 (2006).

52 Di Pietro, A. and Roncero, M. I., Cloning, expression, and role in pathogenicity of pg1 encoding the major extracellular endopolygalacturonase of the vascular wilt pathogen *Fusarium oxysporum*. *Mol Plant Microbe Interact* **11** (2), 91 (1998).

53 Welinder, K. G., Bacterial catalase-peroxidases are gene duplicated members of the plant peroxidase superfamily. *Biochimica et Biophysica Acta* **1080** (3), 215 (1991).

54 Wang, X., Lipid signaling. *Current Opinion in Plant Biology* **7** (3), 329 (2004).

55 Shea, J. M. and Del Poeta, M., Lipid signaling in pathogenic fungi. *Current Opinion in Microbiology* **9** (4), 352 (2006).

56 Beller M, Sztalryd C, Southall N, Bell M, Jäckle H, et al., COPI complex is a regulator of lipid homeostasis. *PLoS Biology* **6** (11), e292 doi:10.1371/journal.pbio.0060292 (2008).

57 Brasaemle, D. L., Thematic review series: adipocyte biology. The perilipin family of structural lipid droplet proteins: stabilization of lipid droplets and control of lipolysis. *Journal of Lipid Research* **48** (12), 2547 (2007).

58 Patel, R. T., Soulages, J. L., Hariharasundaram, B., and Arrese, E. L., Activation of the lipid droplet controls the rate of lipolysis of triglycerides in the insect fat body. *The Journal of Biological Chemistry* **280** (24), 22624 (2005).

59 Schweiger, M. et al., Adipose triglyceride lipase and hormone-sensitive lipase are the major enzymes in adipose tissue triacylglycerol catabolism. *The Journal of Biological Chemistry* **281** (52), 40236 (2006).

60 Hwang, C. S., Flaishman, M. A., and Kolattukudy, P. E., Cloning of a gene expressed during appressorium formation by *Colletotrichum gloeosporioides* and a marked decrease in virulence by disruption of this gene. *The Plant Cell* **7** (2), 183 (1995); Wang, C. and St Leger, R. J., The *Metarhizium anisopliae* perilipin homolog MPL1 regulates lipid

metabolism, appressorial turgor pressure, and virulence. *The Journal of Biological Chemistry* **282** (29), 21110 (2007).

61 Yu, J. H., Wieser, J., and Adams, T. H., The *Aspergillus FlbA* RGS domain protein antagonizes G protein signaling to block proliferation and allow development. *The EMBO Journal* **15** (19), 5184 (1996).

62 Cambareri, E. B., Jensen, B. C., Schabtach, E., and Selker, E. U., Repeat-induced G-C to A-T mutations in *Neurospora*. *Science* **244** (4912), 1571 (1989).

63 Thines, E., Weber, R. W., and Talbot, N. J., Signal transduction leading to appressorium formation in germinating conidia of *Magnaporthe grisea*: Effects of second messengers discylglycerols, ceramindes and sphingomycelin. *FEMS Microbiol. Lett.* **156**, 91 (1997).

64 Nolan, T. et al., The post-transcriptional gene silencing machinery functions independently of DNA methylation to repress a LINE1-like retrotransposon in *Neurospora crassa*. *Nucleic Acids Res* **33** (5), 1564 (2005).

65 Farman, M. L., Tosa, Y., Nitta, N., and Leong, S. A., MAGGY, a retrotransposon in the genome of the rice blast fungus *Magnaporthe grisea*. *Mol Gen Genet* **251** (6), 665 (1996).

66 Selker, E. U., Repeat-induced gene silencing in fungi. *Adv Genet* **46**, 439 (2002).

67 Tamaru, H. and Selker, E. U., A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* **414** (6861), 277 (2001).

68 Freitag, M. et al., HP1 is essential for DNA methylation in *Neurospora*. *Mol Cell* **13** (3), 427 (2004).

69 Lee, J., Jurgenson, J. E., Leslie, J. F., and Blowden, R. L., Alignment of genetic and physical maps of *Gibberella zeae*. *Applied and Environmental Microbiology* **74** (8), 2349 (2008).

70 Chicas, A. et al., Small interfering RNAs that trigger posttranscriptional gene silencing are not required for the histone H3 Lys9 methylation necessary for transgenic tandem repeat stabilization in *Neurospora crassa*. *Mol Cell Biol* **25** (9), 3793 (2005).

71 Borkovich, K. A. et al., Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism. *Microbiol Mol Biol Rev* **68** (1), 1 (2004).

72 Michielse, C. B. et al., Insight into the molecular requirements for pathogenicity of *Fusarium oxysporum* f. sp. *lycopersici* through large-scale insertional mutagenesis. *Genome Biol* **10** (1), R4 (2009).

73 Takken, F. L. et al., A one-step method to convert vectors into binary vectors suited for Agrobacterium-mediated transformation. *Current genetics* **45** (4), 242 (2004).

74 van der Does, H. C. et al., Expression of effector gene *SIX1* of *Fusarium oxysporum* requires living plant cells. *Fungal Genet Biol* **45** (9), 1257 (2008).

75 Wellman, F.L., A technique for studying host resistance and pathogenicity in tomato *Fusarium* wilt. *Phytopathology* **29**, 945 (1939).

76 Houterman, P. M. et al., The mixed xylem sap proteome of *Fusarium oxysporum* infected tomato plants. *Molecular Plant Pathology* **8** (2), 215 (2007).

77 Rep, M. et al., A small, cysteine-rich protein secreted by Fusarium oxysporum during colonization of xylem vessels is required for I-3-mediated resistance in tomato. *Molecular Microbiology* **53** (5), 1373 (2004).

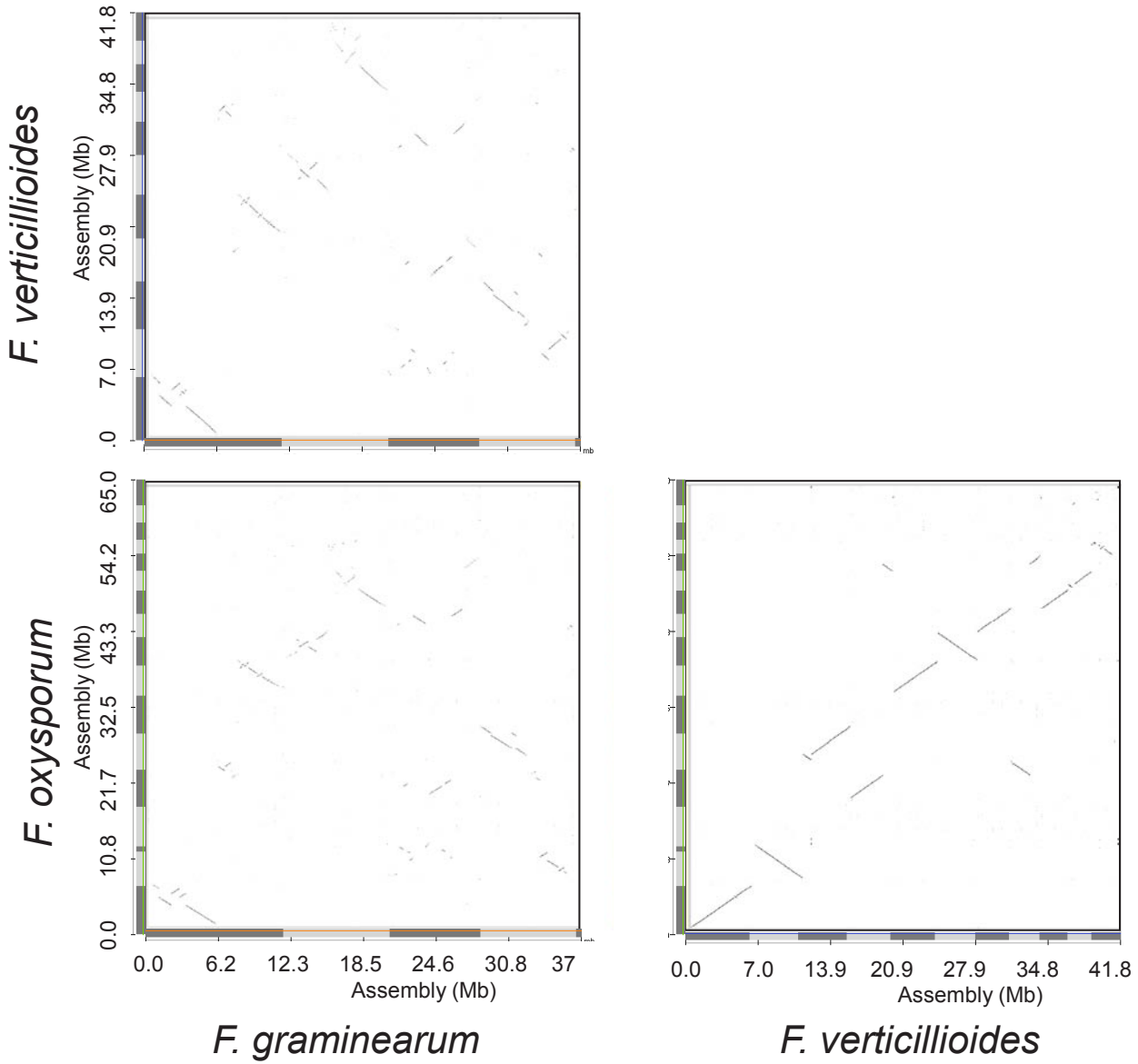

Figure S1. Pair-wise dotplots among three *Fusarium* genomes. The x and y axes represent assembled genomes. The altered gray and light grey bars are defined chromosomes in each assembly (See detailed mapping information at: http://www.broadinstitute.org/annotation/genome/fusarium_group/maps/Index.html). Comparisons are based on blastn alignments of genomic sequence (cutoff 1e-10) using *Fg* (against *Fv* and *Fo*) and *Fv* (against *Fo*) genomic DNA as query sequences.

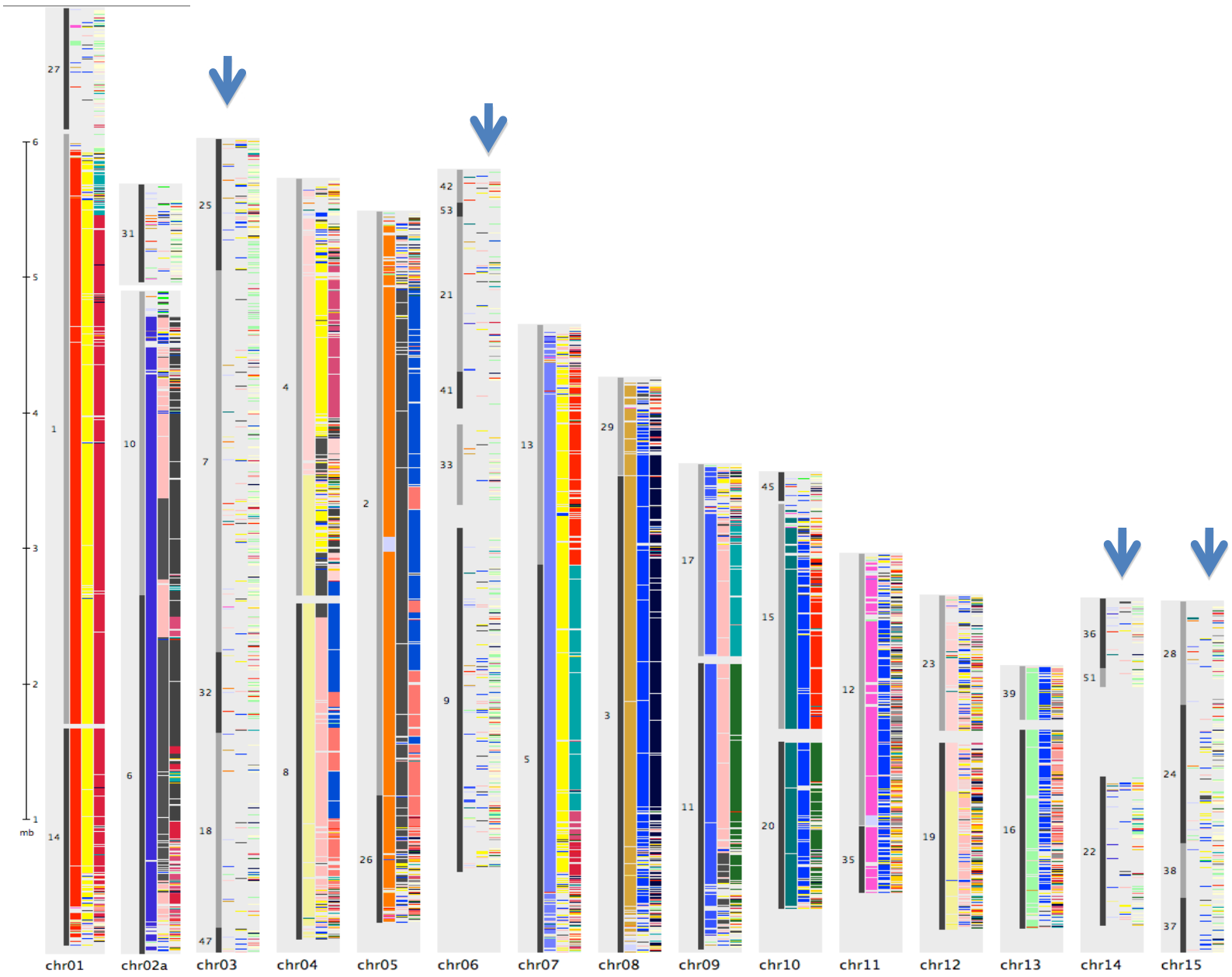

Figure S2. Syntnic alignments among four *Fusarium* genomes with Fol as the reference genome. The columns in each panel are *Fol* scaffolds mapped to the optical maps, the *Fv, Fg* and *Fs* genomes mapped to the *Fol* genomic sequences based on blastn alignments (cutoff 1e-10). The arrows point to the LS chromosomes in *Fol* that lack significant homologous sequences in all other three genomes.

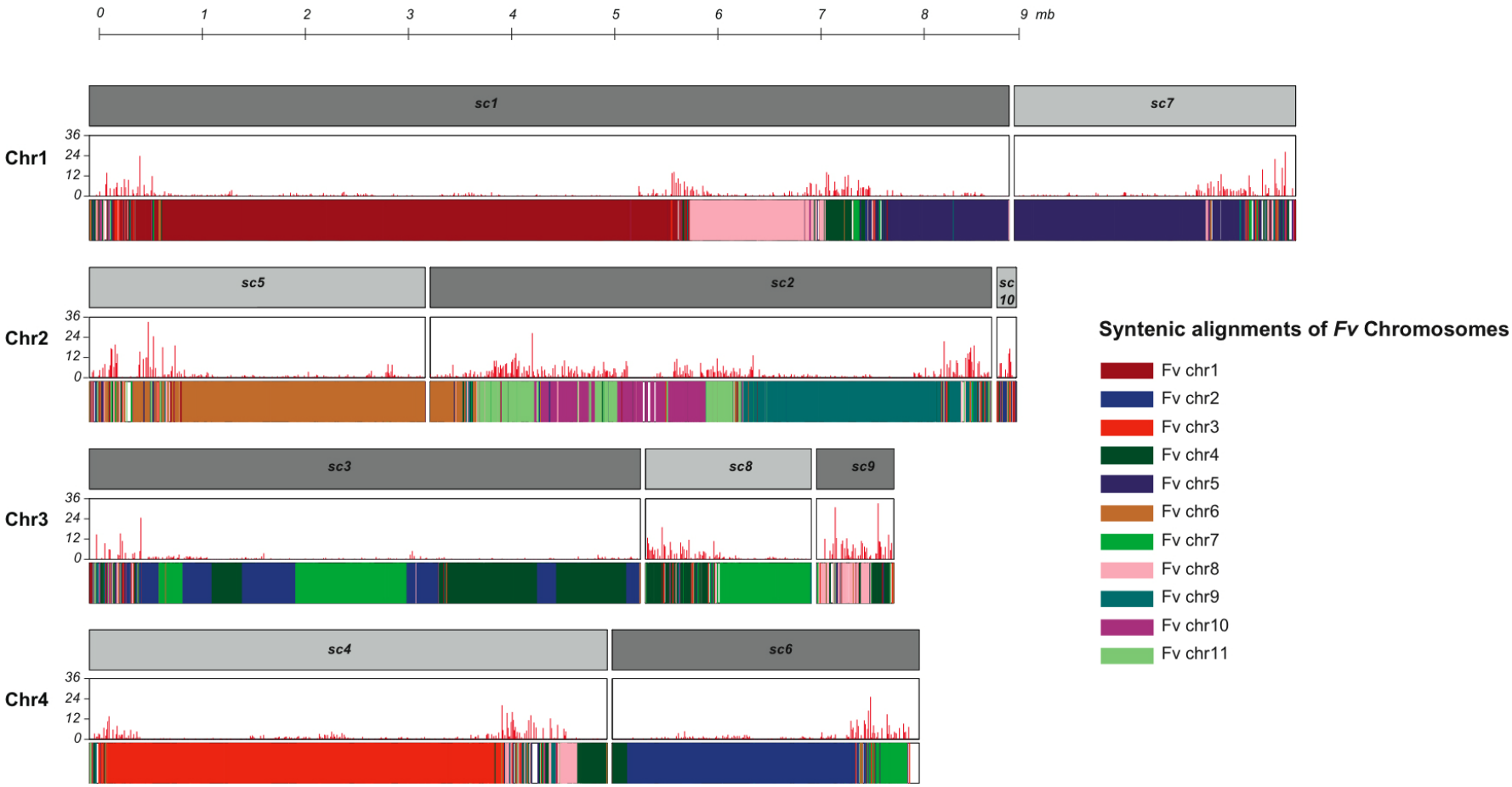

Figure S3. Whole genome alignments of *Fg* to *Fv.* The alignments display end to end synteny in large blocks with the exception of *Fv* chromosome ends and reveal multiple chromosome fusions in *Fg*. The previously described highly polymorphic and recombinogenic regions of *Fg* (*5*) correspond to *Fv* chromosome ends, including the implied interstitial fusion sites.

.

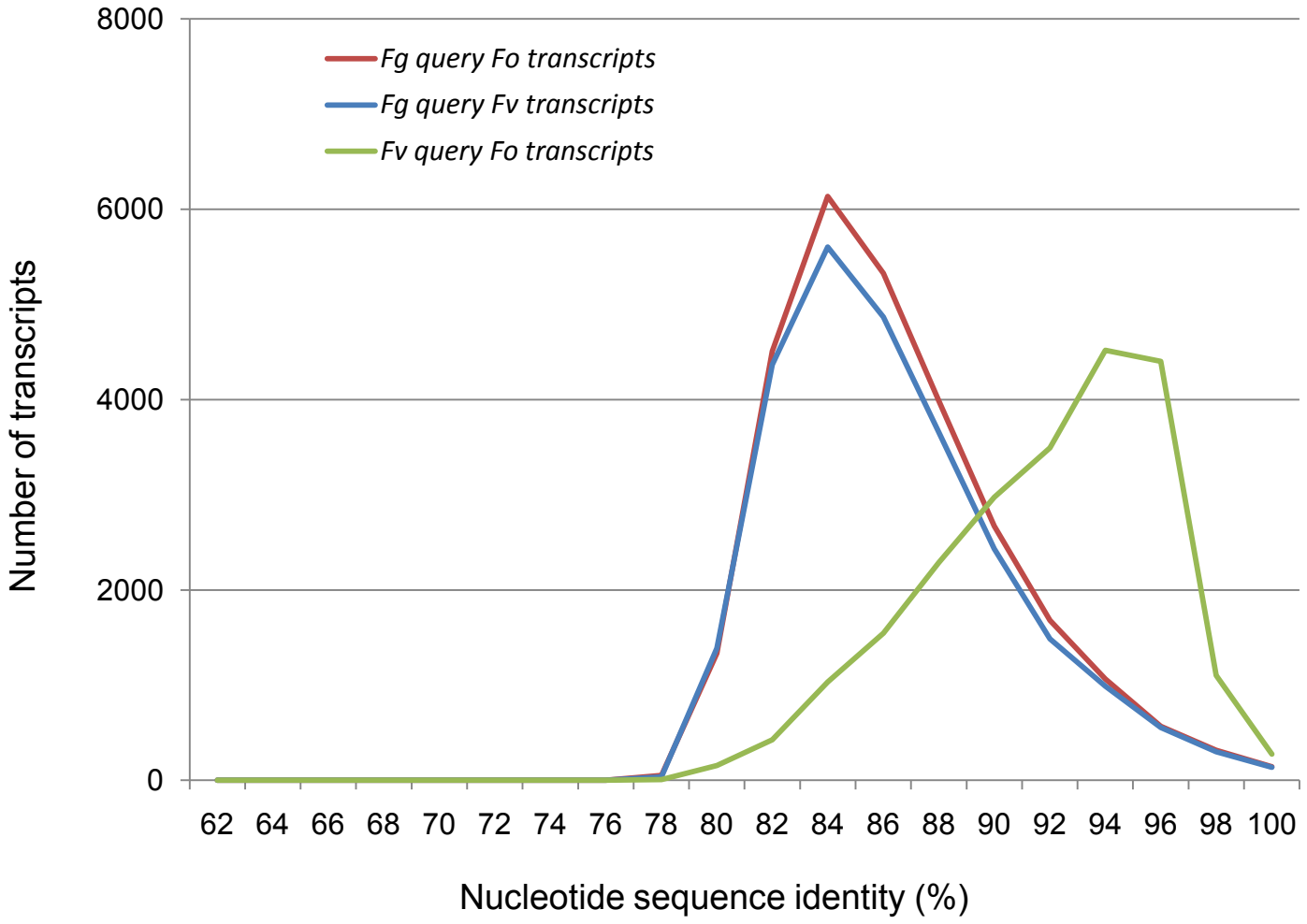

Figure S4. Pair-wise nucleotide sequence identity of coding sequences among three *Fusarium* transcripts. Based on blastn alignments of orthologous genes (cutoff 1e-10) using *Fg* (against *Fv* and *Fo*) and *Fv* (against *Fo*) transcripts as query sequences.

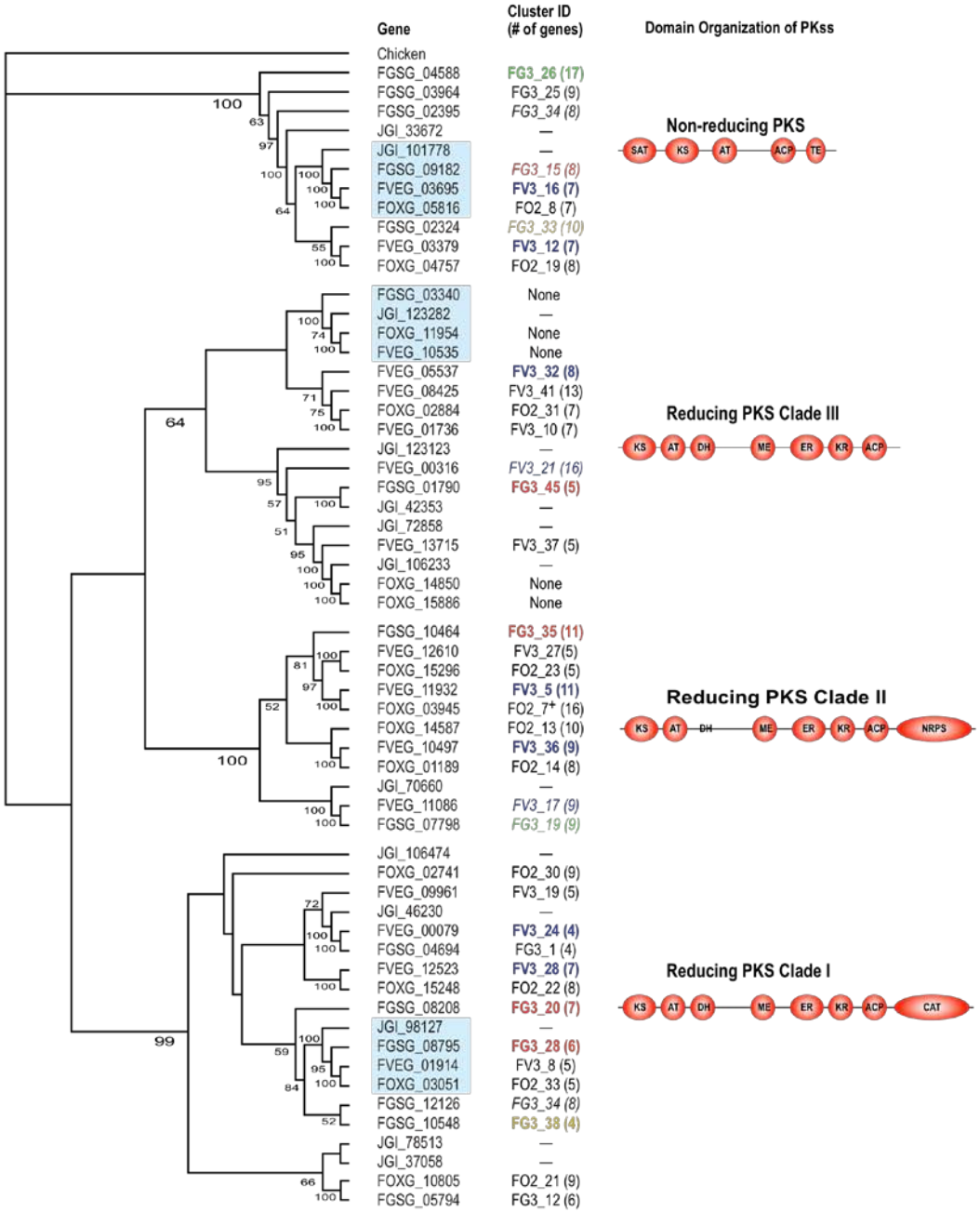

Figure S5. Cladogram representing a phylogenetic analysis of 58 *Fusarium* PKS KS (ketosynthase) and AT (acyl transferase) domains. ClustalW-aligned amino acid sequences were used to construct a gene genealogy, using parsimony in PAUP* 4.0b10. Bootstrap values were generated with 1000 replications to test for significance. The gene cluster identifier is listed next to the PKS gene that defines the cluster. The number of genes included in each cluster is listed in parentheses. The known SMB gene clusters are in italics. The clusters in bold letters are co-expressed as determined by microarray expression data [43]. The clusters in green are specifically expressed *in planta*. The clusters in red are specifically expressed during sexual development; and the clusters in yellow are expressed in culture. -- = cluster not tested; None = no cluster identified with the required four SMB related genes within a 20 kb window. ACP = acyl carrier protein domains; AT = acyl transferase; CAT = carnitine transferase; DH = dehydrogenase, ER = enoylreductase; KR = keto reductase; KS = ketosynthase; ME = methyltransferase; SAT = starter unit acetyltransferase, and TE = thioesterase.

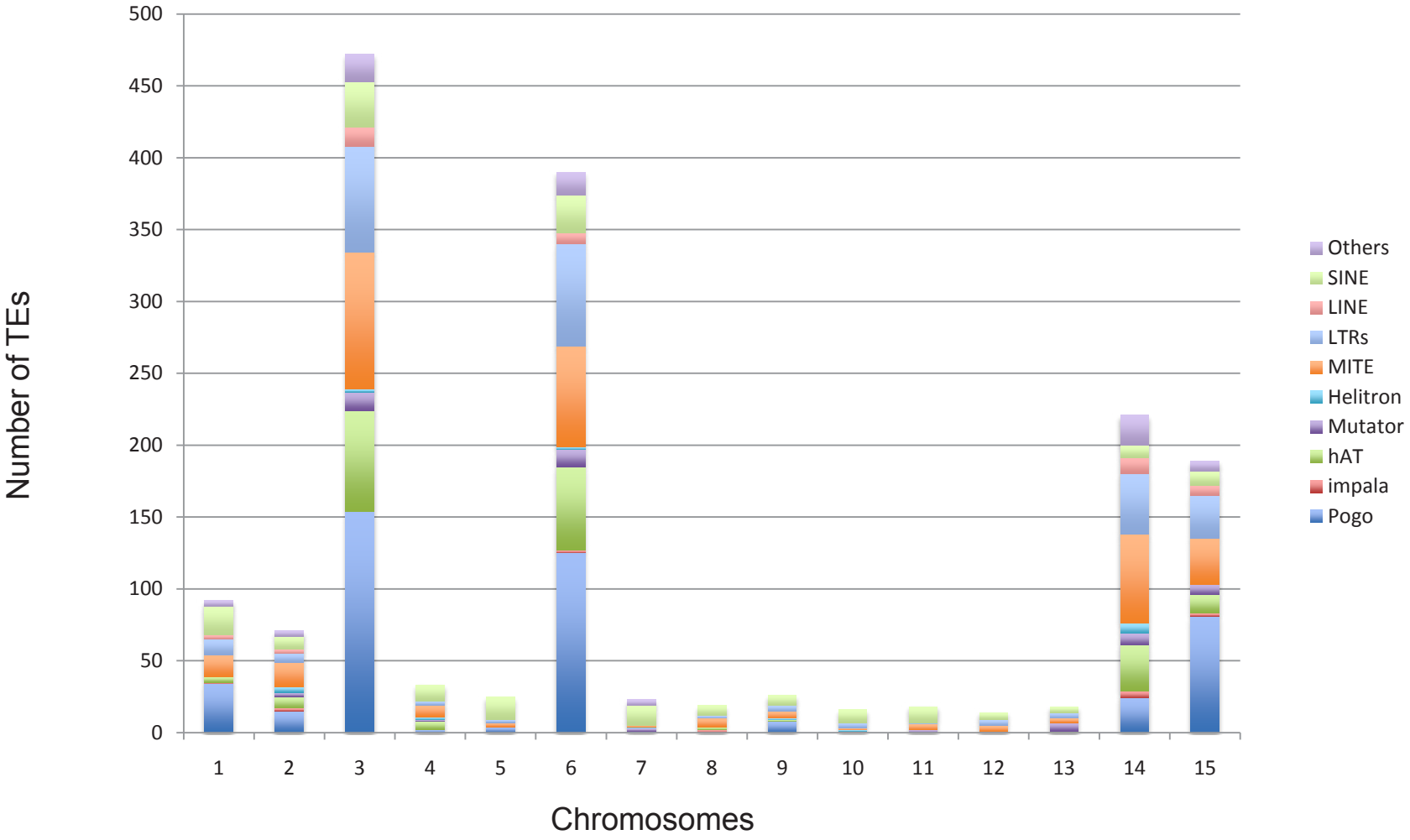

Figure S6. Distribution of TEs in *Fol* chromosomes. Full-length transposable elements were annotated using a combination of computational predictions based on BLAST analysis for transposase genes and manual inspection.

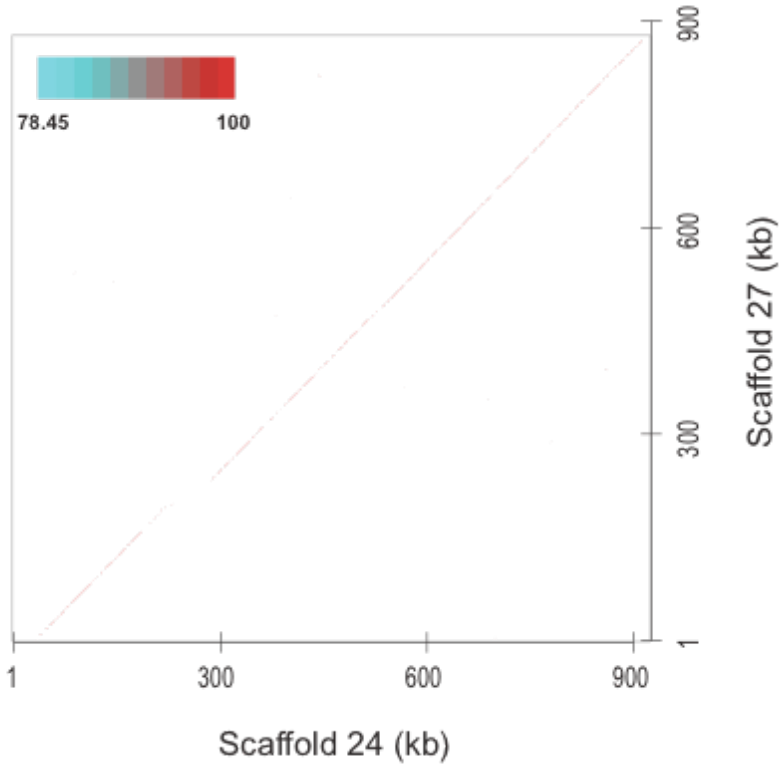

Figure S7. Dot-plot of inter-chromosomal segmental duplication as an example showing high sequence identity reflecting recent duplication, based on blastn (1e-20). The color legend indicates the level of sequence identity.

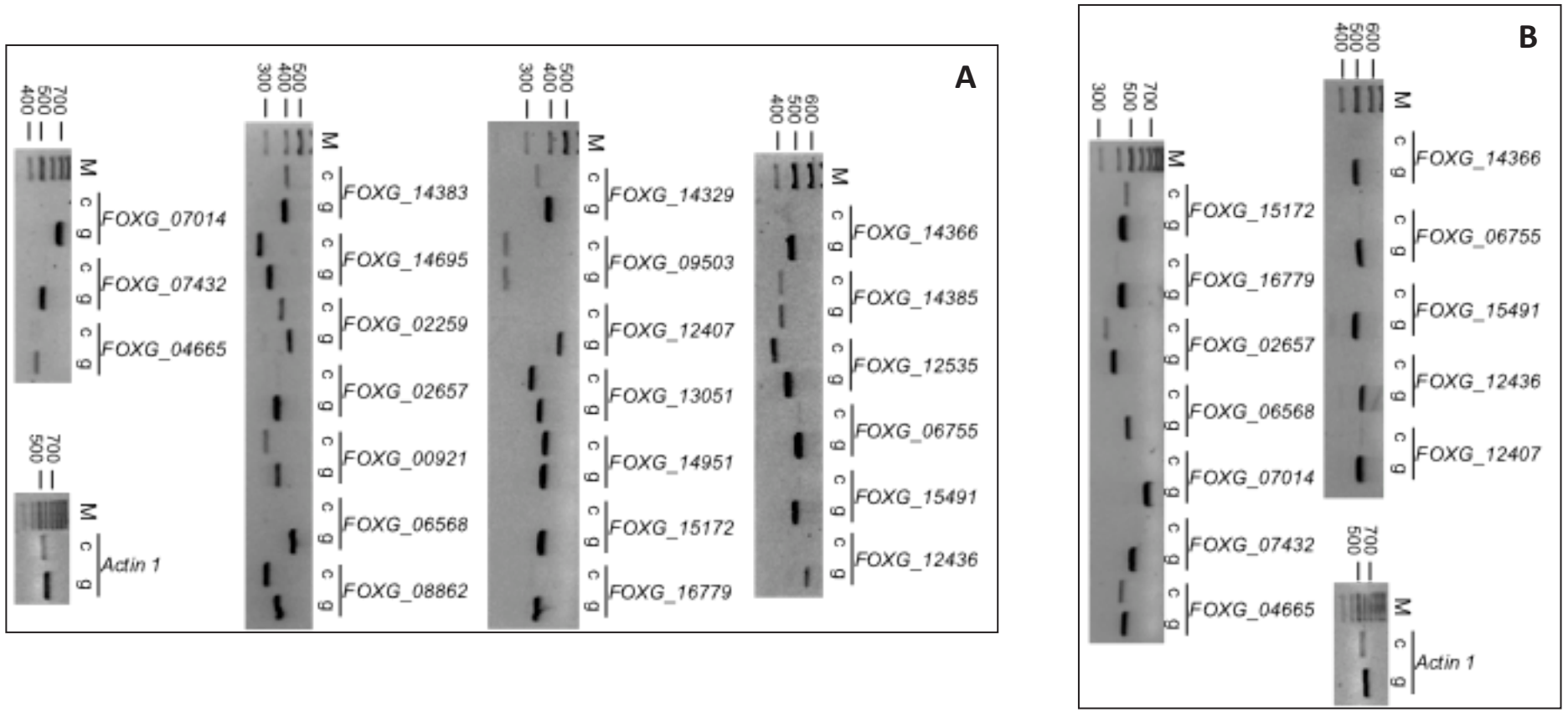

Figure S8 RT-PCR products showing the expression pattern of the indicated *F. oxysporum* CWDE genes (see Supplementary Information for gene annotations) during infection of tomato plants. cDNAs (c) were generated from total RNA isolated from roots of infected tomato plants three (A) or seven (B) days after inoculation, and used as a template for PCR amplification with gene specific primers (see Supplementary Table S13). Genomic DNA (g) was used as control for transcript size. Sizes of marker bands (M) are in bp. Predicted genes *FOXG_09503, FOXG_14383, FOXG_14385, FOXG_14951* and *FOXG_15172* lack introns (see Supplementary Table S16 for transcript sizes). Control cDNA obtained from uninoculated plants failed to produce any amplification bands (results not shown).

FOXG_04665 and FOXG_15172 are only expressed on day 7 (Part B). The very faint bands in FOXG_06755 and FOXG_15491 migrate at the same position as the gDNA, although the primers for these gene were designed to flank an intron. Therefore, the band corresponds to genomic rather than to cDNA.

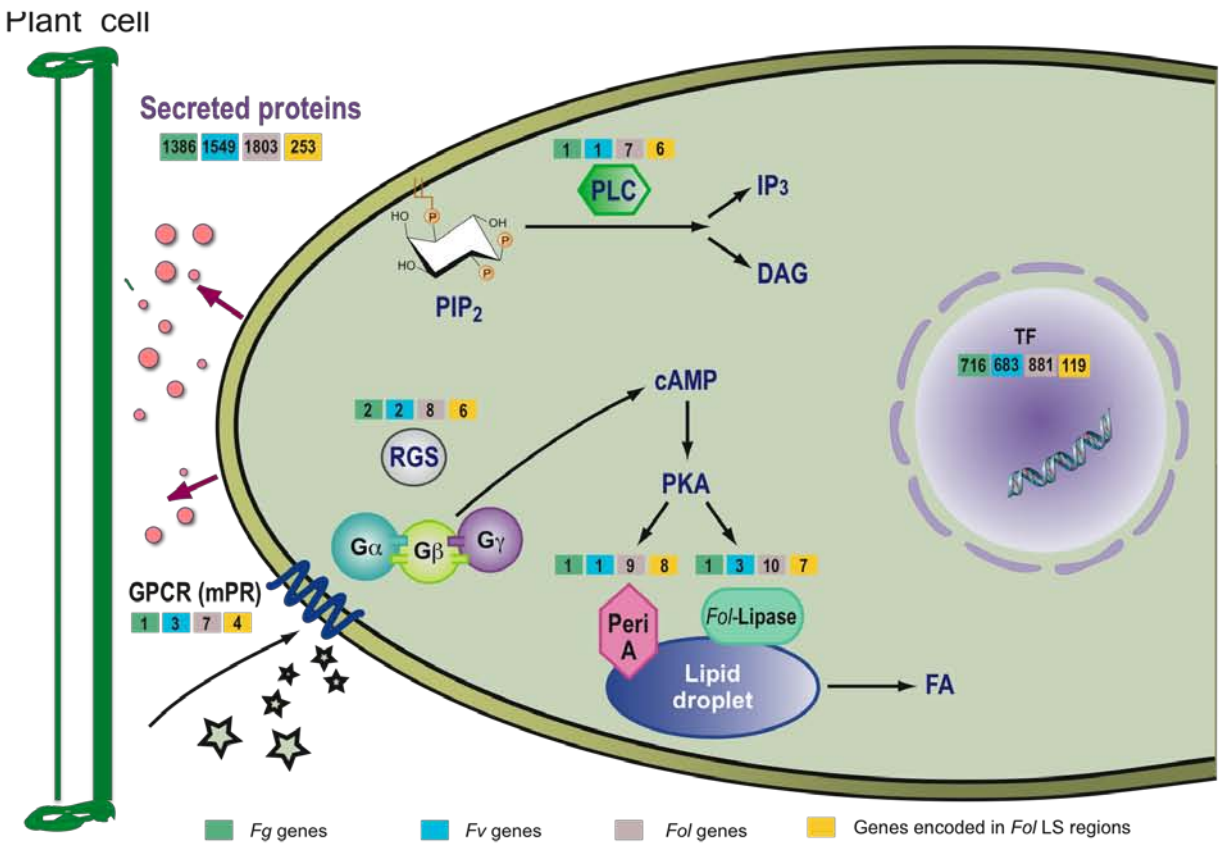

Figure S9. Proteins predicted from genes in the *Fol* LS regions are enriched for proteins predicted to be secreted (p<7e-5) and proteins of the lipid metabolism signalling pathways (p<7.6e-5).

The number of genes encoded on the *Fol* LS regions, shown in yellow, out of the total of 2448 genes encoded on LS chromosomes. In addition, genes identified from *Fg* [45], *Fv* (blue), and *Fol* (purple) in the same categories are shown. See Table S18 for details.

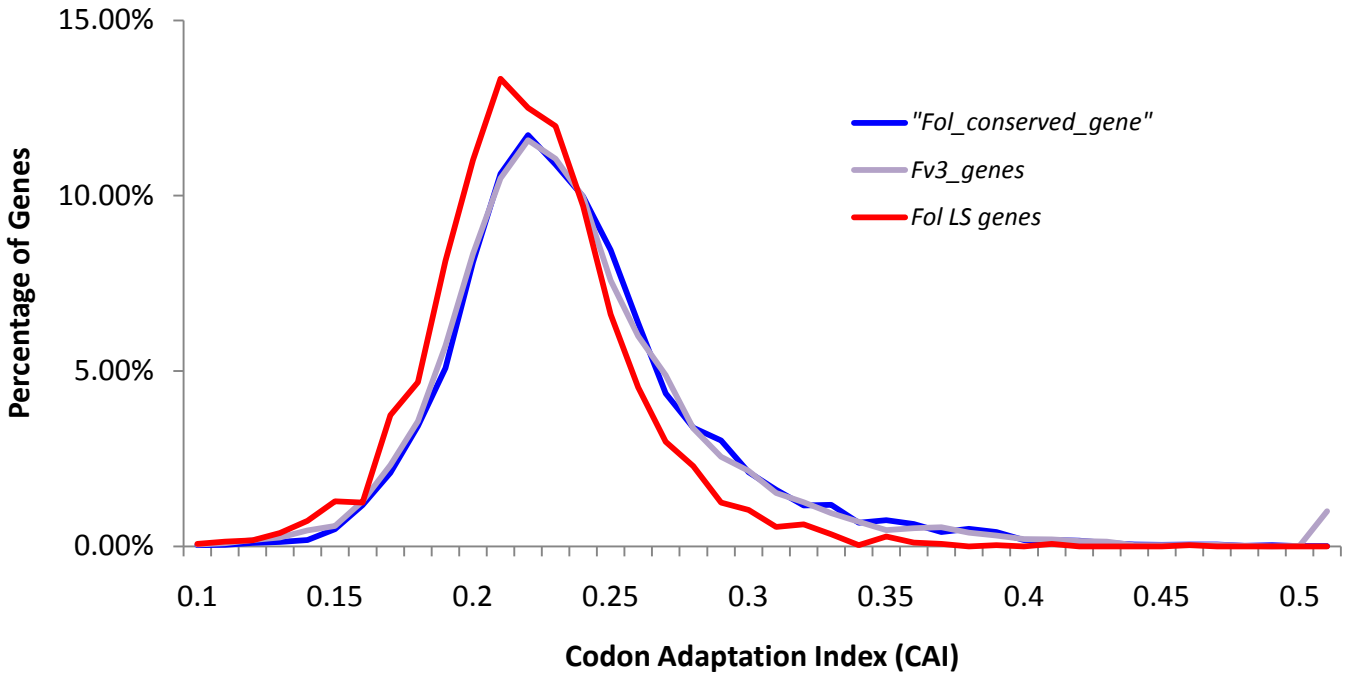

Figure S10. The Codon Adaptation Index (CAI) distribution of genes encoded *Fol* LS regions compared to genes encoded in *Fol* conserved region and the *Fv* genes. The CAI derived from the RSCU estimations is computed using the EMBOSS tool 'cai' (http://oryx.ulb.ac.be/embosshelp/cai.html).

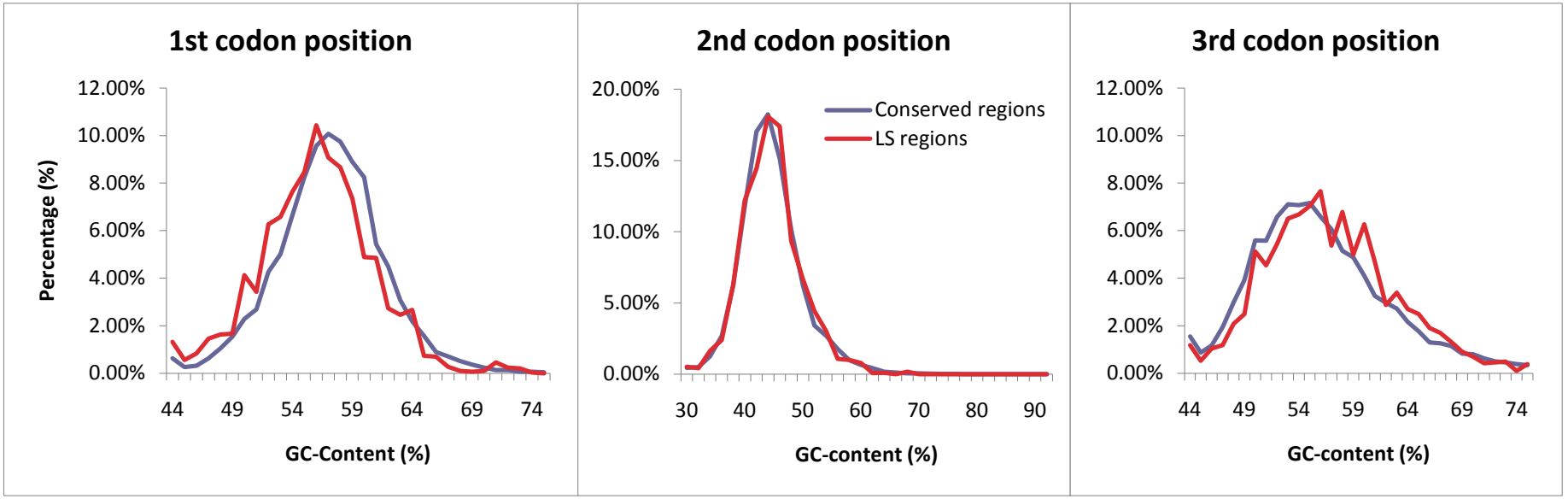

Figure S11. GC-content distribution of *Fol* genes encoded in the conserved versus LS regions.

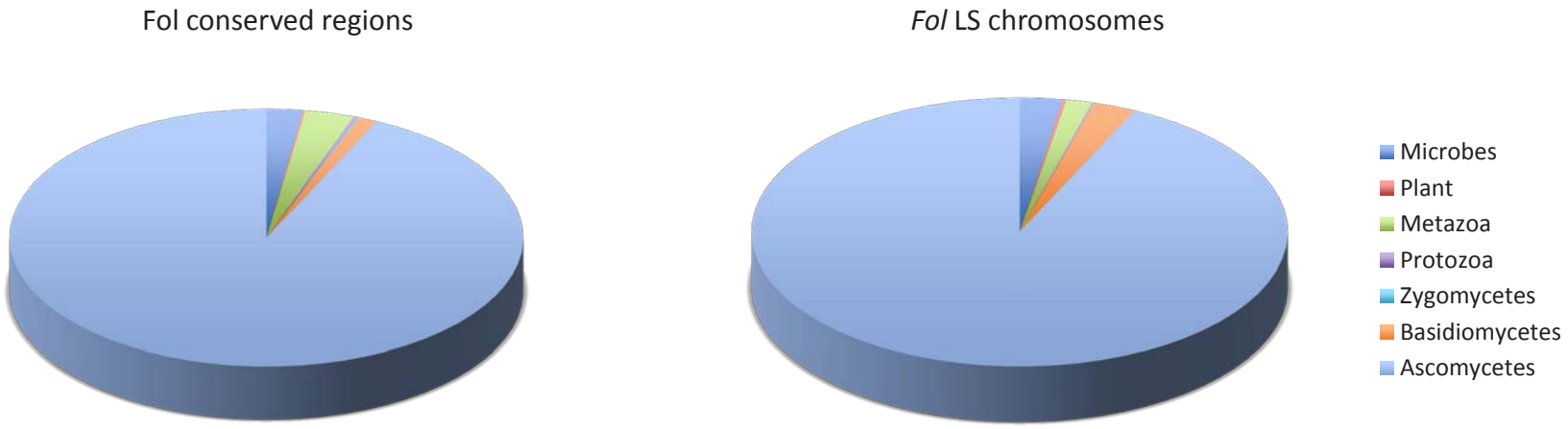

Figure S12. Homologous profile of *Fol* genes encoded in the conserved versus LS regions comparing to proteomes across different kingdoms.
The *Fol* proteins were searched using BLASTP (1e-20) against the NCBI metazoan, plant, microbial gene sets available at ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/GENE_INFO (February 21, 2008 version) and the non-*Fusarium* fungal database including the protein sets from Ascomycete: two fungal genomes from each subphylum Sordariomycetes *(Magnaporthe grisea, Neurospora crassa),* subphylum Leotiomycetes*, (Botrytis cinerea, Sclerotinia sclerotiorum),* and Eurotiomycetes (*Aspergillus fumigatus, A. oryzae)*; Basidiomycete: *Ustilago maydis, Coprinus cinereus, Cryptococcus neoformans* serotype A; and a zygomycete fungal *Rhizopus oryzae* protein set.

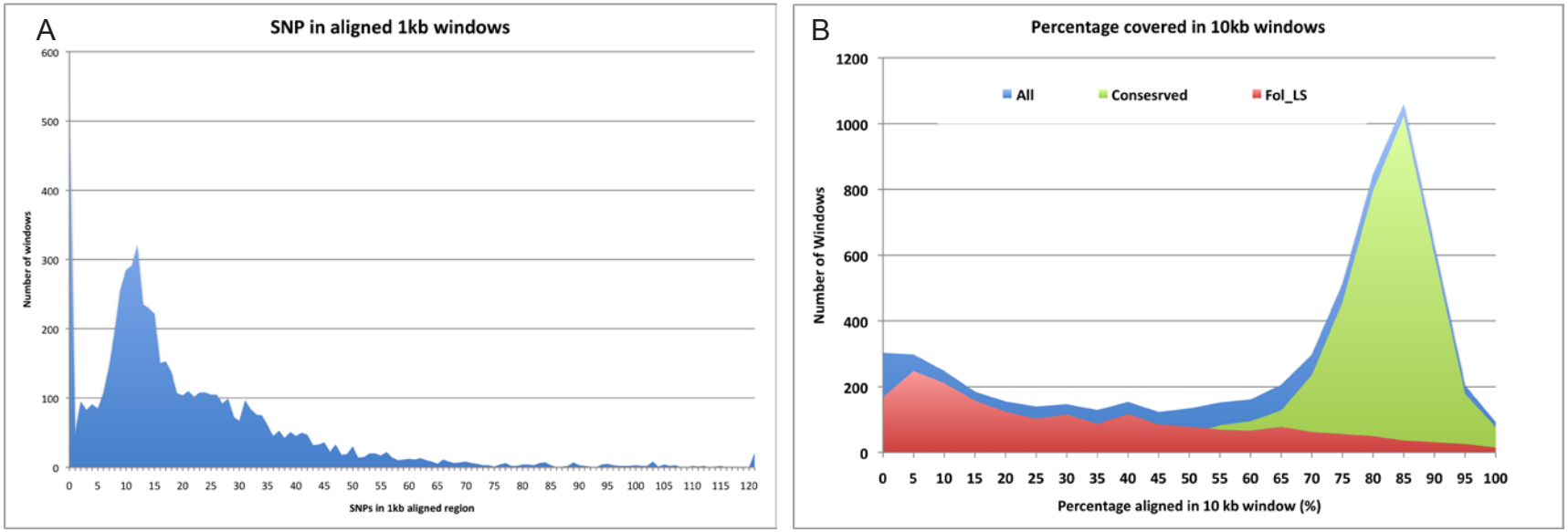

Figure S13. The SNP density (A) and the sequence coverage (B) of the reference assembly *Fol* aligned by Illumina reads from the strain FO5176.

A total of 26.7 million Illumina reads of 51 bases were aligned to the *Fol* assembly using MAQ (43). About 40% of the reference assembly is not present in the strain. A) . The overall SNP rate is less than 20 SNPs per 1 kb window. B). There are two peaks for the fraction of the reference genome that can be aligned by the Illumina reads. The major peak is centered on 85% (range from 60%-100%), and a smaller peak at the low end from 0-10%. These two peaks are clearly separated between the conserved and the *Fol*-LS regions.

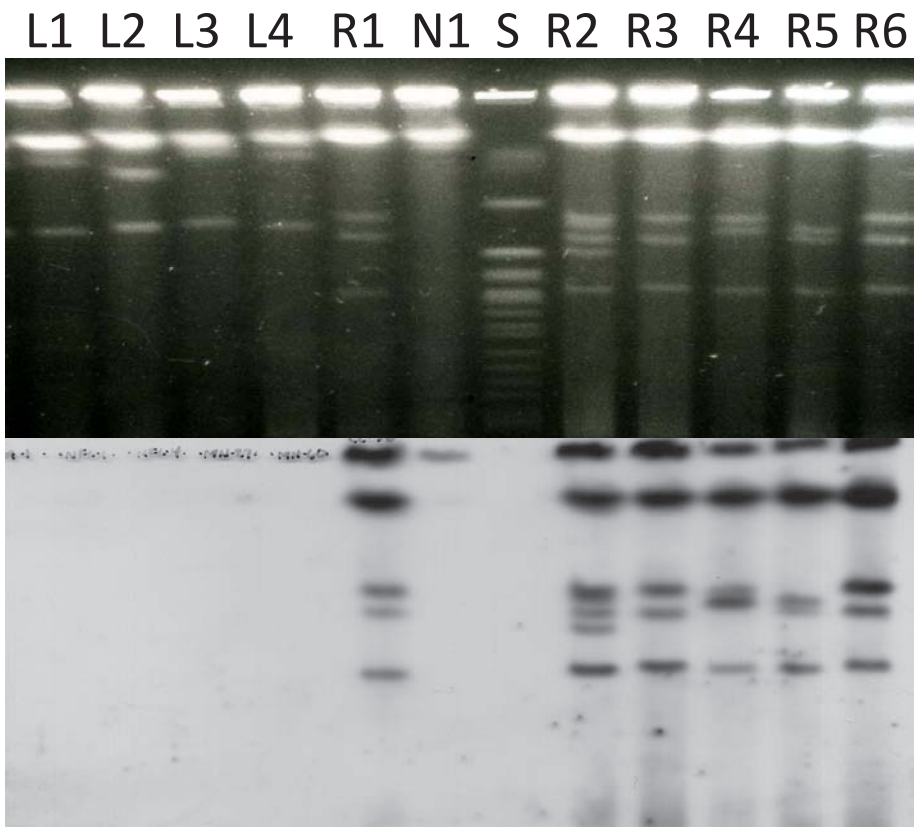

Figure S14. Karyotype variation and lineage- specific chromosomes among tomato-infecting strains of *F. oxysporum*.

(Above) Small chromosomes (< 2.3 Mb) separated from four strains of *F. oxysporum* f. sp. *lycopersici* (L1-L4), six strains of *F. oxysporum* f. sp. *radicis-lycopersici* (R1-R6), and one non-pathogenic strain of *F. oxysporum* from tomato (N1). Size standards are chromosomes from *Saccharomyces cereviseae* (S). (Below) Southern blot hybridized with a clone containing a lineage-specific repetitive sequence.

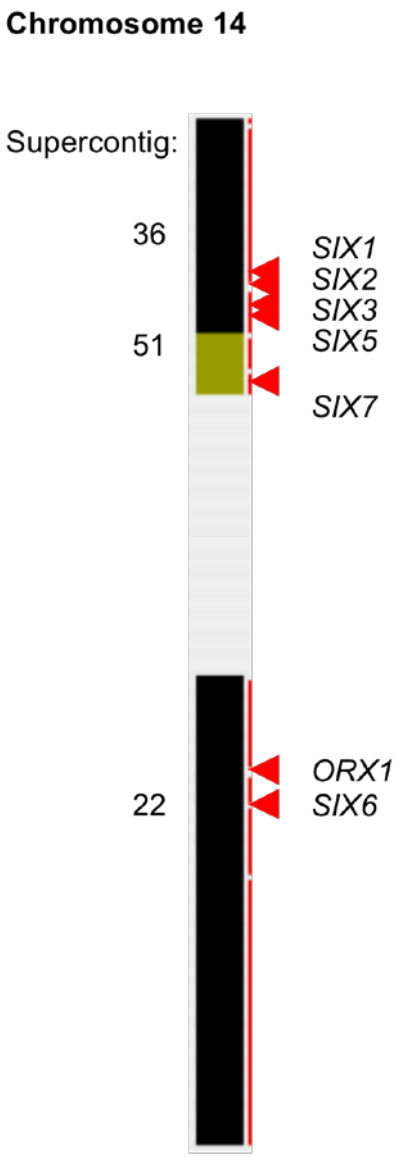

Figure S15. Schematic representation of the scaffolds attributed to chromosome 14 of Fol4287, showing the location of the *SIX (Secreted In Xylem)* genes discussed in this work and *ORX1 (OxidoReductase secreted in Xylem 1)*.

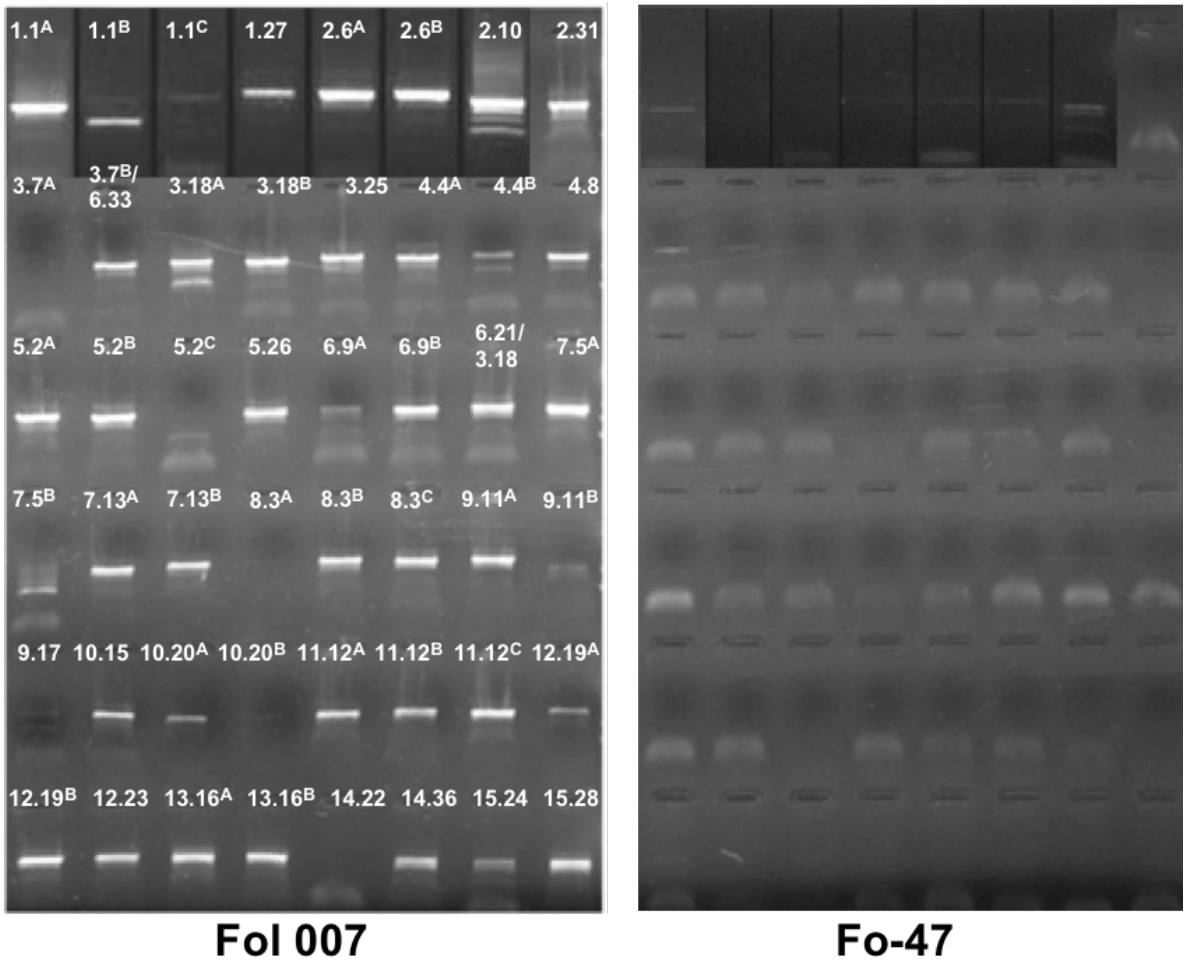

Figure S16. DNA markers based on Foxy insertions. The primer pairs were first tested for the Fol strain used in our study (Fol007), which is closely related to the sequenced isolate, and for Fo-47. Most primer pairs yielded a product with Fol007, but none with Fo-47. Numbers indicate the chromosome (number before the dot) and scaffold (number behind the dot) for which the primer was designed. Some primers are expected to anneal to more than one scaffold /chromosome, due to duplications in the Fol genome. Multiple Foxy insertions on one scaffold are indicated with letters (A, B, C etc.).

Chromosome-specific primers were designed approximately 500 bp upstream of 48 Foxy insertions in the Fol4287 genome sequence released by the Broad Institute (isolate Fol4287) (**Table S17**). Each was used with a reverse primer in Foxy.

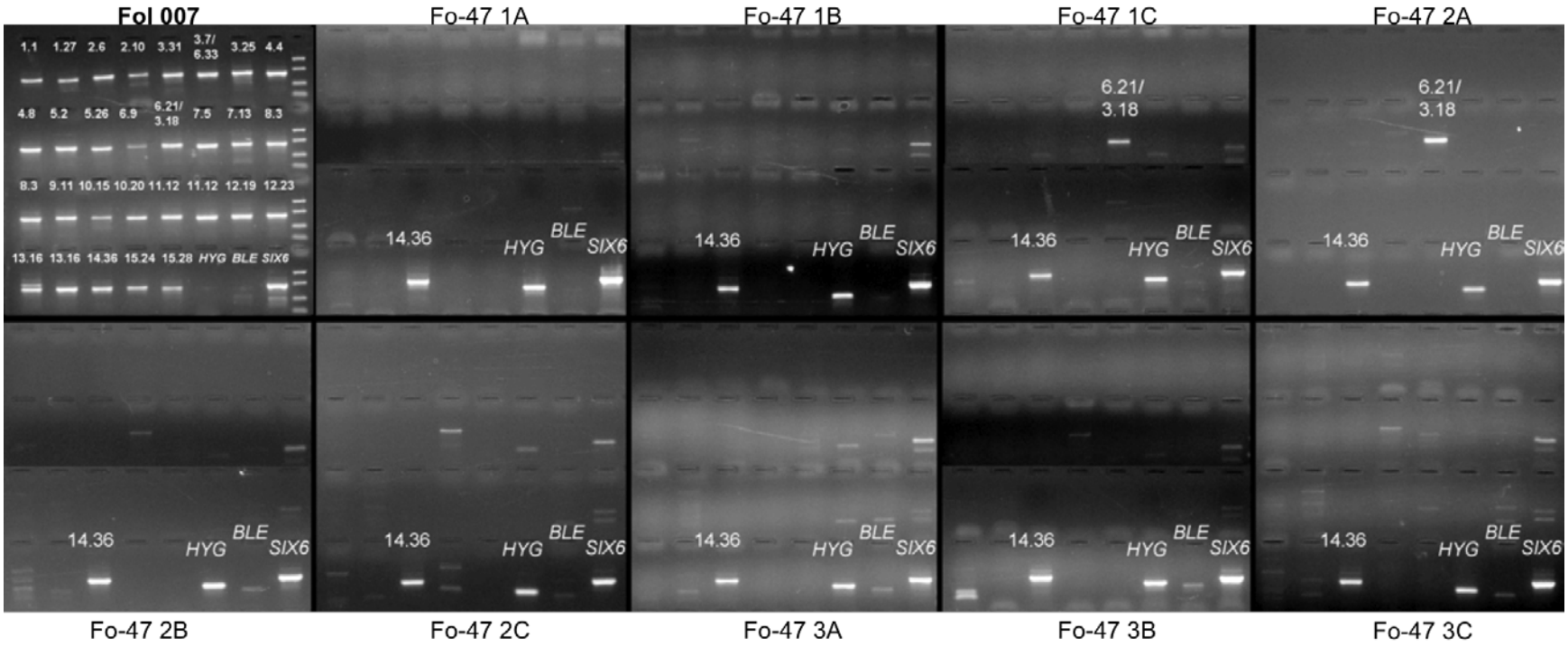

Figure S17. Twenty-nine positive primer pairs were selected for testing the double drug-resistantFo47^{+} strains for the presence of Fol007-derived chromosomes. All double drug-resistant strains were positive for the chromosome 14 marker (14.36, scaffold 36), and two strains (1C and 2A) were also positive for a chromosome 3/6 marker (6.21/3.18, scaffold 21 or 18). In some lanes a specific bands are visible (not labeled). All PCR products of the correct size are indicated with the chromosome/scaffold number.

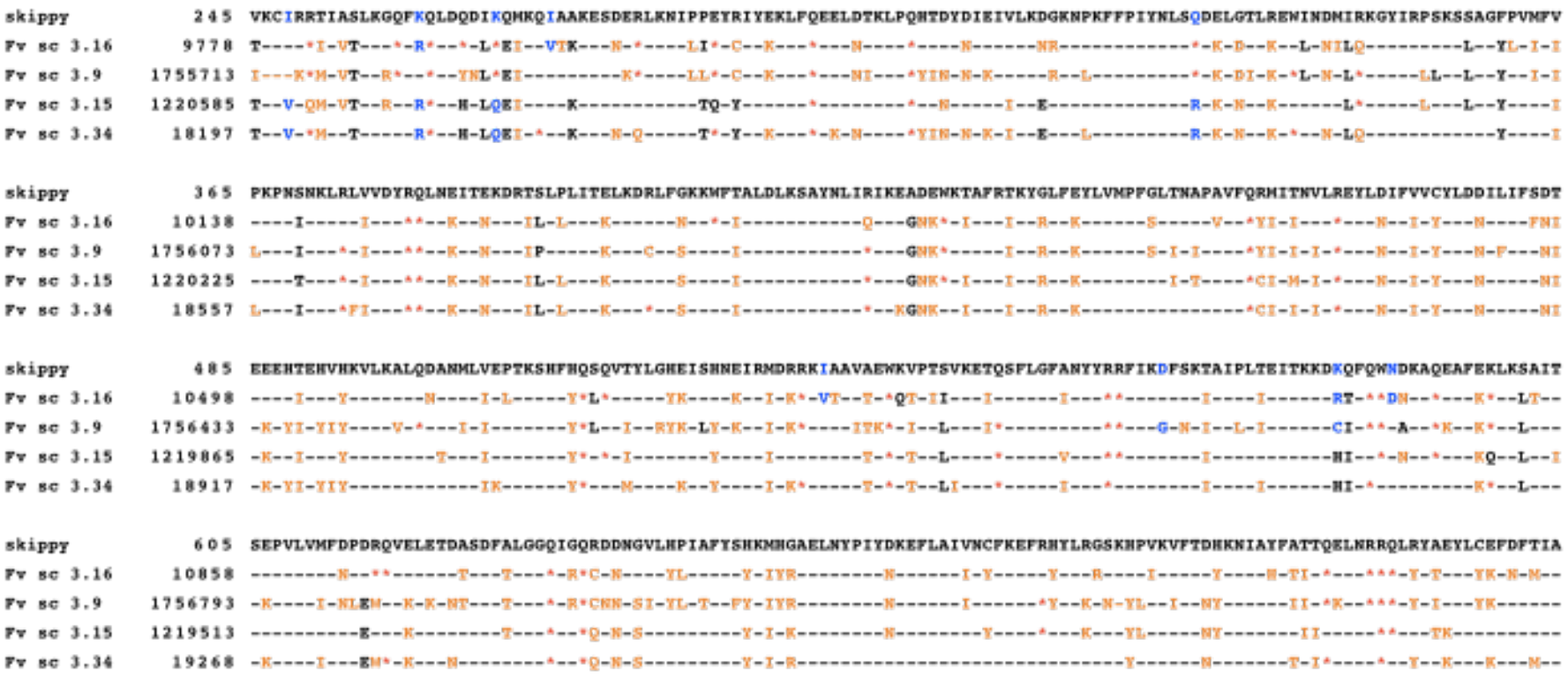

Figure S18. Alignment of four mutated skippy-like TEs from *Fv* to *Fol* Skippy suggests past occurrence of RIP in *Fv*.

Numerous nonsense codons (red) were introduced by RIP-type mutations. Missense mutations explainable by single C:G to T:A mutations from *Fol* Skippy to *Fv* SLRE are shown in orange. Missense mutations that can be explained by RIP of *Fv* SLRE to *Fol* Skippy are shown in blue. All other mutations as shown in black.

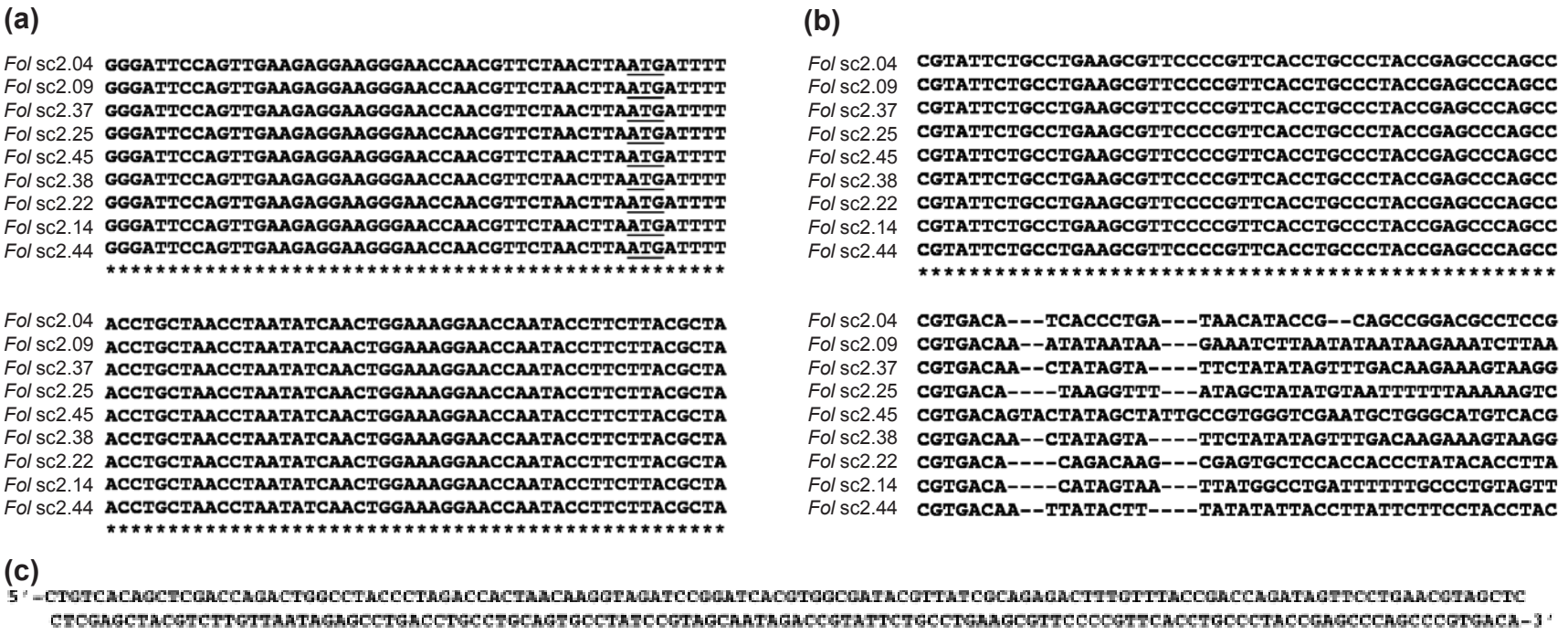

Figure S19. A novel non-RIPped skippy-like retroelement (SLRE) from *Fol*. A similar element had been previously identified as "Skippy" (51) and is related to "Maggy" transposons from *M. grisea* (81). Nine non-mutated *slre* TE coding regions were identified by tblastx searches of the *Fusarium* genome sequences with all previously identified fungal and non-fungal TEs. **(a) partial DNA sequence alignment of skippy elements reveals absence of RIP.** No RIP-type mutations were found across the nine full-length *skippy* elements. Part of the complete alignment is shown from nt 2523592 on the - strand of sc2.04 (the full-length element is 5688 bp long, from sc2.04 nt 2525324-2519636). The predicted Skippy ATG (nt 2523550) is underlined. The preceding *gag* ORF (not shown) is less conserved and characterized by numerous indels. **(b) The 3' boundary of skippy elements.** Top of panel shows 3' end of the completely conserved LTR (ends at CGTGACA in the bottom panel). **(c) Sequence of the identical LTRs.** The 239-bp long LTRs extend from 2519636-2519874 and 2525087-2525325.

(a)

```
Fv_01865    781 FKSRLELWQKH--NRQELPENIVIFRDGVSTGQFAQVLRTELPRIRIACNGKYPKNK---
Fol_03010   783 FKSRLELWQKH--NRQELPENIVIFRDGVSTGQFAQVLRTELPRIRIACNGKYPKNK---
Fg_08752    818 FKSRLALWQMH--NRQQLPENIVIFRDGVSEGQFAQVLQRELPRIRIACNAKYPKNK---
Nc_04730    870 FKTRLELWRSNPANNRSLPENILIFRDGVSEGQFQMVIKDELPLVRAACKLVYPAGK---
Pa_g2992    707 FKSRLRLWQKH--NGAKLPENILIYRDGVSEGQFNMVLTSELPHIRTACSQMYGK-Q---
Fol_12456   784 LKSRLGLWKTKG-KHAALPENILIYRDGVSEGQYDMVLSQELPQLRRACEQMYPAADTKK
Fol_14081   764 LKSRLGLWKTKG-KHAALPENILIYRDGVSEGQYDMVLSQELPQLRRACEQMYPAADTKK
Fol_16455   766 LKSRLGLWKTKG-KHAALPENILIYRDGVSEGQYDMVVSQELPQLRRACEQMYPAADTKK
Ac_01617    772 LKSRLSLWKTKG-KHTALPENILIYRDGVSEGQYDMVLSQELPQLRRACEQVYPTADTKK
Fv_00803    750 FGELIPRWRAN--HPSMVPKHLIYFRDGVGEGQFAYVLDQEVEEIKKYLSTVLPAGQ---
Fol_00711   750 FGELIPRWRAN--HQGLVPKHLIYFRDGVGEGQFAYVLDQEVEEIKKYLATILPGGQ---
Fg_00348    746 FAELLPQWRHN--HPGKIPAHLIYMRDGVSEGQFAHVLEQEVSEIKKFFGGSLPPDK---
Mg_11029    798 FGPLVERWCKT--MR-CAPEHVFYLRDGVSEGQFAHVMALEVAELRKVLNKVG--GN---
consensus   961 fksrl-lWr--------lPenivifRDGVseGQfa-Vl-qElp-irrac--vyp-g----

Fv_01865    836 -APKISIIVSVKRHQARFYPTS-EENAMEKNHNIQNGTVVDRGITEARYWDFYLTAHASI
Fol_03010   838 -APKISIIVSVKRHQTRFYPTS-EENAMEKNHNIQNGTVVDRGITEARYWDFFLTAHASI
Fg_08752    873 -PPRISLIVSVKRHQTRFYPTS-SE-SMTSKNNIENGTIVDRGVTQARYWDFFLTAHSSI
Nc_04730    927 -LPRITLIVSVKRHQTRFFPTD-PKHIHFKSKSPKEGTVVDRGVTNVRYWDFFLQAHASL
Pa_g2992    761 -QPRITLIVSVKRHQTRFYPTD-PQQTHFRSKSPKEGTVVDRGVTNVRYWDFFLQAHASL
Fol_12456   843 GLPRFTIIVCGKRHKTRFYPTT-EQDCDR-SGNTKPGTVVDRGVTEARNWDFFLQAHAAL
Fol_14081   823 GLPRFTIIVCGKRHKTRFYPTT-EQDCDR-SGNTKPGTVVDRGVTEARNWDFFLQAHAAL
Fol_16455   825 GLPRFTIIVCGKRHKTRFYPTT-EQDCDR-SGNTKPGTVVDRGVTEARNWDFFLQAHAAL
Ac_01617    831 GLPRFTIIICGKRHKTRFYPTT-EQDCDR-SGNTKPGTVVDRGVTEARSWDFFLQAHAAL
Fv_00803    805 -MPKFTVIVATKRHHIRFFPQ-----RGDKNGNPLPGTLVEREVTHPFMFDFYLNSHVAI
Fol_00711   805 -MPKFTVIVATKRHHIRFFPQ-----RGDKNGNPLPGTLVEREVTHPFMFDFYLNSHVAI
Fg_00348    801 -IPKMTVIIATKRHHVRFFPQ-----RGDKNGNPLPGTLVEKEVTHPFMFDFYLNSHVAI
Mg_11029    850 -NPKITVMVATKRHHIRFFPKPGDSSSGDRNGNALPGTVVERVVTHPFHYDFYLCSHVAI
consensus  1021 -lPritiiv--KRH-tRFyPts-e-----k-gn--pGTvVdrgvT-ar-wDFfLqaHaai

Fv_01865    894 KGTARPAHYTVLLDEIFQAKFK-------SEAANELEKFTHELCYLFGRATKAVSICPPA
Fol_03010   896 KGTARPAHYTVLLDEIFRSKFK-------SEAANELEKFTHELCYLFGRATKAVSVCPPA
Fg_08752    930 KGTARPAHYTVLLDEVFRAKYG-------AEAANELERYAHELCYLFGRATKAVSICPPA
Nc_04730    985 QGTARSAHYTVLVDEIFRADYG-------NKAADTLEQLTHDMCYLFGRATKAVSICPPA
Pa_g2992    819 QGTARPAHYTVLIDEIFRPSYG-------AQAANNLEQVTHDMCYLYGRATKAVSICPPA
Fol_12456   901 QGTARPCHYYIVHDEIFRQIYAKSIPVPFQSIADIVEDLTHNMCYLL---LENPGIPFPS
Fol_14081   881 QGTARPCHYYIVHDEIFRQIYAKSIPVPFQSIADIVEDLTHNMCYLL---LENPGIPFPS
Fol_16455   883 QGTARPCHYYIVHDEIFRQIYANSIPLPFLSIADIVEDLTHNMCYLFGRATKSVSLCPPA
Ac_01617    889 HGTARPCHYYIVHDAIFRQIYAKLIPSPFQNIADIVEDLTHNMCYLFGRATKAVSLCPPA
Fv_00803    859 QGTARPVHYHIILDEMN-------------MPVNDLQKMIYQQCYSYARSTTPVSLHPAV
Fol_00711   859 QGTARPVHYHVILDEMN-------------MPVNDLQKMIYQQCYSYARSTTPVSLHPAV
Fg_00348    855 QGTARPVHYSVILDEMA-------------MPVNDLQKMIYQQCYSYARSTTPVSLHPAV
Mg_11029    909 QGTARPTHYQVIHDEVG-------------YSPDELQKMLYQQCYQYARSTTPVSLHPAI
consensus  1081 qGTARp-HY-vllDeifr--yg----------aneleklthnlCYlfgratk-vsicppa
```

(b)

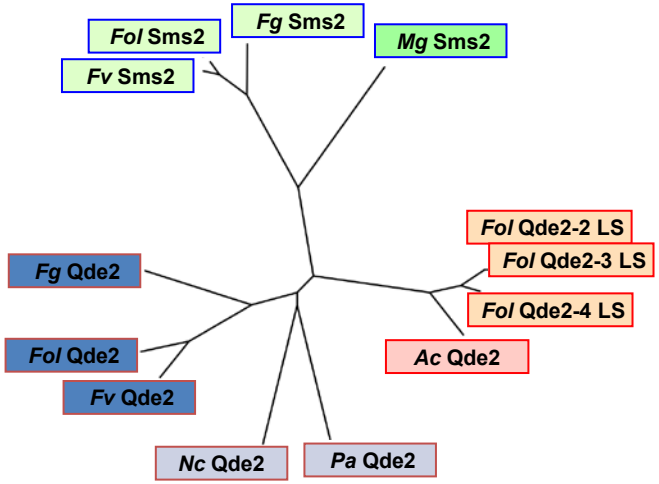

Figure S20. Three conserved *Fol* Qde2 proteins are localized on LS chromosomes. (a) Partial alignment of the QDE-2 PIWI domains. *Fusarium* Qde2 proteins cluster with the *Neurospora* and *Podospora* homologues (top five lines), as do *Fusarium* Sms2 homologues and *Magnaporthe* Sms2 (bottom four lines). Three additional *Fol* Qde2s are more related to *Ajellomyces* Qde2 than the *Fusarium* Qde2s (four center lines). Identical (blue), conserved (cyan) and similar (green) residues are colored; completely different residues are shown in grey. ClustalW was used with default settings. **(b) Relationship between *Fusarium* QDE-2 homologues.** A single ClustalW unrooted tree was constructed with the full-length predicted Qde2 and Sms2 proteins from *Fol*, *Fv*, and *Fg* (see Table 1) and their closest non-*Fusarium* matches among sequenced fungi *(Magnaporthe grisea* Sms2, MGG_11029; *Ajellomyces capsulatum* Qde2, HCAG_01617; *Podospora anserina* QDE-2, PODANSg2992; *Neurospora crassa* QDE-2, NCU04730).

Table S1. Genome sequencing strategy.

Genome	*F. verticillioides*			*F. oxysporum*		
Library	**Reads**	**Physical Coverage (Fold)**	**Sequence Coverage * (Fold)**	**Reads**	**Physical Coverage * (Fold)**	**Sequence Coverage (Fold)**
4kb Plasmid	66,479	3	2.44	444,726	15	4.44
10kb Plasmid	144,437	17	1.11	102,176	9	0.88
40kb Fosmid	70,511	34	1.08	110,003	37	0.96
Plasmid reads (Syngenta)	235,835		3.83			
Total	**517,262**	**54**	**8.46**	**656,905**	**106**	**6.28**

* Q20 base coverage

Table S2. *F. oxysporum* assembly anchored to the optical maps

Optical Contigs	Map estimated size (kb)	Scaffolds	Scaffold size (kb)
I	7,451	14, 1, 27	6,853
II	6,444	6, 10, 31	5,575
III	5,497	47, 18, 32, 7, 25	5,631
IV	5,419	8, 4	5,212
V	5,205	26,2	4,913
VI	5,143	9, 33, 41, 21, 53, 42	4,585
VII	4,376	5, 13	4,346
VIII	3,992	3, 29	3,983
IX	3,434	11, 17	3,304
X	3,422	20, 15, 45	2,896
XI	2,929	35, 12	2,336
XII	2,855	19, 23	2,232
XIII	2,426	16, 39	1,750
XIV	2,269	22, 51, 36	1,645
XV	2,053	37, 38, 24, 28	2,434
Total	**62,917**		**57,697**

Table S3. Enriched gene families in three *Fusarium* genomes.

	Transcription	Carbohydrate-active enzymes[2]					Transporters [3]		
Species	Factors [1]	GH	GT	PL	CBM	Total	ABC	Partial ABC	PDR
Fusarium graminearum	716	242	102	21	37	402	54	0	19
Fusarium verticillioides	683	300	107	23	46	476	57	0	18
Fusarium oxysporum	881	363	124	24	44	555	67	5	25
Aspergillus fumigatus	562	262	100	14	23	399	42	2	15
Neurospora crassa	420	171	74	4	14	263	28	0	7
Magnaporthe grisea	461	231	92	5	36	364	41	1	8
Saccharomyces cerevisiae	295	46	68	0	2	116	22	0	10
Candida albicans	308	58	69	0	1	128	20	0	9
*P value**	2.6E-06	0.01	0.0004	0	2E-11	0.001	7E-9	0.34	5E-7

* The *p*-value (Z score) is calculated using the distribution of gene families in *Fusarium* compared to three other Ascomycete fungi (*Aspergillus fumigatus, Neurospora crassa and Magnaporthe grisea)*

Table S4. Classification of transcription factors in three *Fusarium* species

TF Family Name[a]	*F. graminearum*	*F. verticillioides*	*F. oxysporum*	*F. oxysporum-LS*
APSES	4	3	3	0
AT-rich interaction region	3	3	3	0
Bromodomain transcription factor	3	3	3	0
C2H2 zinc finger	76	67	73	10
CCR4-Not complex component, Not1	1	1	1	0
Centromere protein B, DNA-binding region	3	2	6	0
Cold-shock protein, DNA-binding	1	0	0	0
DDT	1	1	1	0
Forkhead	4	4	5	0
GATA type zinc finger	7	6	8	1
Grainyhead/CP2	1	1	1	0
HMG	39	36	39	5
Heat shock factor (HSF)-type, DNA-binding	3	0	5	2
Helix-turn-helix type 3	1	1	1	0
Helix-turn-helix, AraC type	8	4	4	0
Helix-turn-helix, Psq	0	0	1	0
Heteromeric CCAAT factors	7	6	7	0
Homeobox	13	10	13	0
Homeodomain-like	41	43	63	15
Lambda repressor-like, DNA-binding	2	4	6	2
MADS-box	3	2	4	1
Mating-type protein MAT alpha 1	1	1	1	0
Myb	17	20	22	0
NDT80/PhoG like DNA-binding	3	3	4	0
Negative transcriptional regulator	2	2	2	0
Nucleic acid-binding, OB-fold	48	46	42	3
RFX DNA-binding domain	1	1	1	0
SART1	1	1	1	0
SGT1	1	1	1	0
Transcription factor TFIIS	2	1	2	0
Transcription factor jumonji	6	6	8	2
Winged helix repressor DNA-binding	28	28	35	3
YL1 nuclear protein	1	1	1	0
Zinc finger, BED-type predicted	1	1	1	0
Zinc finger, CCHC-type	12	14	21	6
Zinc finger, DHHC-type	5	3	5	0
Zinc finger, GRF-type	1	0	2	0
Zinc finger, MIZ-type	2	2	2	0
Zinc finger, PARP-type	0	1	0	0
Zinc finger, NF-X1-type	2	0	1	0
Zinc finger, Rad18-type putative	3	3	3	0
Zn2Cys6	313	311	370	27
bHLH	16	15	46	18
bZIP	21	19	55	24
p53-like transcription factor	6	4	6	0
ssDNA-binding transcriptional regulator	2	2	2	0
Total TFs (Genome Ratio %)	**716 (**5.37)	**683 (**4.82)	**881 (**4.97)	**119 (**4.86)

[a]Individual TFs were classified into families based on the DNA binding domain in each TF. [b]The ratio of TFs to the total proteome. [c]This total number may not be same to the sum of TF numbers for each TF family because some TFs have more than one type of DNA-binding motif.

Table S5. Comparison of pectin lyase enzymes

PL family	1	3	4	9	11	20	Total
Aspergillus fumigatus	6	3	3	0	0	1	13
Aspergillus niger CBS 513.88	6	0	2	0	0	0	8
Neurospora crassa OR74A	1	1	1	0	0	1	4
Magnaporthe grisea 70-15	2	1	1	0	0	1	5
Fusarium oxysporum f. sp. lycopersici	11	7	3	2	1	0	24
Fusarium verticillioides	11	7	3	2	0	0	23
Fusarium graminearum	9	7	3	1	0	1	21

Table S6. The five families of *Fusarium* ABC transporters*.

	F. gramineaum	*F. verticilloidies*	*F. oxysporum*
ABCA	FGSG_08373.3	FVEG_02321.3	FOXG_03449.2
		FVEG_12730.3+	FOXG_15400.2+
ABCB	FGSG_01885.3	FVEG_07368.3	FOXG_04248.2
	FGSG_10911.3	FVEG_11590.3	FOXG_12846.2
	FGSG_03805.3	FVEG_12943.3	FOXG_16894.2
	FGSG_05527.3	FVEG_06970.3	FOXG_09366.2
	FGSG_03032.3	FVEG_10569.3	FOXG_11983.2
	FGSG_02749.3	FVEG_08945.3	FOXG_10296.2
	FGSG_04336.3	FVEG_11327.3	FOXG_13902.2
	FGSG_00541.3	FVEG_00998.3	FOXG_00515.2
	FGSG_03323.3	FVEG_08641.3	FOXG_09840.2
	FGSG_06881.3	FVEG_05103.3	FOXG_01943.2
	FGSG_08823.3	FVEG_02402.3	FOXG_03534.2
	FGSG_02786.3	FVEG_13095.3	FOXG_15633.2
	FGSG_12704.3	FVEG_04890.3	FOXG_07972.2
	FGSG_06771.3	FVEG_05216.3	FOXG_02052.2
	FGSG_07516.3	FVEG_12783.3	FOXG_15452.2
	FGSG_11988.3		FOXG_16892.2
			FOXG_11845.2
			FOXG_04837.2
			FOXG_12952.2
			FOXG_04763.2
ABCC	FGSG_06565.3	FVEG_05573.3	FOXG_02392.2
	FGSG_09611.3	FVEG_04167.3	FOXG_06317.2
	FGSG_00669.3	FVEG_00524.3	FOXG_00990.2
	FGSG_11028.3	FVEG_12505.3	FOXG_17424.2
	FGSG_07325.3	FVEG_07940.3	FOXG_01579.2
	FGSG_05571.3	FVEG_06924.3	FOXG_09320.2
	FGSG_09017.3	FVEG_06686.3	FOXG_09082.2
	FGSG_08308.3	FVEG_01831.3	FOXG_02981.2
	FGSG_06141.3	FVEG_04778.3	FOXG_07857.2
	FGSG_09707.3	FVEG_06313.3	FOXG_08360.2
	FGSG_02316.3	FVEG_11763.3	FOXG_13025.2
	FGSG_02139.3	FVEG_07622.3	FOXG_04493.2
	FGSG_04440.3	FVEG_03149.3	FOXG_08622.2
	FGSG_10995.3	FVEG_09941.3	FOXG_05625.2
	FGSG_10547.3	FVEG_10533.3+	FOXG_10860.2
	FGSG_00046.3	FVEG_00329.3	FOXG_11953.2
		FVEG_06048.3+	FOXG_04142.2
		FVEG_07254.3	FOXG_04825.2
		FVEG_03281.3	
		FVEG_03317.3	
ABCD	FGSG_06012.3	FVEG_04642.3	FOXG_07718.2
	FGSG_01526.3	FVEG_09605.3	FOXG_11187.2
	FGSG_07383.3	FVEG_07897.3	FOXG_01535.2

ABCG			
	FGSG_02847.3	FVEG_13230.3	FOXG_15760.2
	FGSG_09329.3	FVEG_03775.3	FOXG_05904.2
	FGSG_11272.3	FVEG_11813.3	FOXG_13381.2
	FGSG_03735.3	FVEG_10580.3	FOXG_11989.2
	FGSG_04580.3	FVEG_11089.3	FOXG_13653.2
	FGSG_08312.3	FVEG_01835.3	FOXG_02985.2
	FGSG_08309.3	FVEG_01830.3	FOXG_02979.2
	FGSG_10577.3	FVEG_13032.3	FOXG_17147.2
	FGSG_08830.3	FVEG_02410.3	FOXG_03541.2
	FGSG_02870.3	FVEG_13167.3	FOXG_15712.2
	FGSG_10935.3	FVEG_11615.3	FOXG_12868.2
	FGSG_05076.3	FVEG_06058.3	FOXG_08610.2
	FGSG_01388.3	FVEG_09464.3	FOXG_11327.2
	FGSG_05589.3	FVEG_06901.3	FOXG_16340.2
	FGSG_08027.3	FVEG_09324.3+	FOXG_07314.2
	FGSG_03882.3	FVEG_13012.3	FOXG_06754.2
	FGSG_11240.3	FVEG_10503.3	FOXG_06510.2
	FGSG_10706.3		FOXG_07190.2
			FOXG_11897.2
			FOXG_17197.2
			FOXG_10675.2
			FOXG_16543.2
			FOXG_04577.2
			FOXG_09300.2

*Shaded genes are orthologous for all three *Fusarium* sp. Red color highlights the transporters unique to *Fusarium*

Table S7. Annotation of the *Fusarium* G protein subunits

Protein Class	Nc Protein	NCU Number	Foxy Number	Fg Number	Fv Number
G alpha Subunit	GNA-1	NCU06493	FOXG_09359.2	FGSG_05535.3	FVEG_06962.3
G alpha Subunit	GNA-2	NCU06729.3	FOXG_05338.2	FGSG_09988.3	FVEG_02792.3
G alpha Subunit	GNA-2	NCU06729.3	FOXG_08807.2		
G alpha Subunit	GNA-3	NCU05206.3	FOXG_06321.2	FGSG_09614.3	FVEG_04170.3
G beta Subunit	GNB-1	NCU00440.3	FOXG_11532.2	FGSG_04104.	FVEG_10291.3
G gamma Subunit	GNG-1	NCU00041.3	FOXG_02181.2	FGSG_07235.3	FVEG_05349.3
RACK1 Homologue	CPC-2	NCU05810.3	FOXG_05557.2	FGSG_09870.3	FVEG_02582.3
Krh1/Gpb1	No Hits	No Hits	No Hits	No Hits	No Hits
Krh2/Gpb2	No Hits	No Hits	No Hits	No Hits	No Hits

Table S8. Predicted G protein coupled receptors (GPCRs) *

Species	*Neurospora crassa*	*Magnaporthe grisea*	*Aspergillus nidulans*	*Fusarium graminearum*	*Fusarium oxysporum*	*Fusarium verticillioides*
Pheromone Receptors	NCU00138 NCU05758	MG06452 MG04711	AN7743 AN2520	FGSG_07270 FGSG_02655	FOXG_02147 FOXG_10633	FVEG_05310 FVEG_09280
cAMP Receptor-Like Proteins	NCU00786 NCU04626 NCU09427	MG11962 MG06257 MG06738	AN3765 AN8262	FGSG_05239 FGSG_09693 (5) FGSG_07716 FGSG_01861 FGSG_03023	FOXG_08375 FOXG_08466 FOXG_04274 FOXG_10838 FOXG_14393(5) FOXG_15087(5) FOXG_15757(4) FOXG_15051(2) FOXG_13170	FVEG_06210 FVEG_06299 FVEG_07395 FVEG_13831(4) FVEG_13224(2) FVEG_12377
Carbon Sensors	NCU06312	MG08803 MG00258	No Hits	FGSG_05006	FOXG_09179	FVEG_06780
Stm1-Related Proteins	NCU00300 NCU09195	MG04698 MG02855	AN5720 AN10166	FGSG_05579 FGSG_08496	FOXG_09314 FOXG_03300	FVEG_06916 FVEG_02168
Microbial Opsins	NCU10055 NCU01735	MG09015	AN3361	FGSG_07554 FGSG_03064 FGSG_01440	FOXG_15406 FOXG_15870(5) FOXG_12142 FOXG_11274	FVEG_12735 FVEG_09519 FVEG_10716
Homologues of *Magnaporthe grisea* PTH11	NCU06531 NCU00700 NCU08624 NCU05854 NCU07649 NCU07591 NCU09201 NCU07538 NCU05189 NCU05307 NCU08447 NCU02903 NCU09823 NCU09796 NCU04106 NCU07769 NCU04931 NCU08718 NCU09022 NCU05101 NCU08431 NCU08429 NCU06891 NCU05187 NCU05829	MG05871 MG10473 MG07553 MG06755 MG09022 MG07565 MG07946 MG11006 MG09070 MG07806 MG03584 MG05214 MG09863 MG10407 MG10571 MG01867 MG09455 MG10050 MG05352 MG07420 MG10442 MG02160 MG02001 MG10257 MG01905 MG07987 MG10438 MG06171 MG07851 MG04935 MG05386 MG09865 MG05514 MG06535 MG01190 MG10581 MG03009 MG10747 MG03935 MG04682 MG09416 MG02692 MG07857 MG00826	AN5639 AN1930 AN2587 AN7774 AN0178 (4) AN7232 AN7395 AN5059 AN2249 AN9444 AN9036 AN5664 AN11159 AN8328 (4) AN1738 AN3395 (5) AN6415 (6) AN8951 AN1540 (5) AN9306 AN8971 AN10886 AN10369 (5) AN2386 (6) AN10357 AN7400 AN4452(5) AN6413 AN8943 AN4378(16) AN2575(5) AN5069 AN5312 AN2683 (6) AN9387 (5) AN0011 (5) AN4213 (4) AN7270(5) AN7523 (6) AN9266(4) AN1557(4) AN3349(8) AN2649 AN2726 (6)	FGSG_03707 FGSG_02155 FGSG_05821 FGSG_03897(6) FGSG_07839 FGSG_00994 FGSG_08408 FGSG_04529 FGSG_11381 FGSG_07663 FGSG_04731 FGSG_02374 FGSG_05039 FGSG_03464 FGSG_07757 FGSG_03005 FGSG_04825 (6) FGSG_04159 FGSG_06541 FGSG_03800 FGSG_02614 FGSG_07136 FGSG_03237 FGSG_00101 (6) FGSG_07489 FGSG_10028 (17) FGSG_03504 FGSG_09370 FGSG_07655 (5) FGSG_03932 (6) FGSG_10085 (6) FGSG_11343 FGSG_03688 FGSG_01989 FGSG_07601 FGSG_11161 FGSG_03164 FGSG_04693 (6) FGSG_03151 (6) FGSG_03409 FGSG_11529 FGSG_03009 FGSG_09352 FGSG_07841	FOXG_15891 FOXG_04477 FOXG_09718 FOXG_11987 FOXG_12029 FOXG_10751 FOXG_04316 FOXG_04048 (5) FOXG_03203 FOXG_14725 FOXG_08949 FOXG_07137 FOXG_16067 FOXG_15611 FOXG_08656 FOXG_09988 FOXG_14513 (6) FOXG_11475 FOXG_02583 FOXG_03452 (6) FOXG_16509 FOXG_01664 FOXG_05380 (15) FOXG_16884 FOXG_17564 FOXG_09687 FOXG_09958 FOXG_04850 (4) FOXG_12335 FOXG_02421 FOXG_17561 FOXG_14891 (6) FOXG_07081 FOXG_14645 (4) FOXG_13506 FOXG_16469 (4) FOXG_05656 (3) FOXG_05684 (6) FOXG_10677 FOXG_08793 (5) FOXG_09789 FOXG_13475 FOXG_04322 FOXG_03973 (5)	FVEG_08756 FVEG_13472 FVEG_09407 FVEG_02076 FVEG_07442 FVEG_07166 (5) FVEG_06016 FVEG_06583 (5) FVEG_08029 FVEG_10345 FVEG_13718 FVEG_13358 FVEG_05766 FVEG_02749 (15) FVEG_10611 (6) FVEG_13015 FVEG_14001 FVEG_11908 FVEG_10583 FVEG_05605 FVEG_07611 FVEG_08525 (6) FVEG_09326 FVEG_00078 (6) FVEG_03296 (4) FVEG_03544 (4) FVEG_03509 (3) FVEG_13899 FVEG_13048 FVEG_08692 (6) FVEG_13491 FVEG_07449 FVEG_05876 (5) FVEG_11641 (6) FVEG_13696 FVEG_14101 FVEG_10934 (5) FVEG_08880 FVEG_10980 FVEG_13747 (6) FVEG_11209 (6) FVEG_07701 (6) FVEG_10789 (6) FVEG_08553

		MG06624 MG00435 MG08653 MG10706 MG04170 MG08525 MG00277 MG02365 MG06595 MG06084 MG09437 MG01890 MG01871 MG03794 MG09667 MG09061	AN0171 AN3257 AN5943 AN7406(3) AN3241 AN2044 AN3348 AN3886(5) AN8548 AN11079(3) AN4642 AN8512 AN8984 AN0857 AN8151 AN6924 AN6946 (5) AN2108 AN2389(3) AN8661 AN1317 (5) AN8727 AN0500 (5) AN0751 AN6622 AN6458 (6)	FGSG_02134 FGSG_05722 FGSG_10630 (6) FGSG_04749 FGSG_07792 FGSG_02981 FGSG_05793 FGSG_03588 FGSG_02401 FGSG_03959 FGSG_03823 FGSG_11598 FGSG_01833 FGSG_03277 FGSG_11385 FGSG_03336 (6) FGSG_00966 FGSG_10946 FGSG_02853 FGSG_02569 FGSG_01806 FGSG_04500 FGSG_04865 FGSG_03215 FGSG_11351 FGSG_11080 (5) FGSG_03091 FGSG_00201 FGSG_10983 FGSG_02334 (6) FGSG_03192 FGSG_03962 FGSG_04815 FGSG_10958 FGSG_02844 FGSG_03310 FGSG_06439 FGSG_02274 FGSG_02402 FGSG_02818 (5) FGSG_03104 FGSG_03476 FGSG_09749 FGSG_13429 FGSG_13487 FGSG_11662 FGSG_12419(5) FGSG_13972 (6) FGSG_04023 FGSG_13461(4) FGSG_12205 FGSG_12474 (3) FGSG_12463(3) FGSG_12490(6)	FOXG_16863 FOXG_16533 (3) FOXG_17122 (5) FOXG_05864 FOXG_12370 (5) FOXG_13042 (6) FOXG_12111 (6) FOXG_15848 (5) FOXG_04515 FOXG_07468 FOXG_13775 (5) FOXG_04594 FOXG_14426 FOXG_15055 FOXG_01279 FOXG_06047 FOXG_09944 (6) FOXG_11631 FOXG_09918 (3) FOXG_12943 FOXG_14011 (6) FOXG_15477 (2) FOXG_14208 FOXG_03205 FOXG_12528 (4) FOXG_13490 FOXG_15682 FOXG_17114 (5) FOXG_07138 (6) FOXG_03847 FOXG_07024 (5) FOXG_04018 (6) FOXG_12023 (6) FOXG_12208 FOXG_07055 (5) FOXG_02918 (4) FOXG_16629 FOXG_02572 FOXG_10585 (5) FOXG_15140 (3) FOXG_12296 (6) FOXG_11781 (4) FOXG_09542 (6) FOXG_10856 (5) FOXG_11694 FOXG_10697 (6) FOXG_09792 (5) FOXG_10478 (6) FOXG_03932 (6) FOXG_08308 FOXG_10038 FOXG_15596 (2) FOXG_15595 (3) FOXG_17695 (4)	FVEG_01609 (4) FVEG_00216 FVEG_03914 FVEG_10190 FVEG_08564 (6) FVEG_00008 FVEG_03741 (6) FVEG_07488 (5) FVEG_12807 (2) FVEG_04399 (5) FVEG_10593 (6) FVEG_08784(4) FVEG_11719 FVEG_12016 FVEG_10615 (5) FVEG_10905 (6) FVEG_10682 (6) FVEG_11941 (5) FVEG_13612 FVEG_13238 (5) FVEG_08439 FVEG_09408 (6) FVEG_13145 (5) FVEG_12373 FVEG_10044 (4) FVEG_10829 (6) FVEG_12935 (3) FVEG_09946 (5) FVEG_09132 (5) FVEG_08689 (2) FVEG_10131 FVEG_09357 (5) FVEG_08466 FVEG_11704 (5) FVEG_06368 (2) FVEG_07719 (5)
Homologues of *Homo sapiens* mPR-like GPCRs	NCU04987 (6) NCU03238	MG05072 MG09091 MG04679	AN4932 AN10630 AN5151	FGSG_04051 (6) FGSG_01064	FOXG_00325 FOXG_12880 FOXG_12495 (6) FOXG_14044(6) FOXG_17118(6) FOXG_14172 (6) FOXG_11596(4)	FVEG_01198 FVEG_11630(6) FVEG_10223(4)
Rat Growth Hormone Releasing Factor-Related Protein	NCU03253(5)	MG00532	AN6680 AN5508 AN3567	FGSG_00527 (5)	FOXG_00532	FVEG_00981 (5)
GprK/AtRGS1-like GPCRs	NCU09883	MG13926 MG11693	AN7795	FGSG_04628	FOXG_13617	FVEG_11041 FVEG_03266 (3)
FlbA	NCU08319	MG14517	AN5893	FGSG_06228 FGSG_03597	FOXG_08482 FOXG_16355 FOXG_06767	FVEG_08855 FVEG_06192

		FOXG_07329 FOXG_06495 FOXG_09613 FOXG_07099 FOXG_17640 (6.5E-39)

* Number of TMs in parenthesis

Table S9. Characterization of the secondary metabolite biosynthesis gene clusters[a]

		PKS	Cluster -id	Size (bp)	Genes	Preferential expression[b]	Known cluster
Non-reducing PKS		FGSG_04588	FG3_26	38179	17	in planta	
		FGSG_03964	FG3_25	29205	9	poor	
		FGSG_02395	FG3_34	29521	8	poor	ZeaA
		FGSG_09182	FG3_15	21303	8	sexual	PGL1
		FVEG_03695	FV3_16	19281	7	Fumonisin	
		FOXG_05816	FO2_8	19694	7		
		FGSG_02324	FG3_33	25247	10	in culture	aurofusarin
		FVEG_03379	FV3_12	18788	7	Fumonisin	
		FOXG_04757	FO2_19	18624	8		
Reducing PKS	Clade I	FGSG_03340	NONE				
		FOXG_11954	NONE				
		FVEG_10535	NONE				
		FVEG_05537	FV3_32	24127	8	Fumonisin	
		FVEG_08425	FV3_41	33102	13		
		FOXG_02884	FO2_31	19644	7		
		FVEG_01736	FV3_10	19430	7		
		FVEG_00316	FV3_21	49188	16	Fumonisin	Fumonisin
		FGSG_01790	FG3_45	17914	5	sexual	
		FVEG_13715	FV3_37	9568	5		
		FOXG_14850	NONE				
		FOXG_15886	NONE				
	Clade II	FGSG_10464	FG3_35	27100	11	sexual	
		FVEG_12610	FV3_27	22802	5		
		FOXG_15296	FO2_23	22800	5		
		FVEG_11932	FV3_5	35392	11	Fumonisin	
		FOXG_03945	FO2_7	42649	16	transposon insertion	
		FOXG_14587	FO2_13	32469	10		
		FVEG_10497	FV3_36	28933	9	Fumonisin	
		FOXG_01189	FO2_14	26645	8		
		FVEG_11086	FV3_17	27587	9	Fumonisin	Fusarin C
		FGSG_07798	FG3_19	26974	9	in planta	Fusarin C
	Clade III	FOXG_02741	FO2_30	22204	9		
		FVEG_09961	FV3_19	12742	5		
		FVEG_00079	FV3_24	14557	4	Fumonisin	
		FGSG_04694	FG3_1	14340	4		
		FVEG_12523	FV3_28	23139	7	Fumonisin	
		FOXG_15248	FO2_22	25367	8		
		FGSG_08208	FG3_20	31544	7	sexual	
		FGSG_08795	FG3_28	22447	6	sexual	
		FVEG_01914	FV3_8	21577	5		
		FOXG_03051	FO2_33	21645	5		
		FGSG_12126	FG3_34	29521	8	poor	ZeaB
		FGSG_10548	FG3_38	17838	4	poor	
		FOXG_10805	FO2_21	28119	9		

		FGSG_05794	FG3_12	19827	6	constitutively expressed	
			FG3_13	19701	12	constitutively expressed	
			FG3_21	27016	12	in planta	trichothecene
			FG3_29	21271	10		
TS			FV3_33	21271	9	Fumonisin	
			FO2_17	18689	8		
			FG3_40	29834	11	sexual	
			FG3_44	26907	8	poor	

[a]. Light blue shaded SMB clusters share orthologous PKS based on the PKS phylogeny (Figure S5) and are also conserved in the flanking genes identified in the cluster. None = no cluster identified with the required four SMB related genes within a 20 kb window.

[b]. The co-expression of the clusters under various experiment conditions using published *Fg* microarray data (13). The clusters in green are specifically expressed in *planta*. The clusters in red are specifically expressed during sexual development; and the clusters in brown are expressed in culture. Fumonisin indicates co-expression of *Fv* gene clusters under culture conditions conducive to fumonisin production in a liquid medium.

Table S10. Pair-wise comparison among three *Fusarium* species.

	F. oxysporum [a]	*F. verticillioides*	*F. graminearum*
F. oxysporum		90%	80%
F. verticillioides	37,302 kb (90%) [b]		80%
F. graminearum	29,018 kb (80%)	28,755 kb (79%)	

[a] Average sequence identity
[b] Total length of syntenic regions (percentage of the genome)

Table S11. Transposable elements in three *Fusarium* genomes.

a.

	ClassI/LTR retrotransposons			DNA transposons						
	SINE	LTRs	Total	Pogo	hAT	Helitron	MITEs	Impala	Others	Total
Fg	360	7,632	8,552	1,076	345	0	0	0	2520	1.421
Fv	248	55,000	56,075	372	999	0	0	0	0	2,018
Fol	159,408	274,097	433,505	491,352	629,444	163,307	21,180	14,563	342,368	1,506,939

b.

	ClassI/LTR retrotransposons			DNA transposons						
	SINE	LTRs	Total	Pogo	hAT	Helitron	MITEs	Impala	Others	Total
Fol_Cons	46.01%	50.98%	49.15%	20.78%	18.54%	29.84%	5.38%	0%	22.76%	5.19%
Fol LS	*53.99%*	49.02%	50.85%	79.22%	81.46%	70.16%	94.62%	100%	77.24%	94.81 %

a). Each type of TE in *Fg*, *Fv* and *Fol* genomes is reported as a total number of bases in each genome.

b). Each type of TE in *Fol* is further located into either *Fol* LS regions or the conserved portion of the genome (*Fol* Cons) and is reported as the percentage distribution between these two regions.

Table S12. Over-represented GO categories of genes encoded in *Fol* LS regions

Go	name	FDR	FWER	p-Value	#Test	#Ref	#nonAnnotTest	#nonAnnotRef
GO:0005576	extracellular region	0.00349266	0.0135177	7.05E-05	19	70	576	6682
GO:0044421	extracellular region part	8.84E-04	0.00154564	1.05E-05	12	23	583	6729
GO:0005625	soluble fraction	4.48E-04	6.72E-04	3.55E-06	12	20	583	6732
GO:0005615	extracellular space	8.87E-04	0.00177332	1.20E-05	10	15	585	6737
GO:0004568	chitinase activity	0.00413742	0.0174305	9.25E-05	8	12	587	6740
GO:0006032	chitin catabolic process	0.00495101	0.0232431	1.39E-04	8	13	587	6739
GO:0006046	N-acetylglucosamine catabolic process	0.00495101	0.0232431	1.39 E-04	8	13	587	6739
GO:0060089	molecular transducer activity	0.00349266	0.0138736	7.61E-05	23	97	572	6655
GO:0016310	phosphorylation	0.00315689	0.0102075	5.20E-05	36	189	559	6563
GO:0004672	protein kinase activity	0.00315689	0.00937104	4.43E-05	29	135	566	6617
GO:0004871	signal transducer activity	0.00349266	0.0138736	7.61E-05	23	97	572	6655
GO:0008986	pyruvate, water dikinase activity	3.65E-06	3.65E-06	2.49E-08	9	3	586	6749
GO:0006094	gluconeogenesis	0.00258178	0.00643367	4.10E-05	9	14	586	6738
GO:0046364	monosaccharide biosynthetic process	0.00579484	0.0316186	1.75E-04	9	18	586	6734
GO:0019319	hexose biosynthetic process	0.00579484	0.0316186	1.75E-04	9	18	586	6734
GO:0016781	phosphotransferase activity, paired acceptors	3.65E-06	3.65E-06	2.49E-08	9	3	586	6749
GO:0006090	pyruvate metabolic process	0.00579484	0.0316186	1.75E-04	9	18	586	6734
GO:0004231	insulysin activity	1.00E-07	2.51E-08	1.46E-10	9	0	586	6752
GO:0000003	reproduction	0.00579484	0.0327718	1.88E-04	16	57	579	6695
GO:0003006	reproductive developmental process	2.32E-04	2.90E-04	1.90E-06	9	8	586	6744
GO:0022414	reproductive process	0.00315689	0.00996744	4.71E-05	13	33	582	6719
GO:0007548	sex differentiation	1.64E-06	8.20E-07	6.73E-09	9	2	586	6750
GO:0004040	amidase activity	0.0011712	0.00263172	1.81E-05	10	16	585	6736
GO:0005871	kinesin complex	0.00695187	0.0408539	1.99E-04	4	1	591	6751

Table S13. Annotation of secreted proteins

Species	Total Secreted	Anchored CPGRP	GPI	Small Cysteine-rich	GH	NEP
Fusarium graminearum	1386	47	88	97	138	4
Fusarium verticillioides	1549	46	86	95	172	4
Fusarium oxysporum	**1803**	**46**	**91**	**126**	**210**	**7**
p-value *	**2E-05**	**0.85**	**3E-05**	**1e-100**	**6E-04**	**0**
F. oxysporum LS specific	253	0	0	26	31	3

* *p*-value is computed using T-test comparing genes in Fol to other two Fusarium genomes. Except GPI anchored proteins, genes Fol LS regions contribute to all other gene family expansions.

Table S14: Enriched glycosyl hydrolases (GHs) in *Fol*.

Family	*F. oxysporum*	*F. verticillioides*		*F. graminearum*		Putative activities
	Num[1]	*Num*	*%*[2]	*Num*	*%*	
GH5	23	19	21%	13	77%	Activity against arabinogalactans
GH43	32	24	36%	18	75%	Arabinanases; β-1,3-galactanases
GH79	6	2	200%	1	500%	β-glucuronidase
GH28	15	9	67%	6	150%	Polygalacturonases, against pectin
GH35	7	4	75%	3	133%	β-galactanases
GH32	14	7	100%	5	180%	Invertases/endo- and exo-inulinases
GH18	30	20	50%	19	58%	Chitinases
GH76	12	9	33%	7	71%	Activity against fungal mannans

[1]The number of the each glycoside hydrolase (GH) gene family is characterized by search against a library of catalytic and carbohydrate-binding modules of carbohydrate-active enzymes (see also www.cazy.org) using blastp.
[2] The increase in *Fol* with respect to the number of genes in the other two genomes.

Table S15. *Fol* carbohydrate-active enzyme (CAZY) genes examined for the presence of transcripts during plant infection.

Predicted gene	CAZY family	Putative activity	Chr(s)	Copy no.	Transcript in *planta*[1]
FOXG_08862	GH28	exo-polygalacturonase pgx4	2a	1	Yes
FOXG_13051	GH28	endo-polygalacturonase pg5	9	1	Yes
FOXG_14695	GH28	endo-polygalacturonase pg1	12	1	Yes
FOXG_12535	GH28	endo-polygalacturonase	3,6	4	Yes
FOXG_14951	GH61	endoglucanase	2b,3	2	Yes
FOXG_14383	GH32	sucrose-6-phosphate hydrolase	1,15	2	Yes
FOXG_14385	GH32	inulinase	1,15	2	yes
FOXG_02657	GH79	β-glucuronidase	8	1	yes
FOXG_06568	GH79	β-glucuronidase	3,6	3	no
FOXG_04665	GH78	α-L-rhamnosidase	4	1	yes
FOXG_15491	GH78	α-L-rhamnosidase	15	2	no
FOXG_09503	GH64	β-1,3-glucanase	9	1	yes
FOXG_12407	GH64	β-1,3-glucanase	3,6	2	no
FOXG_14366	GH76	α-1,6-mannosidase	1,15	2	no
FOXG_06755	GT1	UDP-glucose sterol transferase	3,6	4	no
FOXG_07014	GT4	trehalose phosphorylase	6	1	no
FOXG_07432	GT4	glycosyltransferase	3,4	2	no
FOXG_12436	GT8	glycosyltransferase	2,15	3	no
FOXG_00921	GH18	chitinase 18-3	1	1	yes
FOXG_14329	GH18	chitinase 18-8	1,15	2	yes
FOXG_15172	CBM12	chitin-binding domain	1,15	2	yes

[1]As determined by reverse transcriptase PCR analysis of cDNA obtained from tomato plants infected with *Fg* f. sp. *lycopersici* strain 4287 three days after inoculation (see also Figure S17). Plant CAZYs from GH families encoded in LS regions are coloured pink, those from GT families are coloured green and fungal CAZYs are labelled light blue.

Table S16. Gene-specific primers used for RT-PCR and product sizes.

Gene	gDNA (bp)	cDNA (bp)	Forward primer (5'→ 3')	Reverse primer (5'→ 3')
FOXG_00921	372	311	TTAGTGATGAGTGGGCGGATG	TGGTAGATGAATGCGGACAGC
FOXG_02259	450	395	CCAATGAGCTTCCCGACCTC	CTGGTAGCGGAGAGTCAAGC
FOXG_02657	374	317	AATACCTCTGGGCAATCGAACT	GTCATCGGCGGTAGCATTAGA
FOXG_04665	456	281	TCTGGCATGATGTGGTCGTCT	GAGTATGCGGTCTGCGTGTC
FOXG_06568	472	297	AGTCTAGGCACCCGTGTTTGT	GCCCAGAAGAATCCAGTCAAC
FOXG_06755	476	288	TGCGATATCTGTCCCGATAGG	TTTGAGGATGCCGTTATGAGTG
FOXG_07014	670	474	GCCTTCAAACATGGTCGAAAGT	GCGAAAGCCAACTTGTAGGAG
FOXG_07432	505	444	CTCCTCGGGCGGTTACAAAG	TCATGGATAGACCGTCGTTGG
FOXG_08862	387	328	GTACAGCATTGCCTCGCCAC	CGGGTTTCTCATTCGCAGGTT
FOXG_09503	219	219	TCTGCGAGGGCTTGAAGCGT	CCCATTCTGCCTGCGTGTTG
FOXG_12407	482	355	GATCAGCTTCTCTGCGAGGG	TGTCTCATCTTAATACCGTGTAC
FOXG_12436	495	414	GTTCGACATAAGGATAATACGGA	TATTTCGGAGCCCAGATACTTG
FOXG_12535	426	368	CTGGAAAGACTCTGGACCTGA	TGTGTCCGCCCAGTTTGTCG
FOXG_13051	389	341	GCCTGGTCGCCTCCGTACT	TCTTCTTGCCGCCGTTGCTG
FOXG_14329	391	330	AGCCCAACGAGCCGTCAAGT	AACTGCCGCCAGCGTATGAG
FOXG_14366	470	362	CGTCTGGAATAAGGTTGTCAAG	CCCGCTTTTTCTGCTATTGTGA
FOXG_14383	400	400	GGCTCCTCCGCTTGCTCCT	TGCTGCCATCTTTGCTAGTCG
FOXG_14385	407	407	GAGCAAGAGAAACCGCACGAT	ATCACGCCCAAAGTCGAGCC
FOXG_14695	328	277	GCAGCGTCACTGACTACTCC	GTTAGAACCTTCGCCATCCCA
FOXG_14951	365	365	GCTGGCTCGTTCAGATCCCT	CGGGAGCAGCAGCGGAGAT
FOXG_15172	443	443	GACGCACGCAAGAACGCCTA	ACGCTGTGACACCGCATGGA
FOXG_15491	437	379	CGATTTGGCATGATGTGGTTGT	GCCTCTTGGTGACACGTATTC
FOXG_16779	439	381	GCCCAGCAGCAGTTTACATCT	TGTCCGCAGAGCCTAGATAGA
FOXG_01569 Actin 1	478	429	GAGGGACCGCTCTCGTCGT	GGAGATCCAGACTGCCGCTCAG

Table S17. Lipid metabolism-related gene expansion in *Fol* LS regions*

Species	*F. oxysporum*	*F. verticillioides*	*F. graminearum*	*Neurospora crassa*	*Magnaporthe grisea*	*Aspergillus nidulans*
Homologues						
***Homo sapiens* mPR-like GPCRs**	FOXG_00325	FVEG_01198	FGSG_04051 (6)	NCU04987 (6)	MG05072	AN4932
	FOXG_12880	FVEG_11630(6)	FGSG_01064	NCU03238	MG09091	AN10630
	FOXG_12495 (6)	FVEG_10223(4)			MG04679	AN5151
	FOXG_14044(6)					
	FOXG_17118(6)					
	FOXG_14172 (6)					
	FOXG_11596(4)					
Perilipin-like gene (*CAP20*)	FOXG_07916	FVEG_04839	FGSG_05177	NCU03370	MGG_11916	AN10518
	FOXG_06858					
	FOXG_07131					
	FOXG_07164					
	FOXG_15915					
	FOXG_16152					
	FOXG_16227					
	FOXG_16658					
	FOXG_16721					
Phosphoinositide-specific phospholipase C (PLC)	FOXG_07611	FVEG_04539	FGSG_05898	NCU001266	MGG_02497	AN6389
	FOXG_12081			NCU06245		
	FOXG_06661			NCU09655		
	FOXG_07227			NCU02175		
	FOXG_15032					
	FOXG_06598					
	FOXG_14448					
	FOXG_07102					
Lipase - carboxylic esters	FOXG_15813	FVEG_13281	FGSG_09122	NCU09575	MGG_02497	AN6389
	FOXG_17621	FVEG_12978				
	FOXG_13461	FVEG_13937				
	FOXG_14397					
	FOXG_14401					
	FOXG_14404					
	FOXG_15281					
	FOXG_15077					
	FOXG_15080					
	FOXG_15083					
FlbA	FOXG_08482	FVEG_08855	FGSG_06228	NCU08319	MGG_14517	AN5893
	FOXG_16355	FVEG_06192	FGSG_03597			
	FOXG_06767					
	FOXG_07329					
	FOXG_06495					
	FOXG_09613					
	FOXG_07099					
	FOXG_17640 (6.5E-39)					
Kinases (cdc2)	FOXG_09513	FVEG_07140	FGSG_05393	NCU01266	MGG_02444	AN0664
	FOXG_03267	FVEG_02139	FGSG_08468			
	FOXG_09498	FVEG_07125	FGSG_05406			
		FVEG_10803	FGSG_03132			
	FOXG_17471					
	FOXG_06983					
	FOXG_07291					
	FOXG_06877					
	FOXG_06534					
	FOXG_06733					
	FOXG_03267					
	FOXG_16245					
	FOXG_16169					
	FOXG_06842					
	FOXG_16211					
	FOXG_16133					

*****Yellow highlights the orthologous genes among three *Fusarium* genomes, while all others are Fol LS expansions.

Table S18. Proteins involved in chromatin silencing pathways*.

1. DNA methyltransferase homologues

a. RID homologues (DMT domain of unknown enzymatic function)

Species	Locus	Name	Chromosome	EST
Fg	FGSG_08648.3	FgRID	2	none
Fv	FVEG_02018.3	FvRID	6	none
Fol	FOXG_03151.2	FolRID	8	none

b. DIM-2-type DNA methyltransferase homologues

Species	Locus	Name	Chromosome	EST
Fg	FGSG_10766.3	FgDIM2	3	none
Fv	FVEG_11429.3	FvDIM2	7	incomplete
Fol	FOXG_12693.2	FolDIM2	9	none

2. Histone H3 Lysine9 methyltransferase DIM-5 homologues

Species	Locus	Name	Chr	EST
Fg	FGSG_02778.3	Fg Dim5	1	incomplete
Fv	FVEG_08911.3	Fv Dim5	5	complete
Fol	FOXG_10263.2	FolDim5	7	complete

3. Heterochromatin Protein 1 (HP1) homologues

Species	Locus	Name	Chr	EST
Fg	FGSG_08763.3	FgHP1	2	none
Fv	FVEG_01876.3	FvHP1	6	complete
Fol	FOXG_03020.2	FolHP1	8	complete

4. RNA-dependent RNA Polymerase homologues

Species	Locus	Name	Chr	EST
Fg	FGSG_06504.3	FgQde1	4	incomplete
Fv	FVEG_05646.3	FvQde1	3	none
Fol	FOXG_02461.2	FolQde1	5	none
Fg	FGSG_08716.3	FgSad1	2	none
Fv	FVEG_01945.3	FvSad1	6	none
Fol	FOXG_03081.2	FolSad1	8	none
Fg	FGSG_01582.3	FgRdp3	1	none
Fv	FVEG_09667.3	FvRdp3	1	incomplete
Fol	FOXG_11123.2	FolRdp3	1	incomplete

5. Dicer homologues

Species	Locus	Name	Chr	EST
Fg	FGSG_09025.3	FgDcl1	4	none
Fv	FVEG_06696.3	FvDcl1	7	none
Fol	FOXG_09093.2	FolDcl1	9	none
Fg	FGSG_04408.3	FgDcl2	2	none
Fv	FVEG_11254.3	FvDcl2	9	incomplete
Fol	FOXG_13826.2	FolDcl2	10	incomplete

6. Argonaute homologues

Species	Locus	Name	Chr	EST
Fg	FGSG_08752.3	FgQde2	2	Incomplete
Fv	FVEG_01865.3	FvQde2	6	Incomplete

Fol	FOXG_03010.2	FolQde2	8	Incomplete
Fol	**FOXG_12456.2**	**FolQde2-2**	**3**	**none**
Fol	**FOXG_14081.2**	**FolQde2-3**	**6**	**none**
Fol	**FOXG_16455.2**	**FolQde2-4**	**14**	**none**
Fg	FGSG_00348.3	FgSms2	1	none
Fv	FVEG_00803.3	FvSms	1	none
Fol	FOXG_00711.2	FolSms2	1	none

* Loci were identified by using *Fg* or *Neurospora crassa* predicted and known protein sequences as baits in tblastn searches of the *Fusarium* genomes with default settings. All genes except for the three non-orthologous *QDE2* homologues of *Fol* are located on non-LS chromosomes. The non-orthologous *Fol QDE2* proteins (shown on bold) are more closely related to *Ajellomyces capsulatus*, *Aspergillus* spp. and *Pyrenophora tritici-repentis*.

Table S19. Pearson correlation coefficient (*r*) of codon usage between chromosome pairs.

	Chr1	Chr2	Chr4	Chr5	Chr7	Chr8	Chr9	Chr10	Chr11	Chr12	Chr13	Chr3	Chr6	Chr14
Chr2	0.999													
Chr4	0.999	0.999												
Chr5	0.999	1.000	0.999											
Chr7	0.999	0.999	1.000	1.000										
Chr8	0.999	0.999	1.000	1.000	1.000									
Chr9	0.999	0.999	0.999	0.999	0.999	0.999								
Chr10	0.999	0.999	1.000	0.999	1.000	0.999	0.999							
Chr11	0.994	0.996	0.998	0.997	0.997	0.997	0.997	0.998						
Chr12	0.994	0.996	0.997	0.996	0.996	0.997	0.996	0.997	0.999					
Chr13	0.993	0.996	0.997	0.996	0.996	0.997	0.996	0.997	0.999	0.999				
Chr3	0.992	0.992	0.990	0.992	0.990	0.991	0.990	0.991	0.989	0.989	0.989			
Chr6	0.992	0.993	0.991	0.993	0.991	0.992	0.991	0.992	0.990	0.990	0.990	1.000		
Chr14	0.987	0.990	0.988	0.990	0.987	0.988	0.987	0.988	0.987	0.986	0.988	0.997	0.997	
Chr15	0.993	0.992	0.991	0.993	0.992	0.992	0.991	0.992	0.988	0.989	0.987	0.998	0.998	0.993

Table S20. Amino acids with significant codon bias in the *Fol* LS chromosomes

Amino acid	Codon	Number of codons in conserved regions (%)		Number of codon in *Fol* LS regions (%)		P-value*
Q	CAA	166721	53.74%	22461	48.79%	1.00E-86
	CAG	143544	46.26%	23576	51.21%	1.00E-86
	total	310265		46037		
C	TGT	87368	45.16%	11409	42.75%	1.00E-13
	TGC	106112	54.84%	15281	57.25%	1.00E-13
	Total	193480		26690		
A	GCA	127397	24.79%	19537	24.72%	0.35
	GCC	135711	26.40%	22477	28.44%	1.00E-30
	GCG	94290	18.35%	15659	19.81%	1.00E-20
	GCT	156573	30.46%	21357	27.02%	1.00E-80
	total	513971		79030		
G	GGC	131445	29.97%	22031	32.31%	1.00E-30
	GGG	72143	16.45%	13128	19.25%	1.00E-70
	GGA	123164	28.09%	17921	26.28%	1.00E-20
	GGT	111775	25.49%	15107	22.16%	1.00E-78
	Total	438527		68187		
V	GTC	129752	32.55%	18820	32.07%	0.02
	GTA	64982	16.30%	10286	17.53%	1.00E-10
	GTG	89308	22.41%	14147	24.11%	1.00E-19
	GTT	114564	28.74%	15430	26.29%	1.00E-30
	Total	398606		58683		
T	ACC	118093	34.71%	16778	34.20%	0.06
	ACA	129291	38.01%	17878	36.44%	1.00E-07
	ACT	117785	34.62%	15832	32.27%	1.00E-18
	ACG	93106	27.37%	15351	31.29%	1.00E-60
	Total	340182		49061		
E	GAA	153574	46.16%	23891	45.15%	1.00E-05
	GAG	179115	53.84%	29019	54.85%	1.00E-05
	Total	332689		52910		

* p-value

determined using Fisher's exact test to examine the significance of codons used between conserved versus LS regions. Pink indicates codons with increased usage among the *Fol* LS genes. Blue indicates the codons with decreased usage among the *Fol* LF genes.

Table S21 Distance matrix of *Fol* LS homologous genes among selected fungal genomes*

	*Fo*LS genes	*Fs* homo	*Fo* conserved homo	*Fv* homo	*Fg* homo	*M. grisea* homo	*N crassa* homo	*A. nidulans* homo	*A. oryzae* homo
Fo LS genes	-	0.436	0.487	0.542	0.538	0.682	0.721	0.714	0.703
Fs homologs	1058	-	0.148	0.195	0.177	0.260	0.292	0.300	0.294
*Fo*_conserved homologs	1180	358	-	0.111	0.115	0.238	0.265	0.269	0.269
Fv homologs	1313	473	269	-	0.082	0.201	0.209	0.227	0.229
Fg homologs	1304	430	278	199	-	0.194	0.203	0.229	0.222
M. grisea homologs	1654	630	576	487	470	-	0.099	0.126	0.126
N crassa homologs	1747	707	643	506	491	241	-	0.153	0.154
A. nidulans homologs	1731	727	651	550	555	305	372	-	0.080
A. oryzae homologs	1703	713	651	554	537	305	374	194	-

* The distance is measured by the presence and absence of homologous gene in each genome through blastp search (1e-20). Only genes encoded in the conserved regions of Fol genome are used for the search within Fol genome. Below diagonal: Total character differences; Above diagonal: Mean character differences.

Table 22: Peptides from Six5, Six6 and Six7 identified with mass spectrometry.

Protein	Peptide	Peptide mass (MS) [1]	Amino acid sequence(MS/MS) [1]
Six5	AANGKIDGDTFYNLCINNGGADSTCFDCSK	spot 17	p12 [2]
	IDGDTFYNLCINNGGADSTCFDCSK	spot 17	
	CINNGGADSTCFDCSK	spot 17	spot 17
	NVIYCDAA	spot 17	p12 [2]
	FAGIYLK	spot 17	spot 17 [3]
	TSPAYWFADR	spot 17	spot 17
	SYLWAQTSPAYWFADR	spot 17	
	RDHQYCACQSGSGDSIDIDATTQLQNDNSK	spot 17	
Six6	SFCVANPR	spot 16	spot 16
	ITDTPCQPR	spot 16	spot 16
	EKITDTPCQPR	spot 16	spot 16
	DTLPVSTCPAGQKYDR	spot 16	
	EGCTTTSVNPAGYHHLGTIVYDINK	spot 16	
	EGCTTTSVNPAGYHHLGTIVYDINKNPIEVDK	spot 16	
	ISYFGEPGNVNEGIGGSTSYFSSDNFQFSK	spot 16	
Six7	TTFVEVR	spot 21	spot 21
	QGQCFSTTGSTPPRPPPAAR	spot 21	spot 21
	EVTFDITQNVNTFTSAASTPWTEGVGLSNIR	spot 21	
	VFGTAEAQVVLLPDAPGTSR	spot 21	

[1] spot numbers in the 2-D gel with pH range 3-10 as published in Houterman *et al.* (2007) [76].

[2] p12: Found among previously unidentified peptides in 12 kDa protein band reported in Rep *et al.* 2004 [77].

[3] also found in spot 18 of the gel with pH range 6-11[76].

Table S23: *SIX* genes used in this study.

Gene	Accession	Reference [1]	Length [2]	Cys[3]	Introns	Position on chromosome 14 [4]
SIX1	AJ608702	Rep *et al*. 2004 [77]	263 [189]	8	0	scaffold 36: 116409-116848
SIX2	AJ608703	Houterman *et al*. 2007 [76]	212 [172]	8	0	scaffold 36: 108933-109802
SIX3	AM234063	Houterman *et al*. 2007 [76]	144	2	0	scaffold 36: 5017-5484
SIX5	FJ767863	unidentified #2 PH2007 [76]	102	6	3	scaffold 36: 3273-3407
SIX6	FJ755835	unidentified #3 PH2007 [76]	195	7	0	scaffold 22: 761399-762125
SIX7	FJ755836	unidentified #5 PH2007 [76]	201	3	0	scaffold 51: 65216-65875
ORX1	AM236600	Houterman *et al*. 2007 [76]	588	2	7	scaffold 22: 88830929-833119

[1] PH2007: previously described as unidentified, in Houterman *et al.*, 2007 [76].
[2] In amino acids, without predicted signal peptide; between brackets without predicted prodomain.
[3] Number of cysteine residues.
[4] In strain 4287 (http://www.broad.mit.edu/).

Table S24. Double resistant colonies per million (10^6) colony forming units.

	Experiment						
	1 (a)	2 (b)	3 (b)	4 (c)		5 (c)	
receiving strain				average	± st. dev.	average	± st. dev.
Fol 7	0.13	0.69		0.04	± 0.03	0	± 0
Fol 8	0.12			0.06	± 0.04	0	± 0
Fol 9	0.08	11.16					
Fom H1*	0.05	0					
Fom H2*	30			0	± 0	0	± 0
Fom H3*	10	6.98		1.17	± 0.33	6.98	± 4.25
Fo H1*	1.77	2.73					
Fo H2*	0.35		0.84	1.13	± 0.10		
Fo H3*	1.85	0	2.1	0.79	± 0.45		
Foc H1	0	0					
Foc H2	0						

empty box: not determined

(a): Mixed spores were used to incubate four different co-incubation plates. Spores from the co-incubation were pooled before plating on selective medium.

(b): Mixed spores were used to incubate only one co-incubation plate.

(c): Mixed spores were used to incubate three different co-incubation plates. Spores from the co-incubation plates were plated separately on selective medium and the average number of double resistant colonies was calculated.

* double resistant colonies from this experiment were used for further analyses.

Table S25A. Chromosomal –specific primers based Foxy insertion sites in each chromosome

Primer name*	Sequence	Band in Fol007**	Selected as markers
chr1sc1fwd1L	ATATGAAAGGGGTTCAAGGA	no (smaller band)	
chr1sc1fwd2K	GATCTAATTGAGTTATGAGCC	no	
chr1sc1fwd3L	TCCTGAGAAACTTTCTCAGC	YES	X
chr1sc27fwd1K	TTACCGAGCAGGAGACGAGC	YES	X
chr2sc6fwd1L	AAAAGAATTGCTCGCGCTCT	YES	X
chr2sc6fwd2K	GAATTCTCTCGCGCATCACT	YES	
chr2sc10fwd1L	GAATTCCTTGTAGAGGTGCT	YES	X
chr2sc31fwd1K	TCTACCCGGGTCTTCCAACG	YES	X
chr3sc7fwd1L	GTACCGATAATATGCCATAC	no	
chr3sc7fwd2K-chr6	GTATGCCACTCAAAACAATC	YES (also anneals to chr6sc33)	X
chr3sc18fwd1L	GAGCAGATGGGCGTTCCCTA	YES	
chr3sc18fwd2K	GTGTACGAATCTTACAATGAC	YES	
chr3sc25fwd1L	ATTTACAGGGTTGCGCGAAC	YES	X
chr4sc4fwd1K	GGCACTATCAACTGTTAATC	YES	X
chr4sc4fwd2L	AGTCCCTTGCGTCCAACCGA	YES (weak)	
chr4sc8fwd1K	TAGATCCTCCCCTGGTGTCC	YES	X
chr5sc2fwd1L	ACCTCCCGAGAAGGTTATCA	YES	X
chr5sc2fwd2K	CAAACTCTAATACTCAACGG	YES	
chr5sc2fwd3L	CCCATGCAGCTTCAAAGTCG	no	
chr5sc26fwd1K	AAAGTTGGCGAACTTAGGGA	YES	X
chr6sc9fwd1L	ATTCATTGTCACATGCAGCC	YES (weak)	X
chr6sc9fwd2K-chr3	CATCAACCAGTTATGACATC	YES	
chr6sc21fwd1L	ACAAGACGAAAAAGGCATGG	YES (also anneals to chr3sc18)	X
chr7sc5fwd1K	GAGGAGATCAAGGACATTTT	YES	X
chr7sc5fwd2L	TCAGCGGGTTCGAGGGCAAT	no (smaller band)	
chr7sc13fwd1K	GGAAAGTCTTCCGTGGCTTG	YES	
chr7sc13fwd2L	CTTGTCATTTCTGAGCCTTG	YES	X
chr8sc3fwd1K	ACCTTGGAGGAAATAAACTG	no	

chr8sc3fwd2L	CACTTCGAATACGTAGGCTG	YES	X
chr8sc3fwd3K	ACGTGCAGTGCATTCATGGC	YES	X
chr9sc11fwd1L	TATGCTGTTCCCCAGTCGCT	YES	X
chr9sc11fwd2K	TGAAGTGGACTAAGGAGGAG	no (smaller band)	
chr9sc17fwd1L	CTCCTTACTTCAACGGCGTC	no	
chr10sc15fwd1L	GCATAAACGCCGAGCCTCCA	YES	X
chr10sc20fwd1K	GTTCTGAATGTCCTGAGGAA	YES (weak)	X
chr10sc20fwd2L	GGACCTATCAAGTTCCGTGT	no	
chr11sc12fwd1K	GGGCATAATTAAGCAGTGAT	YES	X
chr11sc12fwd2L	GGACGGAGGGTAACAGGTAC	YES	
chr11sc12fwd3K	GCGAAAACCTTGCTTTTAAG	YES	X
chr12sc19fwd1L	GTTATAACGAACTGCGGTGG	YES (weak)	
chr12sc19fwd2K	GGGCATTTATATGGTTCAGT	YES	X
chr12sc23fwd1L	GAACGCTCGCGTGATGAAGC	YES	X
chr13sc16fwd1K	TACTGCTTGCGTCGGGGATG	YES	X
chr13sc16fwd2L	AGGCATCGTTTAATCGTTGG	YES	X
chr14sc22fwd1K	ATCTCATAATCTGGCGGCTC	no	
chr14sc36fwd1L	ACTATTAACAATCACGAGGAC	YES	X
chr15sc24fwd1L	ACCGCACAAGACTCAACTAT	YES (weak)	X
chr15sc28fwd1K	TTTAGGGCCCACGTGCAGAC	YES	X
HYG-F	GTTGGTCAAGACCAATGCGGAG		
HYG-R	TGCACAGGGTGTCACGTTGC		
BLE-F	TCCTTCCCATCCCTTATTCC		
BLE-R	CACGAAGTGCACGCAGTTG		
Foxy-primer	GAGAGAATTCTGGTCGGTG		X

* The number behind 'chr' refers to the chromosome and the number behind 'sc' refers to the super contig on which the primer was designed. Two primers also anneal to a second supercontig on another chromosome (both on chr. 3 and 6).
** Used in combination with Foxy-primer as a reverse primer.

Table S 25B Primers based on Foxy insertions covering the duplicated region in Chr3/6

Primer name*	Sequence	Anneals on: chr 3	Anneals on: chr 6	Used in suppl. fig.	Band present in: ** Fol007	Fo-47	1B	1C	2A	2B
chr3sc32fwd1K	GTATGCCACTCAAAACAATC	SC 32	SC 33		X					
chr3sc18fwd3L	TTGCAGGAGAATATCTAGAG	SC 18	SC 21		X			X	X	
chr3sc18fwd4K	TATCAAGAGATATGCAGAGC	SC 18	SC 21		X			X	X	
chr3sc18fwd1L	GAGCAGATGGGCGTTCCCTA	SC 18	SC 21	S20 A	X			X	X	
chr3sc18fwd5K	ACAGGAAGAGCACCACCAAA	SC 18	SC 21		X			X	X	
chr6sc21fwd1L	ACAAGACGAAAAAGGCATGG	SC 18	SC 21	S20 A/B	X			X	X	
chr3sc18fwd6K	AGTGATAGAAAGTTGAGAAAG	SC 18	SC 21		X			X	X	
chr3sc18fwd7L	TGCAGGGTGCCCTAGATTG	SC 18	SC 21		X			X	X	
chr3sc18fwd8K	TGCATGCAGGATAATCAATC	SC 18	SC 21		X			X	X	
chr3sc18fwd9L	TATGCGTGAAGTTCTTCGGC	SC 18	SC 21, 41		X	X	X	X	X	X
chr3sc18fwd2K	GTGTACGAATCTTACAATGAC	SC 18	SC 21, 41	S20 A	X			X	X	
chr3sc18fwd11L	TCTAATCAAACGCTTCAGTC	SC 18	SC 21, 41		X			X	X	

* the number behind 'chr' refers to the chromosome and the number behind 'sc' refers to the super contig on which the primer was designed.
** used in combination with Foxy-primer as a reverse primer.