Host specificity in Fusarium oxysporum

van Dam, P.

Publication date
2017
Document Version
Final published version
License
Other

Host specificity in *Fusarium oxysporum*

Peter van Dam

Host specificity in *Fusarium oxysporum*

Peter van Dam

2017

Host specificity in *Fusarium oxysporum*

Academisch Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit van Amsterdam op gezag van de Rector Magnificus prof. dr. ir. K.I.J. Maex ten overstaan van een door het College voor Promoties ingestelde commissie, in het openbaar te verdedigen in de Agnietenkapel op vrijdag 8 december 2017, te 14:00 uur

door
Peter van Dam
geboren te Rotterdam

Promotiecommissie

Promotor:	Prof. dr. B.J.C. Cornelissen (Universiteit van Amsterdam)
Copromotor:	Dr. M. Rep (Universiteit van Amsterdam)
Overige leden:	Prof. dr. M.W. Prins (Universiteit van Amsterdam)
	Prof. dr. S. Brul (Universiteit van Amsterdam)
	Prof. dr. A.F.J.M. van den Ackerveken (Universiteit Utrecht)
	Prof. dr. B. Lievens (Katholieke Universiteit Leuven)
	Dr. P.M. Bleeker (Universiteit van Amsterdam)
Faculteit:	Faculteit der Natuurwetenschappen, Wiskunde en Informatica

The research described in this thesis was carried out at the Molecular Plant Pathology laboratory at the Swammerdam Institute for Life Sciences, Faculty of Science of the University of Amsterdam. The work was funded by a Netherlands Genomics Initiative grant to Martijn Rep (project 93512007).

Voor papa

Cover design: Machiel Beijaert
Printing and binding: Boekendeal.nl

ISBN 9789492597076

Table of contents

Chapter 1

General introduction

General introduction

Fusarium oxysporum (Fo) is an ascomycetous fungus that is found commonly in soils worldwide. It is apparently asexual and is considered a species complex harbouring several 'phylogenetic species' (Laurence *et al.*, 2014). Taken together, pathogenic isolates from the Fo species complex (FOSC) infect a wide range of plant species including important crops. Individual pathogenic isolates of *F. oxysporum*, however, usually only infect one or a few plant species (Armstrong and Armstrong, 1981). The question what genetic factors underlie this host specificity is the main focus of this thesis.

Fusarium oxysporum as a pathogen

Colonization of a plant by Fo starts by growth of the fungus towards and into the root, followed by entry into the xylem vessels (Fig. 1A). It has frequently been observed, however, that even entry into xylem vessels does not necessarily result in disease (Gao *et al.*, 1995; Jiménez-Fernández *et al.*, 2013; Gordon, 2017). The plant's initial response upon recognition of the presence of the fungus includes production of anti-microbial compounds and pathogenesis-related (PR) proteins (Rep *et al.*, 2002). Formation of tyloses and pectic gums is also induced to block the xylem vessels and prevent the fungus from spreading further. This response occurs both in incompatible and compatible interactions but in the latter the reaction appears to be too slow to prevent systemic infection. Hyphal growth through the vascilature allows rapid upward movement (Michielse and Rep, 2009). The deterioration of xylem tissue eventually results in reduced water flow through the vessels, causing the typical wilt symptoms observed in affected plants.

It is important to realise that the vast majority of *F. oxysporum* strains is not pathogenic and some can even be beneficial to plants (Alabouvette *et al.*, 1979; Correll *et al.*, 1986). Most research attention, however, has been paid to pathogenic strains of Fo because they cause serious vascular wilt and cortical rot disease in a wide variety of agricultural crops. Strains are classified into host-specific forms (*formae speciales*, ff. spp.) and are often further subdivided into races or pathotypes based on their capacity to infect different cultivars of a plant species (Armstrong and Armstrong, 1978; Del Mar Jiménez-Gasco and Jiménez-Díaz, 2003; Inami *et al.*, 2012). The disease does not only affect vegetable crops; also flowers, field crops and plantation crops suffer from Fo infections (Michielse and Rep, 2009). Examples of economically important *formae speciales* include *melonis* (abbrev. Fom, on muskmelon; Fig. 1B-C), *tulipae* (on tulips), *cubense* (on banana) and *asparagi* (on asparagus plants). So far, little is known about the genetic basis for host-specificity in the FOSC. Intriguingly, strains belonging to different *formae speciales* can be more related based on conserved gene sequence similarity than strains belonging to the same *forma specialis* (Kistler, 1997; Lievens *et al.*, 2009). For example, based on conserved genes such as the *EF1-alpha* gene, non-pathogenic

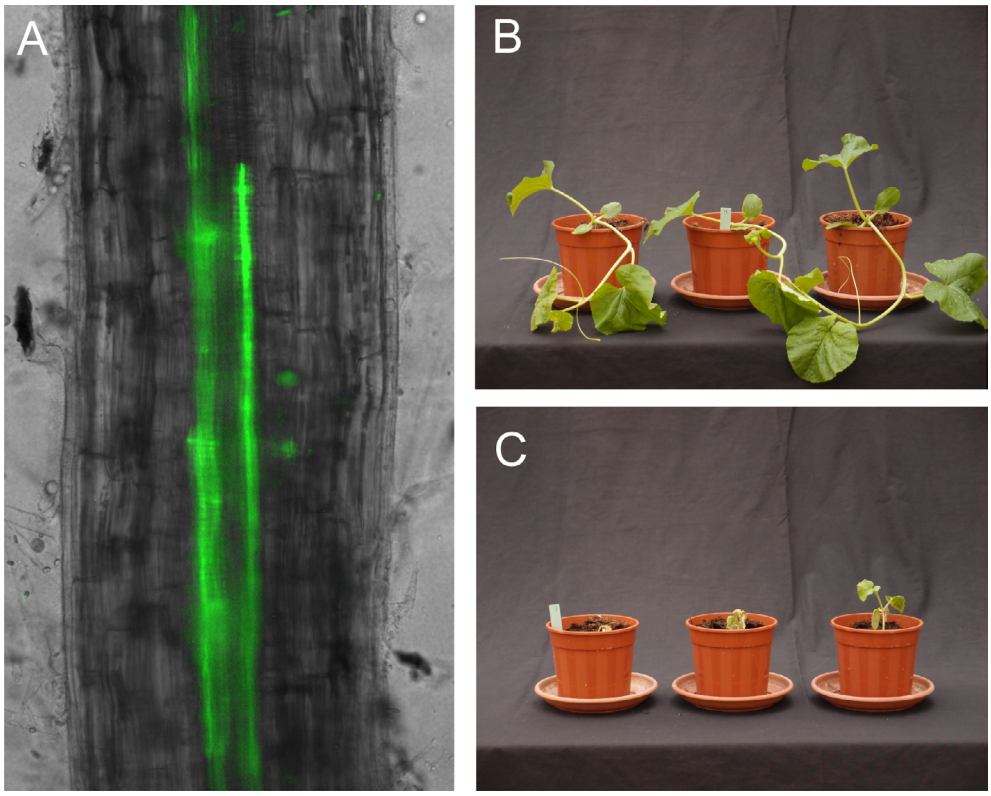


Fig. 1: *Fusarium oxysporum*-induced disease development.
(A) Fluorescently tagged Fo f. sp. *radicis-cucumerinum* (in green) colonizing the vasculature of a cucumber seedling, nine days after inoculation. **(B)** Healthy (mock-treated) melon plants compared to **(C)** melon plants infected by Fo f. sp. *melonis,* two weeks after inoculation.

strains are dispersed among pathogenic strains in phylogenetic trees and can be very closely related to pathogenic strains (Baayen *et al.*, 2000).

Control of *Fusarium* disease

Currently, there are no effective curative treatments for *Fusarium* disease control (Lievens *et al.*, 2006). Use of resistant varieties or rootstocks is the only practical measure for controlling the disease in the field (Pavlou *et al.*, 2002; Cohen *et al.*, 2014, 2015). Other methods are being explored, such as the use of antagonistic microorganisms like Pseudomonas spp., Trichoderma spp. and endophytic strains of *F. oxysporum* (Dubey *et al.*, 2007; Alabouvette *et al.*, 2009) and the creation of disease-suppressive conditions by incorporating composts into the soil (Reuveni *et al.*, 2002; Rose *et al.*, 2003; Yogev *et al.*, 2006). In glasshouses, soil sterilization by fumigation with methyl bromide can be performed (Pavlou *et al.*, 2002; Michielse and Rep, 2009).

Fo, like many other plant pathogenic microorganisms, can be spread through contaminated soils, plant tissues and seeds. Seed transmission of pathogenic strains of *F. oxysporum* was first found in watermelon (Martyn, 2014), both internally (Fulton and Winston, 1913) and externally (Porter, 1928). It has since been described for ff. spp. affecting numerous other crop plants, including each of the major cucurbit wilt *formae speciales* (watermelon, melon, cucumber and bottle gourd (Kuniyasu, 1980; Martyn and Vakalounakis, 2012)) and other plants such as chickpea, lettuce, cotton and basil (Chiocchetti *et al.*, 1999; Garibaldi *et al.*, 2004; Pande *et al.*, 2007; Bennett *et al.*, 2008). Severe outbreaks and quick spread of *Fusarium* disease is known in – amongst others – banana ("Panama disease" caused by Fo f. sp. *cubense* tropical race 4) and date palm ("Bayoud disease" caused by Fo f. sp. *albedinis*) (Daayf *et al.*, 2003; Ordonez *et al.*, 2015).
Fo can be very persistent in soils due to the formation of chlamydospores (Nelson, 2012). Rapid and reliable identification of the presence of pathogenic strains in seed batches is highly important to prevent spread of the disease and outbreaks in new locations around the world. Timely diagnosis of host-specific Fo forms can therefore be of decisive influence and could prevent unnecessary efforts to suppress harmless fungal populations (Van Der Does *et al.*, 2008). However, the design of *forma specialis*-specific markers is difficult due to the polyphyletic nature of most ff. spp. (see Chapter 4).

Effector proteins facilitate host colonization

During colonization of its host, Fo secretes enzymes and small, apparently non-enzymatic proteins. Proteomics on the xylem sap of infected susceptible tomato plants revealed the presence of 14 small fungal proteins inside this xylem sap, named "Secreted In Xylem" (Six) proteins (Houterman *et al.*, 2007; Schmidt *et al.*, 2013). These Six proteins are presumed to be "effectors" that facilitate the colonization process by suppression of, or protection against, the plant's immune system (Giraldo and Valent, 2013). For *SIX1, SIX3, SIX5* and *SIX6*, a contribution to virulence has been found in Fol (Rep *et al.*, 2004; Houterman *et al.*, 2009; L. Ma *et al.*, 2013; Gawehns *et al.*, 2014; Ma *et al.*, 2015). Some small secreted proteins act as avirulence factors because they are recognized by plant resistance (R) proteins (Rep *et al.*, 2004; Stergiopoulos and de Wit, 2009).
Studies on effector proteins in Fo have so far mainly focused on the *Fusarium*-tomato pathosystem. I set out to generate genome sequences of strains belonging to other *formae speciales*, which form the principle data component in this thesis. I particularly focussed on a group of *formae speciales* affecting the *CuCurbitaceae* family of crop plants. Several *formae speciales* causing *Fusarium* wilt in cucurbits have been described: f. sp. *cucumerinum* (on cucumber), f. sp. *melonis* (on melon), f. sp. *niveum* (on watermelon), f. sp. *lagenariae* (on bottle gourd), f. sp. *momordicae* (on bitter gourd) and f. sp. *luffae* (on *Luffa* spp.) (Leach *et al.*, 1937; Owen, 1956; Matuo and Yamamoto, 1967; Armstrong and Armstrong, 1981; Sun

and Huang, 1982; Kim *et al.*, 1993; Namiki *et al.*, 1994). Finally, an additional *forma specialis* has been described: f. sp. *radicis-cucumerinum* (Vakalounakis, 1996). This f. sp. is different from the others because it causes root and shoot rot and it is capable of doing this in multiple species of the *Curbitaceae* family, including cucumber, melon, watermelon and several gourds (Vakalounakis, 1996; Punja and Parker, 2000; Cohen *et al.*, 2015). I chose these ff. spp. because I wished to know whether strains that infect related plant species (cucurbits) possess similar effector gene sets.
By comparing the genome sequences generated from several isolates belonging to these ff. spp. with each other and to those already available from other studies, I hoped to identify what genetic factors allow a strain to cause disease in one species, but not in another. My hypothesis was that the combination of effector genes of a strain together determine virulence – and sometimes avirulence – towards a given plant species.

Effector identification

Effectors are often small and lack recognizable domains. Methods of computationally recognizing effector genes include the selection of predicted proteins that meet some or all of the following requirements: small size, secreted, high number of cysteines, evidence of diversifying selection and *in planta* transcriptional induction (Sperschneider, Dodds, *et al.*, 2015; Sperschneider, Gardiner, *et al.*, 2015).
Besides these criteria, effector genes in Fo have been found to be contextually associated with certain types of transposable elements (TEs; Fig. 2). Most notably, a *miniature impala (mimp)* sized ~220 nucleotides was found in all cases in the upstream region of Fol *SIX* genes (Schmidt *et al.*, 2013). *Mimps* are non-autonomous DNA transposons that can be mobilized through the action of the Impala transposase (Dufresne *et al.*, 2007). Bergemann *et al.* (2008) showed that even novel (*in vitro* created) *mimp* elements are recognized by this transposase, based on their 27 nucleotide long terminal inverted repeats (TIRs). Deletion of *mimp* elements in the promoter of two *SIX* genes did not result in highly altered gene expression, nor did it have an effect on (a)virulence of Fol (Schmidt *et al.*, 2013).
The presence of *mimp* elements in the genome of *F. oxysporum* has proven to be a useful tool to predict putative effector genes without having to resort to genome annotation in many different genomes (Schmidt *et al.*, 2013, 2016; van Dam *et al.*, 2016). In Schmidt *et al.* (2016), this approach was used on a set of Fo f. sp. *melonis* strains (causing *Fusarium* wilt in melon plants), and eight novel candidate effectors were found. One of the predicted effector genes was *AVRFOM2*, the gene that encodes the avirulence protein recognized by the melon Fom-2 resistance protein. The identification of *AVRFOM2* can help to select melon cultivars that are resistant or tolerant to avoid melon wilt caused by Fo.
This method is further explored in Chapter 2, where TE-based effector gene prediction is done for genomes of *F. oxysporum* belonging to a variety of *formae speciales*.

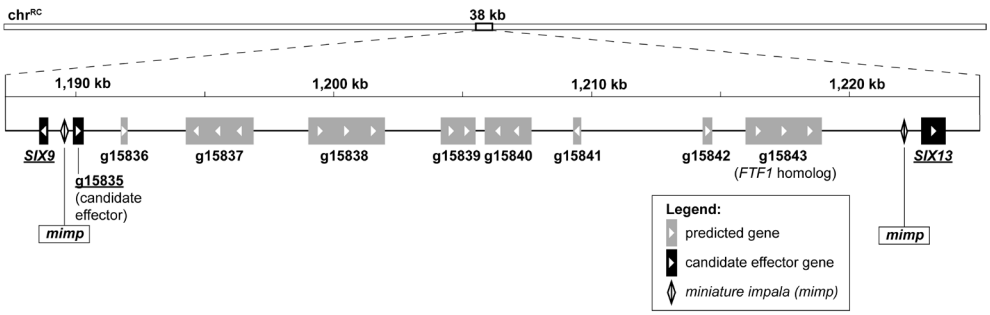


Fig. 2: A region on *F. oxysporum* f. sp. *radicis-cucumerinum*'s chr^{RC} contains two *mimps* that are positioned upstream of the homologs of *SIX9* and *SIX13* as well as a third predicted effector gene (marked in black). A *Fusarium Transcription Factor 1* (*FTF1*) homolog is also found in this region. This transcription factor is associated with effector gene induction during *in planta* growth (Niño-Sánchez *et al.*, 2016; van der Does *et al.*, 2016).

The compartmentalized genome of *Fusarium oxysporum*

F. oxysporum genomes are structurally compartmentalized, as is the case in many plant-pathogenic fungi (Raffaele and Kamoun, 2012). The relatively slow-evolving "core" genome handles the housekeeping of the organism and is highly syntenic between different Fo isolates as well as between Fo and related species, such as *F. verticillioides*. The second, "accessory" part consists of separate chromosomes or extensions of core chromosomes (Ma *et al.*, 2010). Accessory chromosomes show non-Mendelian inheritance and are present in some, but not all, individuals in a population (Möller and Stukenbrock, 2017). The accessory genome of Fo is not required for vegetative growth, but does contain most, if not all, of the genes that allow the fungus to invade and cause disease in its host plant. Gene density is lower on accessory chromosomes, while repeats and transposons are found much more abundantly here (Ma *et al.*, 2010). The sequences on these chromosomes are not conserved in all Fo strains. Finally, genes on these chromosomes generally have a higher non-synonymous mutation rate (d_n/d_s ratio) than genes on the core genome, suggesting relaxed selection (Sperschneider, Gardiner, *et al.*, 2015). It is thought that this genome partitioning into a stable and variable part allows the pathogen to adapt more quickly to changes with respect to the interaction with its host. Raffaele and Kamoun coined the term "two-speed genome" to describe two rates of evolution within the same genome (Raffaele and Kamoun, 2012).

The reference genome of *F. oxysporum*, that of Fo f. sp. *lycopersici* (Fol) 4287 has previously been sequenced and assembled through a combination of Sanger sequencing and optical mapping, resulting in a chromosome-level assembly. Analysis of the genome revealed that the strain possesses 15 chromosomes and that four entire chromosomes and part of chromosomes 1 and 2 could be considered accessory due to their high TE content and lack of synteny with *F. verticillioides* (Ma *et al.*, 2010). In the case of Fol, all of the 14 previously described *SIX* genes are located on chromosome 14. Since this chromosome was shown to be horizontally transferrable from strain to strain, thereby conferring the capability of infecting tomato

plants to the recipient strain, it is often referred to as the Fol "pathogenicity" chromosome. The function of the other accessory chromosomes not known. Potentially they act as an "evolutionary playground", providing a storehouse for large numbers of TEs and duplicated genes (Davière *et al.*, 2001; Vanheule *et al.*, 2016). At least 24% of the sequence of Fol4287 chromosome 14 is taken up by TEs or TE remnants (Schmidt *et al.*, 2013) and the other accessory regions have a similar density of transposons.

Two-speed genomes have been described in many plant-pathogenic fungal species, including *Leptosphaeria maculans, Zymoseptoria tritici* and *Fusarium solani* (teleomorph: *Nectria haematococca*) (Coleman *et al.*, 2009; Rouxel *et al.*, 2011; Schotanus *et al.*, 2015). Interestingly, trimethylation of lysine 27 of histone 3 (H3K27me3) and trimethylation of histone 3 lysine 9 (H3K9me3) seem to be important for regulation of gene expression in accessory regions. H3K9me3 is important for repression of virulence associated genes in *L. maculans* (Soyer *et al.*, 2014). Presence of these histone modifications is typically associated with gene silencing and has been used to distinguish core and accessory chromosomes in *Zymoseptoria tritici* (Schotanus *et al.*, 2015).

Histone 3 methylation has also been described in *F. oxysporum* (Connolly *et al.*, 2013). Mapping of reads derived from ChIP-seq experiments to the Fol4287 reference genome showed that there seems to be a third "speed", encompassing chromosomes with relatively high H3K27me3 and infection-related gene expression levels but that are conserved in most strains of *F. oxysporum* (Like Fokkens and Shermineh Shahi, unpublished data). This "soft-core" genome compartment consists of chromosomes 11, 12 and 13 in Fol4287. Chromosome 12 is not essential for vegetative or invasive growth, as this chromosome can be lost from Fol4287 without apparent effect on either carbon source usage or pathogenicity on tomato (Vlaardingerbroek, Beerens, Schmidt, *et al.*, 2016). Additionally, a strain of *F. oxysporum* f. sp. *cucumerinum* (strain Focuc037) was identified in our lab that lacks any sequences similar to Fol chromosome 12 (van Dam *et al.*, 2016).

Horizontal transfer and loss of chromosomes

No sexual structures such as perithecia, asci or ascospores have been described under natural or controlled conditions in Fo. Interestingly, strains do belong to either of the two mating type idiomorphs (MAT1-1-1 and MAT1-1-2), which may be the traces from a sexual history (Yun *et al.*, 2000). Even though it is possible that the presumed asexuality of Fo is due to a lack of environmental cues needed to trigger initiation of a sexual cycle, sex is thought to play a small role in the evolution and diversification of the species (Koenig *et al.*, 1997; Bentley *et al.*, 1998; Gordon, 2017). Asexual (clonal) reproduction can be advantageous in certain situations: co-adapted gene combinations are maintained in the population and fit genotypes can propagate more rapidly (Möller and Stukenbrock, 2017).

Today's agricultural practices provide a powerful selection environment for virulent

genotypes. When passed through this evolutionary filter of virulence to a particular host, *formae speciales* of *F. oxysporum* typically reveal very limited genetic diversity (Gordon, 2017), particularly in genes defining virulence (Rocha *et al.*, 2015). Horizontal transfer of chromosomes within the species complex and perhaps across species boundaries is believed to have contributed to genetic diversity and the generation of new (pathogenic) variants (Ma *et al.*, 2013; Kang *et al.*, 2014). It is still not clear how this process operates, especially since hyphal fusion of genetically dissimilar strains generally does not lead to viable heterokaryons (Glass and Dementhon, 2006). However, vegetative incompatibility responses are suppressed during conidial anastomosis tube (CAT) fusion in *Colletotrichum lindemuthianum* (Ishikawa *et al.*, 2012). The fusion of nuclei and he resulting formation of heterokaryotic structures possessing genetic information from two fused Fo strains is therefore hypothesized to happen during conidial anastomosis (Vlaardingerbroek, Beerens, Rose, *et al.*, 2016). CAT fusion has been described to happen frequently between genetically identical fungal spores, but also occurs between genetically distinct members of the same species and is even possible between different species (Roca *et al.*, 2004; Mehrabi *et al.*, 2011).

The phenomenon of horizontal chromosome transfer (HCT) was experimentally shown for the first time by Ma *et al.* (2010), where chromosome 14 and the smallest chromosome of Fol007 were horizontally transferred between strains belonging to different vegetative compatibility groups (VCGs). HCT was later also shown for Fol core chromosomes 7 and 8 (Ma *et al.*, 2010; Vlaardingerbroek, Beerens, Rose, *et al.*, 2016). Since transfer of core chromosomes led to the loss of the homologous region of the recipient genome, the term chromosome exchange was used for this particular process (Vlaardingerbroek, Beerens, Rose, *et al.*, 2016). Interestingly, exchange of core chromosomes was always accompanied by transfer of chromosome 14. This is an indication that one of each homologous chromosome pair in fused nuclei might be selectively degraded and that chromosome 14, that has no counterpart in the recipient's genome (in this case Fo47), is left intact.

Chromosomes and parts thereof can also be lost from the genome. This has been described for chromosomes 14 and 12 of Fol4287 (Vlaardingerbroek, Beerens, Schmidt, *et al.*, 2016). Fol strains that had been selected for loss of chromosome 14 were no longer pathogenic towards tomato plants. Surprisingly, strains with a large deletion of chromosome 14, including several candidate effector genes such as *SIX6*, *SIX9*, *SIX11* and *ORX1*, remained pathogenic on tomato (Vlaardingerbroek, Beerens, Schmidt, *et al.*, 2016).

In Fo, individual effector gene deletions usually result in only a modest loss of virulence, implying that a number of effectors contributes to host plant colonization and/or disease development. The examples presented above clearly illustrate the power of HCT and (partial) chromosome loss as an experimental approach to study the genetic factors underlying pathogenicity and host specificity of *F. oxysporum*. In this thesis, I present data showing for the first time horizontal transfer as well as loss of a pathogenicity chromosome in a different *forma specialis* than *lycopersici*.

Thesis outline

In **Chapter 2**, an effector prediction pipeline is applied to identify 104 putative effector genes in 59 (45 newly sequenced) *F. oxysporum* genomes. By looking at which effector genes are present or absent, and by analyzing sequence alignments of these genes, we were able to conclude that *F. oxysporum* strains infecting the same host generally possess a similar set of effectors and that these effectors are often identical within a *forma specialis*. In other words, effectors determine – or at least predict – a strain's host range.

Chapter 4 zooms in on the genome of Forc016, an isolate of Fo f. sp. *radicis-cucumerinum*. This isolate was sequenced with PacBio long read sequencing technology, which allowed most chromosomes to be assembled into single contigs. The assembly contained two contigs that together make up the Forc pathogenicity chromosome, which is functionally equivalent to Fol chromosome 14. Through horizontal chromosome transfer to a non-pathogenic strain (Fo47), as well as selective loss from Forc016 itself, we were able to show that this is indeed the chromosome responsible for invasive growth and root rot disease in cucurbits. Interestingly, we find that candidate effector genes and transposons are concentrated in a central region of the pathogenicity chromosome, and that Forc-*SIX6* is necessary for full virulence on cucumber. Most of the Forc pathogenicity chromosome is syntenic, with close to 100% sequence identity, to a chromosome in a melon-infecting strain. However, the central, effector-enriched regions of the two chromosomes seem to have been subject to a high number of recombination, deletion and/or translocation events.

Chapter 4 takes the knowledge obtained in **Chapter 2** a step further, seeking to apply effector genes as markers that allow molecular discrimination of Fo host range using PCR and quantitative PCR based techniques. Since effectors are causally linked to pathogenicity and host range, they form the best target for marker development. We aimed to find a marker set for each of the *formae speciales* that together affect the *Cucurbitaceae* plant family and tested their applicability on a set of isolates from around the world.

As described in **Chapters 2** and **4**, certain transposable elements have predictive value for locating effector genes in *F. oxysporum* genomes. Thus far, these specific elements (*miniature impalas,* or *mimps*) have only been described within the *F. oxysporum* species complex and therefore seem to be restricted to the FOSC. In **Chapter 5**, a number of strains initially thought to be *F. oxysporum* isolated from diseased flower bulbs is analyzed. Although several of these turned out to be other *Fusarium* species than Fo, they did possess *mimps* as well as *SIX* gene homologs in their genome assemblies. In this chapter, we evaluate the distribution of different *mimp* families in the *Fusarium* genus and explore the possibility of inter-species horizontal chromosome transfer events.

Finally, the results described in this thesis and their implications are discussed in **Chapter 6**.

References

Alabouvette, C., Olivain, C., Migheli, Q., and Steinberg, C. (2009) Microbiological control of soil-borne phytopathogenic fungi with special emphasis on wilt-inducing *Fusarium oxysporum*. *New Phytol.* **184**: 529–544.

Alabouvette, C., Rouxel, F., and Louvet, J. (1979) Characteristics of *Fusarium* wilt-suppressive soils and prospects for their utilization in biological control. *Soil-borne Plant Pathog. Acad. Press. New-York* 165–182.

Armstrong, G.M. and Armstrong, J.K. (1981) *Formae speciales* and races of *Fusarium oxysporum* causing wilt diseases. In *Fusarium: diseases, biology, and taxonomy*. Pennsylvania State University Press, University Park, PA, pp. 391–399.

Armstrong, G.M. and Armstrong, J.K. (1978) *Formae speciales* and Races of *Fusarium oxysporum* Causing Wilts of the *CuCurbitaceae*. *Phytopathology* **68**: 19.

Baayen, R.P., O'Donnell, K., Bonants, P.J., Cigelnik, E., Kroon, L.P., Roebroeck, E.J., and Waalwijk, C. (2000) Gene Genealogies and AFLP Analyses in the *Fusarium oxysporum* Complex Identify Monophyletic and Nonmonophyletic *Formae speciales* Causing Wilt and Rot Disease. *Phytopathology* **90**: 891–900.

Bennett, R.S., Hutmacher, R.B., and Davis, R.M. (2008) Seed transmission of *Fusarium oxysporum* f. sp. *vasinfectum* race 4 in California. *J. Cotton Sci.*

Bentley, S., Pegg, K.G., Moore, N.Y., Davis, R.D., and Buddenhagen, I.W. (1998) Genetic Variation Among Vegetative Compatibility Groups of *Fusarium oxysporum* f. sp. *cubense* Analyzed by DNA Fingerprinting. *Phytopathology* **88**: 1283–1293.

Bergemann, M., Lespinet, O., M'Barek, S. Ben, Daboussi, M.J., and Dufresne, M. (2008) Genome-wide analysis of the *Fusarium oxysporum mimp* family of MITEs and mobilization of both native and *de novo* created *mimps*. *J. Mol. Evol.* **67**: 631–642.

Chiocchetti, A., Ghignone, S., Minuto, A., Gullino, M.L., Garibaldi, A., and Migheli, Q. (1999) Identification of *Fusarium oxysporum* f. sp. *basilici* Isolated from Soil, Basil Seed, and Plants by RAPD Analysis. *Plant Dis.* **83**: 576–581.

Cohen, R., Orgil, G., Burger, Y., Saar, U., Elkabetz, M., Tadmor, Y., *et al.* (2015) Differences in the responses of melon accessions to *Fusarium* root and stem rot and their colonization by *Fusarium oxysporum* f. sp. *radicis-cucumerinum*. *Plant Pathol.* **64**: 655–663.

Cohen, R., Tyutyunik, J., Fallik, E., Oka, Y., Tadmor, Y., and Edelstein, M. (2014) Phytopathological evaluation of exotic watermelon germplasm as a basis for rootstock breeding. *Sci. Hortic.* **165**: 203–210.

Coleman, J.J., Rounsley, S.D., Rodriguez-Carres, M., Kuo, A., Wasmann, C.C., Grimwood, J., *et al.* (2009) The genome of *Nectria haematococca*: Contribution of supernumerary chromosomes to gene expansion. *PLoS Genet.* **5**:.

Connolly, L.R., Smith, K.M., and Freitag, M. (2013) The *Fusarium graminearum* Histone H3 K27 Methyltransferase *KMT6* Regulates Development and Expression of Secondary Metabolite Gene Clusters. *PLoS Genet.* **9**: e1003916.

Correll, J.C., Puhalla, J.E., and Schneider, R.W. (1986) Vegetative compatibility groups among nonpathogenic root-colonizing strains of *Fusarium oxysporum*. *Can. J. Bot.* **64**: 2358–2361.

Daayf, F., El Bellaj, M., El Hassni, M., J'Aiti, F., and El Hadrami, I. (2003) Elicitation of soluble phenolics in date palm *(Phoenix dactylifera)* callus by *Fusarium oxysporum* f. sp. *albedinis* culture medium. *Environ. Exp. Bot.* **49**: 41–47.

van Dam, P., Fokkens, L., Schmidt, S.M., Linmans, J.H.J., Kistler, H.C., Ma, L.J., and Rep, M. (2016) Effector profiles distinguish *formae speciales* of *Fusarium oxysporum*. *Environ. Microbiol.* **18**: 4087–4102.

Davière, J.M., Langin, T., and Daboussi, M.J. (2001) Potential role of transposable elements in the rapid reorganization of the *Fusarium oxysporum* genome. *Fungal Genet. Biol.* **34**: 177–192.

van der Does, H.C., Fokkens, L., Yang, A., Schmidt, S.M., Langereis, L., Lukasiewicz, J.M., *et al.* (2016) Transcription Factors Encoded on Core and Accessory Chromosomes of *Fusarium oxysporum* Induce Expression of Effector Genes. *PLOS Genet.* **12**: e1006401.

Van Der Does, H.C., Lievens, B., Claes, L., Houterman, P.M., Cornelissen, B.J.C., and Rep, M. (2008) The presence of a virulence locus discriminates *Fusarium oxysporum* isolates causing tomato wilt from other isolates. *Environ. Microbiol.* **10**: 1475–1485.

Dubey, S.C., Suresh, M., and Singh, B. (2007) Evaluation of *Trichoderma* species against *Fusarium oxysporum* f. sp. *ciceris* for integrated management of chickpea wilt. *Biol. Control* **40**: 118–127.

Dufresne, M., Hua-Van, A., Abd El Wahab, H., M'Barek, S. Ben, Vasnier, C., Teysset, L., *et al.* (2007) Transposition of a fungal miniature inverted-repeat transposable element through the action of a *Tc1*-like transposase. *Genetics*

175: 441–452.
Fulton, H.R. and Winston, J.R. (1913) Watermelon wilt spread by contaminated seed. *NC Agr. Expt. Sta. Bien. Rpt* **14**: 48–51.
Gao, H., Beckman, C.H., and Mueller, W.C. (1995) The nature of tolerance to *Fusarium oxysporum* f. sp. *lycopersici* in polygenically field-resistant marglobe tomato plants. *Physiol. Mol. Plant Pathol.* **46**: 401–412.
Garibaldi, A., Gilardi, G., and Gullino, M.L. (2004) Seed transmission of *Fusarium oxysporum* f. sp. *lactucae*. *Phytoparasitica* **32**: 61–65.
Gawehns, F., Houterman, P.M., Ichou, F.A., Michielse, C.B., Hijdra, M., Cornelissen, B.J.C., *et al.* (2014) The *Fusarium oxysporum* Effector *SIX6* Contributes to Virulence and Suppresses I-2-Mediated Cell Death. *Mol. Plant-Microbe Interact.* **27**: 336–348.
Giraldo, M.C. and Valent, B. (2013) Filamentous plant pathogen effectors in action. *Nat. Rev. Microbiol.* **11**: 800–14.
Glass, N.L. and Dementhon, K. (2006) Non-self recognition and programmed cell death in filamentous fungi. *Curr. Opin. Microbiol.* **9**: 553–558.
Gordon, T.R. (2017) *Fusarium oxysporum* and the *Fusarium* wilt Syndrome. *Annu. Rev. Phytopathol.* **55**: 1–17.
Houterman, P.M., Ma, L., Van Ooijen, G., De Vroomen, M.J., Cornelissen, B.J.C., Takken, F.L.W., and Rep, M. (2009) The effector protein Avr2 of the xylem-colonizing fungus *Fusarium oxysporum* activates the tomato resistance protein I-2 intracellularly. *Plant J.* **58**: 970–978.
Houterman, P.M., Speijer, D., Dekker, H.L., De Koster, C.G., Cornelissen, B.J.C., and Rep, M. (2007) The mixed xylem sap proteome of *Fusarium oxysporum*-infected tomato plants. *Mol. Plant Pathol.* **8**: 215–221.
Inami, K., Yoshioka-Akiyama, C., Morita, Y., Yamasaki, M., Teraoka, T., and Arie, T. (2012) A genetic mechanism for emergence of races in *Fusarium oxysporum* f. sp. *lycopersici*: Inactivation of avirulence gene *AVR1* by transposon insertion. *PLoS One* **7**: e44101.
Ishikawa, F.H., Souza, E.A., Shoji, J.Y., Connolly, L., Freitag, M., Read, N.D., and Roca, M.G. (2012) Heterokaryon incompatibility is suppressed following conidial anastomosis tube fusion in a fungal plant pathogen. *PLoS One* **7**:.
Jiménez-Fernández, D., Landa, B.B., Kang, S., Jiménez-Díaz, R.M., and Navas-Cortés, J.A. (2013) Quantitative and Microscopic Assessment of Compatible and Incompatible Interactions between Chickpea Cultivars and *Fusarium oxysporum* f. sp. *ciceris* Races. *PLoS One* **8**:.
Kang, S., Demers, J., del Mar Jimenez-Gasco, M., and Rep, M. (2014) *Fusarium oxysporum*. In, *Genomics of Plant-Associated Fungi and Oomycetes: Dicot Pathogens*. Springer, pp. 99–119.
Kim, D.H., Martyn, R.D., and Magill, C.W. (1993) Mitochondrial DNA (mtDNA)-Relatedness among *formae speciales* of *Fusarium oxysporum* in the *CuCurbitaceae*. *Phytopathology* **83**: 91–97.
Kistler, H.C. (1997) Genetic Diversity in the Plant-Pathogenic Fungus *Fusarium oxysporum*. *Phytopathology* **87**: 474–479.
Koenig, R.L., Ploetz, R.C., and Kistler, H.C. (1997) *Fusarium oxysporum* f. sp. *cubense* Consists of a Small Number of Divergent and Globally Distributed Clonal Lineages. *Phytopathology* **87**: 915–923.
Kuniyasu, K. (1980) Seed transmission of *Fusarium* wilt of bottle gourd, *Lagenaria siceraria*, used as rootstock of watermelon. *JARQ* **14**: 157–162.
Laurence, M.H., Summerell, B.A., Burgess, L.W., and Liew, E.C.Y. (2014) Genealogical concordance phylogenetic species recognition in the *Fusarium oxysporum* species complex. *Fungal Biol.* **118**: 374–384.
Leach, J.G., Currence, T.M., and others (1937) *Fusarium* wilt of muskmelons in Minnesota. *Univ. Minnesota, Agric. Exp. Stn. Bull.* **129**:.
Lievens, B., Brouwer, M., Vanachter, A.C.R.C., Cammue, B.P.A., and Thomma, B.P.H.J. (2006) Real-time PCR for detection and quantification of fungal and oomycete tomato pathogens in plant and soil samples. *Plant Sci.* **171**: 155–165.
Lievens, B., Houterman, P.M., and Rep, M. (2009) Effector gene screening allows unambiguous identification of *Fusarium oxysporum* f. sp. *lycopersici* races and discrimination from other *formae speciales*. *FEMS Microbiol. Lett.* **300**: 201–215.
Ma, L.-J., van der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.-J., Di Pietro, A., *et al.* (2010) Comparative analysis reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367–373.
Ma, L.-J., Geiser, D.M., Proctor, R.H., Rooney, A.P., O'Donnell, K., Trail, F., *et al.* (2013) *Fusarium* pathogenomics. *Annu. Rev. Microbiol.* **67**: 399–416.
Ma, L.-J., Cornelissen, B.J.C., and Takken, F.L.W. (2013) A nuclear localization for Avr2 from *Fusarium oxysporum* is required to activate the tomato resistance protein I-2. *Front. Plant Sci.* **4**: 94.
Ma, L., Houterman, P.M., Gawehns, F., Cao, L., Sillo, F., Richter, H., *et al.* (2015) The *AVR2-SIX5* gene pair is required

to activate *I-2*-mediated immunity in tomato. *New Phytol.* **208**: 507–518.
Del Mar Jiménez-Gasco, M. and Jiménez-Díaz, R.M. (2003) Development of a Specific Polymerase Chain Reaction-Based Assay for the Identification of *Fusarium oxysporum* f. sp. *ciceris* and Its Pathogenic Races 0, 1A, 5, and 6. *Phytopathology* **93**: 200–209.
Martyn, R.D. (2014) *Fusarium* Wilt of Watermelon: 120 Years of Research. *Hortic. Rev. Vol. 42* **42**: 349–442.
Martyn, R.D. and Vakalounakis, D.J. (2012) *Fusarium* wilt of greenhouse cucurbits: melon, watermelon cucumber. *Fusarium wilts Greenh. Veg. Ornam. Crop. APS Press. St. Paul, MN*.
Matuo, T. and Yamamoto, I. (1967) On *Fusarium oxysporum* f. sp. *lagenariae* n.f. causing wilt of *Lagenaria vulgaris* var. hispida. *Trans. Mycol. Soc. Japan* **8**: 61–63.
Mehrabi, R., Bahkali, A.H., Abd-Elsalam, K.A., Moslem, M., Ben M'Barek, S., Gohari, A.M., *et al.* (2011) Horizontal gene and chromosome transfer in plant pathogenic fungi affecting host range. *FEMS Microbiol. Rev.* **35**: 542–554.
Michielse, C.B. and Rep, M. (2009) Pathogen profile update: *Fusarium oxysporum*. *Mol. Plant Pathol.* **10**: 311–324.
Möller, M. and Stukenbrock, E.H. (2017) Evolution and genome architecture in fungal plant pathogens. *Nat. Rev. Microbiol.* doi: 10.1038/nrmicro.2017.76.
Namiki, F., Shiomi, T., Kayamura, T., and Tsuge, T. (1994) Characterization of the *formae speciales* of *Fusarium oxysporum* causing wilts of cucurbits by DNA fingerprinting with nuclear repetitive DNA sequences. *Appl. Environ. Microbiol.* **60**: 2684–2691.
Nelson, P.E. (2012) Life cycle and epidemiology of *Fusarium oxysporum*. In *Fungal wilt diseases of plants*. New York: Academic, pp. 51–80.
Niño-Sánchez, J., Casado-Del Castillo, V., Tello, V., de Vega-Bartol, J.J., Ramos, B., Sukno, S.A., *et al.* (2016) The *FTF* gene family regulates virulence and expression of *SIX* effectors in *Fusarium oxysporum*. *Mol. Plant Pathol.* 1–18.
Ordonez, N., Seidl, M.F., Waalwijk, C., Drenth, A., Kilian, A., Thomma, B.P.H.J., *et al.* (2015) Worse Comes to Worst: Bananas and Panama Disease—When Plant and Pathogen Clones Meet. *PLOS Pathog.* **11**: e1005197.
Owen, J.H. (1956) Cucumber wilt, caused by *Fusarium oxysporum* f. *cucumerinum*. *Phytopathology* **46**: 153–157.
Pande, S., Narayana Rao, J., and Sharma, M. (2007) Establishment of the Chickpea Wilt Pathogen *Fusarium oxysporum* f. sp. *ciceris* in the Soil through Seed Transmission. *Plant Pathol. J.* **23**: 3–6.
Pavlou, G.C., Vakalounakis, D.J., and Ligoxigakis, E.K. (2002) Control of Root and Stem Rot of Cucumber, Caused by *Fusarium oxysporum* f. sp. *radicis-cucumerinum*, by Grafting onto Resistant Rootstocks. *Plant Dis.* **86**: 379–382.
Porter, D.R. (1928) Infection studies with watermelon wilt caused by *Fusarium niveum* EFS. *Iowa Agric. Exp. Sta. Res. Bull.* **112**: 347–368.
Punja, Z.K. and Parker, M. (2000) Development of *Fusarium* root and stem rot, a new disease on greenhouse cucumber in British Columbia, caused by *Fusarium oxysporum* f. sp. *radicis-cucumerinum*. *Can. J. Plant Pathol.* **22**: 349–363.
Raffaele, S. and Kamoun, S. (2012) Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Microbiol.*
Rep, M., Dekker, H.L., Vossen, J.H., de Boer, A.D., Houterman, P.M., Speijer, D., *et al.* (2002) Mass spectrometric identification of isoforms of PR proteins in xylem sap of fungus-infected tomato. *Plant Physiol.* **130**: 904–917.
Rep, M., Van Der Does, H.C., Meijer, M., Van Wijk, R., Houterman, P.M., Dekker, H.L., *et al.* (2004) 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.
Reuveni, R., Raviv, M., Krasnovsky, A., Freiman, L., Medina, S., Bar, A., and Orion, D. (2002) Compost induces protection against *Fusarium oxysporum* in sweet basil. *Crop Prot.* **21**: 583–587.
Roca, M.G., Davide, L.C., Davide, L.M.C., Mendes-Costa, M.C., Schwan, R.F., and Wheals, A.E. (2004) Conidial anastomosis fusion between *Colletotrichum* species. *Mycol. Res.* **108**: 1320–1326.
Rocha, L.O., Laurence, M.H., Ludowici, V.A., Puno, V.I., Lim, C.C., Tesoriero, L.A., *et al.* (2015) Putative effector genes detected in *Fusarium oxysporum* from natural ecosystems of Australia. *Plant Pathol.* **65**: 914–929
Rose, S., Parker, M., and Punja, Z.K. (2003) Efficacy of Biological and Chemical Treatments for Control of *Fusarium* Root and Stem Rot on Greenhouse Cucumber. *Plant Dis.* **87**: 1462–1470.
Rouxel, T., Grandaubert, J., Hane, J.K., Hoede, C., van de Wouw, A.P., Couloux, A., *et al.* (2011) Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nat. Commun.* **2**: 202.
Schmidt, S.M., Houterman, P.M., Schreiver, I., Ma, L., Amyotte, S., Chellappan, B., *et al.* (2013) MITEs in the

promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genomics* **14**: 119.

Schmidt, S.M., Lukasiewicz, J., Farrer, R., van Dam, P., Bertoldo, C., and Rep, M. (2016) Comparative genomics of *Fusarium oxysporum* f. sp. *melonis* reveals the secreted protein recognized by the *Fom-2* resistance gene in melon. *New Phytol.* **209**: 307–318.

Schotanus, K., Soyer, J.L., Connolly, L.R., Grandaubert, J., Happel, P., Smith, K.M., *et al.* (2015) Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. *Epigenetics Chromatin* **8**: 41.

Soyer, J.L., El Ghalid, M., Glaser, N., Ollivier, B., Linglin, J., Grandaubert, J., *et al.* (2014) Epigenetic Control of Effector Gene Expression in the Plant Pathogenic Fungus *Leptosphaeria maculans*. *PLoS Genet.* **10**:.

Sperschneider, J., Dodds, P.N., Gardiner, D.M., Manners, J.M., Singh, K.B., and Taylor, J.M. (2015) Advances and Challenges in Computational Prediction of Effectors from Plant Pathogenic Fungi. *PLOS Pathog.* **11**: e1004806.

Sperschneider, J., Gardiner, D.M., Thatcher, L.F., Lyons, R., Singh, K.B., Manners, J.M., and Taylor, J.M. (2015) Genome-Wide Analysis in Three *Fusarium* Pathogens Identifies Rapidly Evolving Chromosomes and Genes Associated with Pathogenicity. *Genome Biol. Evol.* **7**: 1613–1627.

Stergiopoulos, I. and de Wit, P.J.G.M. (2009) Fungal effector proteins. *Annu. Rev. Phytopathol.* **47**: 233–263.

Sun, S. and Huang, J. (1982) A new *Fusarium* wilt of bitter gourd in Taiwan. *Plant Dis.* **67**: 226–227.

Vakalounakis, D.J. (1996) Root and Stem Rot of Cucumber Caused by *Fusarium oxysporum* f. sp. *radicis-cucumerinum* f. sp. nov. *Plant Dis.* **80**: 313–316.

Vanheule, A., Audenaert, K., Warris, S., van de Geest, H., Schijlen, E., Höfte, M., *et al.* (2016) Living apart together: crosstalk between the core and supernumerary genomes in a fungal plant pathogen. *BMC Genomics* **17**: 670.

Vlaardingerbroek, I., Beerens, B., Rose, L., Fokkens, L., Cornelissen, B.J.C., and Rep, M. (2016) Exchange of core chromosomes and horizontal transfer of lineage-specific chromosomes in *Fusarium oxysporum*. *Environ. Microbiol.***18:** 3702–3713.

Vlaardingerbroek, I., Beerens, B., Schmidt, S.M., Cornelissen, B.J.C., and Rep, M. (2016) Dispensable chromosomes in *Fusarium oxysporum* f. sp *lycopersici*. *Mol. Plant Pathol.* **17:** 1455-1466

Yogev, A., Raviv, M., Hadar, Y., Cohen, R., and Katan, J. (2006) Plant waste-based composts suppressive to diseases caused by pathogenic *Fusarium oxysporum*. *Eur. J. Plant Pathol.* **116**: 267–278.

Yun, S.-H., Arie, T., Kaneko, I., Yoder, O.C., and Turgeon, B.G. (2000) Molecular Organization of Mating Type Loci in Heterothallic, Homothallic, and Asexual *Gibberella/Fusarium* Species. *Fungal Genet. Biol.* **31**: 7–20.

Chapter 2

Effector profiles distinguish *formae speciales* of *Fusarium oxysporum*

This chapter has been published as:

Peter van Dam, Like Fokkens, Sarah M. Schmidt, Jasper H.J. Linmans, H. Corby Kistler, Li-Jun Ma, and Martijn Rep (2016) Effector profiles distinguish *formae speciales* of *Fusarium oxysporum*. *Environmental Microbiology* **18**: 4087–4102. doi:10.1111/1462-2920.13445

Effector profiles distinguish *formae speciales* of *Fusarium oxysporum*

Abstract

Formae speciales (ff. spp.) of the fungus *Fusarium oxysporum* are often polyphyletic within the species complex, making it impossible to identify them on the basis of conserved genes. However, sequences that determine host-specific pathogenicity may be expected to be similar between strains within the same *forma specialis*. Whole genome sequencing was performed on strains from five different ff. spp. (*cucumerinum*, *niveum*, *melonis*, *radicis-cucumerinum* and *lycopersici*). In each genome, genes for putative effectors were identified based on small size, secretion signal and vicinity to a 'miniature impala' transposable element. The candidate effector genes of all genomes were collected and the presence/absence patterns in each individual genome were clustered. Members of the same *forma specialis* turned out to group together, with cucurbit-infecting strains forming a supercluster separate from other ff. spp. Moreover, strains from different clonal lineages within the same *forma specialis* harbour identical effector gene sequences, supporting horizontal transfer of genetic material. These data offer new insight into the genetic basis of host specificity in the *F. oxysporum* species complex and show that (putative) effectors can be used to predict host specificity in *F. oxysporum*.

Introduction

The *Fusarium oxysporum* (Fo) species complex (FOSC) comprises an important group of filamentous fungi that includes plant-pathogenic strains. The species complex as a whole has a very wide host range, but individual pathogenic strains are restricted to one or a few host species. Accordingly, such strains are grouped into *formae speciales* (ff. spp.) based on host specificity (Armstrong and Armstrong, 1981; Baayen *et al.*, 2000). They cause vascular wilt (due to xylem colonization) or root, bulb or foot rot in over 120 plant species, including many economically important crops like tomato and cucurbits (Armstrong and Armstrong, 1981; Michielse and Rep, 2009).

Each *forma specialis* (f. sp.) of Fo consists of one or several clonal lineages (O'Donnell *et al.*, 1998; Katan, 1999). Strains belonging to different *formae speciales* may be more related than strains belonging to the same *forma specialis* (Kistler, 1997; Lievens *et al.*, 2009), which implies that the host of a strain of *F. oxysporum* can not reliably be determined based on conserved gene sequences, such as the translation elongation factor 1-alpha gene *(EF-1α)* alone. Currently, disease assays remain the primary method of discriminating host range and races (defined by the capacity to infect different cultivars of a plant species and often based on presence/absence of or point mutations in effector genes) of a pathogenic Fo strain (Recorbet *et al.*, 2003; Covey *et al.*, 2014; Martyn, 2014).

Because disease assays are laborious and time consuming, a molecular screening method based on *forma-specialis*-specific DNA sequences is highly desirable. However, knowledge of the genetic basis of host-specificity is limited.

For successful infection of their host, pathogens often rely on effector proteins – small secreted proteins that facilitate the colonization process (Stergiopoulos and de Wit, 2009; Dodds and Rathjen, 2010; Giraldo and Valent, 2013). Infection of different plant hosts likely requires a different set of effectors, which renders these proteins potentially informative for discriminating *formae speciales*. In this study we aimed to uncover which putative effector genes are shared by strains belonging to each of four economically important (Kim *et al.*, 1993b; Vakalounakis *et al.*, 2005) cucurbit-infecting *formae speciales*, notably *cucumerinum* (Foc; cucumber), *melonis* (Fom; musk melon), *niveum* (Fon; watermelon) and *radicis-cucumerinum* (Forc; cucumber and other cucurbits), as well as the tomato-infecting *F. oxysporum* f. sp. *lycopersici* (Fol). We chose these ff. spp. because we wished to know whether strains that infect related plant species *(Cucurbitaceae)* possess similar effector suites.

Of all of the *formae speciales* mentioned above, Fol has been investigated most extensively. A horizontally transferrable 'pathogenicity' chromosome (chromosome 14 of Fol4287) harbours all but one of the 14 known Fol effector genes, named *SIX* (for Secreted In Xylem) (Ma *et al.*, 2010; Schmidt *et al.*, 2013). These genes encode small, cysteine-rich, secreted proteins with no recognizable protein domain and at least some appear to be employed by Fol to manipulate the host's defense responses, thereby promoting the infection process (Rep *et al.*, 2004; Gawehns

et al., 2014; Ma *et al.*, 2015). Together with chromosomes 1B, 2B, 3, 6 and 15, chromosome 14 forms the accessory genome of Fol4287. The accessory genome lacks synteny with other *Fusarium* species such as *F. verticillioides,* is relatively gene poor and harbours many repeats and transposable elements (TEs), characteristics that differentiate it from the core genome (Ma *et al.*, 2010, 2013).

Analysis of the genomic context of *SIX1 – SIX7* revealed an association of these genes with two TEs. A miniature impala (*mimp*) was found in all cases in the upstream region and mFot5 was found frequently downstream of the Open Reading Frame (ORF) (Schmidt *et al.*, 2013). Both *mimps* and mFots are classes of Miniature Inverted-repeat TEs (MITEs). MITEs are short, non-autonomous DNA transposons, thought to be truncated derivatives of autonomous DNA transposons (Feschotte and Pritham, 2007; Lu *et al.*, 2012). They contain 27-30 nucleotide terminal inverted repeats (TIRs). In the case of *mimps*, the sequence of these TIRs is conserved between different *mimp*-subfamilies (Bergemann *et al.*, 2008). A total of 103 *mimps* are present in the genome of Fol4287 (Schmidt *et al.*, 2013). The highest density was found on chromosome 14, the 'pathogenicity chromosome', with 54 *mimps* (21 *mimps*/Mb), compared to 45 (3 *mimps*/Mb) on the other accessory chromosomes and 4 (0.1 *mimps*/Mb) on the core chromosomes. Although *mimp* deletion experiments did not result in altered *SIX* gene expression (Schmidt *et al.*, 2013), the consistent presence of a *mimp* in the promoter region of known Fo effector genes was successfully exploited to identify novel effector candidates (Schmidt *et al.*, 2016). Sixteen candidate effector genes were identified close to a *mimp*, and the products of fourteen of these (*SIX1 – SIX14*) were found in the xylem sap of infected tomato plants using mass spectrometry (Houterman *et al.*, 2007; Schmidt *et al.*, 2013).

We made use of the association between *mimps* and effector genes in Fo genomes to predict the suite of putative effectors present in 59 *F. oxysporum* genomes (45 new assemblies), without the need to rely on genome annotation. We then compared the predicted 'effectoromes' of these 59 different strains to determine whether strains that belong to the same f. sp. have similar effector repertoires and whether we could use presence/absence patterns of putative effector genes to predict the host range of a strain. We find that indeed strains cluster into *formae speciales* based on effector presence/absence profiles. Moreover, identification of ff. spp. based on effector genes is further strengthened by taking their sequences into account, since these are identical or highly similar within a f. sp., but often different between ff. spp.

Results

Strain selection

In order to make a well-founded evaluation of genome and effector variation within and between *formae speciales* of *F. oxysporum*, we selected divergent strains for our study. For

the *formae speciales* Foc, Fom and Fon, a polyphyletic nature has been described (Jacobson and Gordon, 1990; Kim *et al.*, 1992, 1993a; Lievens *et al.*, 2007). Forc has a broader host range and is able to cause root and stem rot in various cucurbit species such as muskmelon and sponge gourd (*Luffa aegyptica*) (Vakalounakis, 1996; Vakalounakis and Fragkiadakis, 1999; Vakalounakis *et al.*, 2005). According to earlier reports it comprises two vegetative compatibility groups (VCGs) but these might constitute a single clonal lineage (Lievens *et al.*, 2007). Fol is also polyphyletic, comprising at least four clonal lineages (Van Der Does *et al.*, 2008; Ma *et al.*, 2010).

Based on diversity in *EF-1α* sequence (indicative of clonal lineages, Supplementary data Fig. S1), geographical origin and disease assays to confirm host-specific pathogenicity, 45 strains (9 Foc, 9 Fom, 9 Fon, 3 Forc, 14 Fol, one non-pathogenic, Table 1) were selected for whole genome Illumina paired-end sequencing. Combined with 14 previously generated Fo assemblies belonging to various ff. spp. (Table 2), these genomes formed the basis for our search for putative effector genes within the FOSC.

Most cucurbit-infecting strains are host specific

Previous studies have indicated that Fo strains belonging to f. sp. *cucumerinum* (Cafri *et al.*, 2005) and *niveum* (Zhou and Everts, 2007) display mild cross-pathogenicity towards musk melon. To assess the level of (cross-)pathogenicity of the cucurbit-infecting strains in our collection, we conducted disease assays with all sequenced Foc, Forc, Fom and Fon strains, plus one Fol strain (Fol029) on susceptible cultivars of cucumber, musk melon, watermelon and tomato, as well as on a Foc-resistant cultivar of cucumber. The results are summarized in Fig. 1.

Fom and Fon strains were highly specific to their described host species, whereas several Foc strains showed some degree of cross-pathogenicity, especially towards musk melon. Disease symptoms on musk melon plants caused by *formae speciales* other than *melonis* (and *radicis-cucumerinum*) were generally not severe, being limited to growth retardation or light wilting symptoms. Within Foc, strains Foc018, 021, 030 (all belonging to VCG0180) and Foc011 (VCG0186) were pathogenic on susceptible, but not on resistant cucumber plants; the other Foc strains also caused symptoms in the resistant plants. Forc strains were strongly pathogenic towards cucumber and melon plants and moderately pathogenic towards Fon-susceptible watermelon plants (cv. Black Diamond) as well as Foc-resistant cucumber plants (cv. Melen). Only Fol029 caused disease in tomato and inoculation with this strain did not cause any symptoms in any of the tested cucurbit plants. This shows that although cross-pathogenicity of Foc and Forc towards related cucurbit plants is possible, this does not extend towards tomato, a member of the *Solanaceae*.

To evaluate the extent to which the different Fo strains are able to colonize the vasculature of the various plant lines and species, slices of surface-sterilized hypocotyls of infected plants

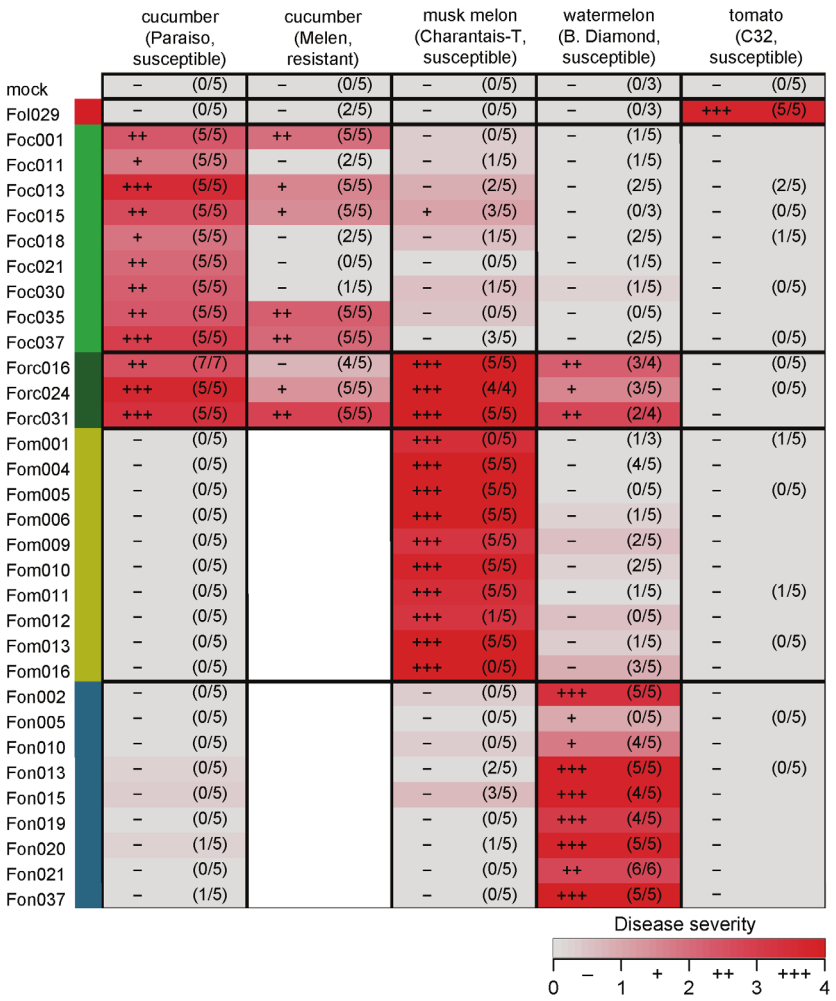


Fig. 1: Host-specificity of cucurbit-infecting strains used in this study.
Disease index (0-4) was scored in multiple bioassays under controlled conditions in the greenhouse after two weeks (cucurbits) or three weeks (tomato). All compatible interactions were tested at least twice in individual bioassays. Between brackets is the number of surface-sterilized hypocotyls from which *F. oxysporum* outgrowth was observed on CDA after 4 days.

were placed on Czapec Dox Agar (CDA) plates. Mycelial outgrowth after four days largely correlated with disease symptoms (Fig. 1). However, sometimes outgrowths were observed from symptomless plants, suggestive of an endophytic interaction.

Genome features

The genomes of the 45 selected strains were sequenced using paired-end Illumina libraries with different insert sizes, as described in the Experimental procedures. The genomes were

sequenced to 60-200X coverage (Table S1, trimmed and cleaned Illumina data), resulting in assemblies of 826 (Forc024) to 3867 (Foc018) scaffolds. The smallest *de novo* assembly has a cumulative size of 48.4 megabasepairs (Mbp), including ambiguous bases (Foc037); the largest *de novo* assembly adds up to 57.7 Mbp (Fom009). 97-98% of highly conserved protein-coding genes are estimated by CEGMA in all assemblies (Parra *et al.*, 2009), which is similar to other *F. oxysporum* reference genome assemblies retrieved from Genbank (Table S1). This highlights the completeness of the assemblies, particularly in core regions of the genomes. Repeat-density, RNA and DNA transposon abundance (Fig. S2) and GC content (Table S1) is largely similar between the assemblies.

The genome of the reference strain Fol4287 is the best-studied and assembled Fo genome to date, with a near-complete chromosome assembly through the use of Sanger sequencing combined with an optical map (Ma *et al.*, 2010). Fig. 2 shows the size and the level of fragmentation in each of the assessed assemblies. This analysis indicates that most strains contain a roughly equally sized core genome of about 41 Mb, with some exceptions that seem to have a slightly larger (e.g. Fom001, Fo f. sp. cubense (Focub) B2, Forl CL57, Fo47, FOSC-3a) or smaller (e.g. Fo5176, Fom010 and Foc037) cumulative core genome size. Differences in reported core genome size between strains may be caused by addition of accessory regions to the ends of core chromosomes, as was observed in chromosomes 1 and 2 of Fol4287 (Ma *et al.*, 2010) or by partial or entire loss of chromosomes. The size of the accessory regions varies considerably (between 4 and 19 Mb), but is usually comparable within a clonal lineage. To make an estimation of the complete genome size, we calculated the number of basepairs mapped against the assembly divided by the median coverage of contigs larger than 100kb. This yielded estimated genome sizes that were considerably larger than the assembly (Table S1), indicating that indeed large-scale duplication events are a relevant factor. This also explains Fol4287's assembly sticking out in size (Fig. 2) due to its relatively complete assembly, most notably in the aforementioned duplicated regions.

Foc, Fom, Fon and Fol are all polyphyletic within the FOSC

To determine to what extent host specificity is polyphyletic among the selected strains, we inferred a phylogeny based on a concatenated alignment of 1195 conserved core genes (see Experimental procedures for more detail). The resulting tree, depicted in Fig. 2, is congruent with previously published phylogenetic analyses (Baayen *et al.*, 2000; Lievens *et al.*, 2007). We observe three major clades, of which clade 1 corresponds to a separate phylogenetic species (Laurence *et al.*, 2014). We find that ff. spp. are generally grouped into distinct clonal lineages and that these lineages are represented in distinct clades or subclades in the tree. One notable exception is Forc, of which all three sequenced strains group into a single clonal line. This is in accordance with previous reports in which 68 different Forc strains all belonged to a single Random Amplified Polymorphic DNA (RAPD) and Amplified Fragment Length

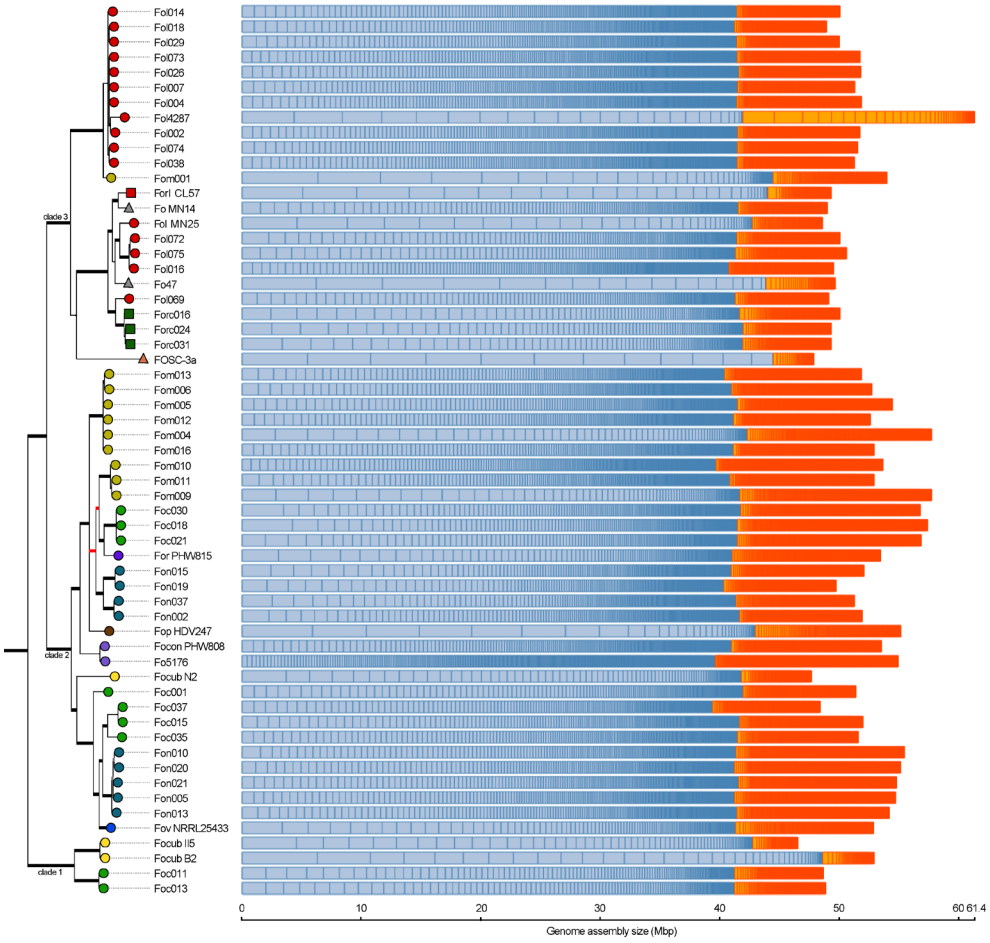


Fig. 2: Most strains analyzed in this study belong to polyphyletic *formae speciales* and contain a core genome of roughly the same size.
1195 genes residing on Fol4287's core genome were selected that have one-to-one orthologs in all *Fusarium* strains, using *F. verticillioides* 7600 as outgroup (outgroup not shown). A concatenated sequence alignment was generated using ClustalO (total length of 2,542,215 nt after trimming, including gaps) and phylogeny was inferred with 100 bootstrap iterations. Branches with most parsimonious bootstrap partitions greater than 90% are indicated in bold; those with a value greater than 70% are bold and shaded red. A coloured circle (wilting), square (root and shoot rot) or triangle (non-pathogenic / endophytic) representing the strain's host range was plotted on the leaves of the tree. Next to the dendrogram, the scaffolds in the genome assembly are plotted for each strain. Each scaffold is represented by a scaled rectangle and coloured blue if it is part of the core genome (based on alignment to the Fol4287 core genome) and orange if it is not (see Materials and Methods for more detail). The size of the core is very similar amongst different strains, in contrast to the size of the accessory genome, which varies even within a clonal line.

Polymorphism (AFLP) group (Vakalounakis and Fragkiadakis, 1999; Vakalounakis *et al.*, 2005; Lievens *et al.*, 2007).
We find that Foc, Fom and Focub are each represented in two different major clades, Foc and Focub even in two different phylogenetic species. Foc strains are distributed over five clonal lineages: one in clade 1 and four in clade 2. Focub (poorly represented in this study with only three strains) is grouped into two lineages, one that belongs to clade 1 and one strain that is placed in clade 2. Fom strains cluster into three clonal lineages, two that belong to different subclades in clade 2 and one strain (Fom001 / NRRL24604) that belongs to clade 3. Fon and Fol are confined to a single major clade, but within this clade we find that different clonal lines belong to distinct subclades. Fon is clustered into three distinct groups within clade 2. Fol harbours four clonal lineages in clade 3, with the majority of the sequenced strains belonging to the same clonal lineage as the reference strain Fol4287, corresponding to vegetative compatibility group 30 (VCG 0030). In summary, all ff. spp. except Forc are polyphyletic, consisting of several clonal lines.

Effector prediction pipeline yields 104 effector candidates in 59 Fo genomes

If host preference of a strain is largely determined by its suite of effectors, we expect strains that infect the same host to have a similar effector repertoire, even if they belong to different phylogenetic clades. To determine the effector repertoire of different strains, we exploited the association of *mimps* with effectors in Fo. Each genome was scanned for the presence of *mimp* TIRs. Subsequently, two methods were used for open reading frame (ORF) identification: i) the sequence 2500 bp downstream of the *mimp* IR was translated in the three possible reading frames and ORFs bigger than 25 codons were extracted; ii) AUGUSTUS 3.1 gene prediction software was run on the 5000 bp downstream of the TIR. In both cases, the threshold for distance from TIR to ATG was set to 2000bp. Method ii allowed for the prediction of putative effectors with a short first exon, like *SIX10*. Supplementary data Fig. S3 depicts a summary of the method.
To evaluate the influence of assembly fragmentation on the number of effectors found, we compared the number of *mimps*, *mimp*-inverted repeats and associated candidate genes identified in two different short read assemblies (based on Illumina HiSeq and Ion Proton sequencing) to the reference assembly of Fol4287 (Sanger/optical mapping). We found that the fragmented assemblies performed only slightly worse than the reference assembly (Table S1 and Table S3) in terms of effector candidate prediction, although clearly less complete *mimps* were found in the Ionproton assembly (21 versus 40).
When we applied the extended pipeline to all 59 genomes, we found in total 2242 ORFs that met the selection criteria (>25 aa; <300 aa; SignalP value >0.550; distance to closest *mimp* TIR <2500 bp). In order to reduce redundancy, we grouped these ORFs into gene families based on a self-BLAST search. This resulted in a final set of 201 *mimp*-associated gene families

encoding small secreted proteins. This set includes all previously described *SIX* genes of Fol except *SIX5* (Six5 returns a SignalP value of 0.444) and *SIX12* (Six12 does not have a signal peptide). Subsequently, we ran BLAST2GO and InterProScan to find gene ontology (GO) terms and protein domains in our set of genes, which resulted in a functional annotation for 88 out of 201 genes. Fifty four genes returned no significant BLAST hit and 59 returned a best BLASTX hit with a hypothetical protein (13 of these had at least one associated GO term). For those genes that had a homolog in any of the annotated Fo genomes, we also report the corresponding gene id in this strain (Table S2*, column L).

Due to the nature of the method, we sometimes picked up fragments of genes, originating from only the first exon or from an internal gene region. We checked for each candidate if the sequence could be aligned to the start site of the corresponding gene sequence from the Broad annotation in order to update the gene model. In 83 out of 201 candidates the reported Broad gene model could be used. For the others, we did not alter the gene model. To groom the list of candidates, we removed duplicate records (5x), very short protein products (<35aa after signal peptide removal; 23x); homologs or fragments of *SIX* genes (17x; all 14 *SIX* genes with correct gene models from Fol were manually added to the list instead, resulting in 215 records in Table S2) and sequences with an unlikely signal peptide based on visual inspection (67x), despite a positive (>0.550) detection of a signal peptide by SignalP. These included transposable elements, transcription factors and integral membrane proteins. This resulted in a final set of 104 candidate effectors (Table S2*).

Predicted proteins in the set with recognisable domains include FOVG_19456 which contains a LysM domain, potentially protecting the fungal cell wall or preventing chitin-triggered immunity in plants (de Jonge and Thomma, 2009; Stergiopoulos and de Wit, 2009; Jiang *et al.*, 2014), Rapid Alkalization Factor (RALF)-like protein 33 (FOIG_11494) and several secreted enzymes with predicted peptidase (FOXG_17323, FOVG_19376, FOXB_07727), polygalacturonase (FOTG_18786, FOWG_18016), glycoside hydrolase (FOCG_17303), carbonic anhydrase (FOMG_18585) or peroxidase (FOVG_19731) activity that could play a role in nutrient acquisition or suppressing plant defences.

Strains with the same host-specificity cluster together based on effector content

Having identified the combined putative 'effectorome' of the 59 Fo strains, we examined which putative effectors are shared amongst members of the same f. sp., multiple ff. spp., all analyzed Fo genomes or with other *Fusarium* species. Presence of a candidate effector gene in a genome was defined as having at least one blastn hit with an *e*-value ≤ 1*e*-03 and an identity score (number of identical nucleotides divided by the query length) of at least 30%. This gave rise to a binary 'effector-barcode' for each genome. Hierarchical clustering of these presence/absence patterns showed clear grouping of strains with the same host-specificity (Fig. 3), with

* Table S2 is not included in this thesis, but can be accessed online via http://onlinelibrary.wiley.com/doi/10.1111/1462-2920.13445/abstract

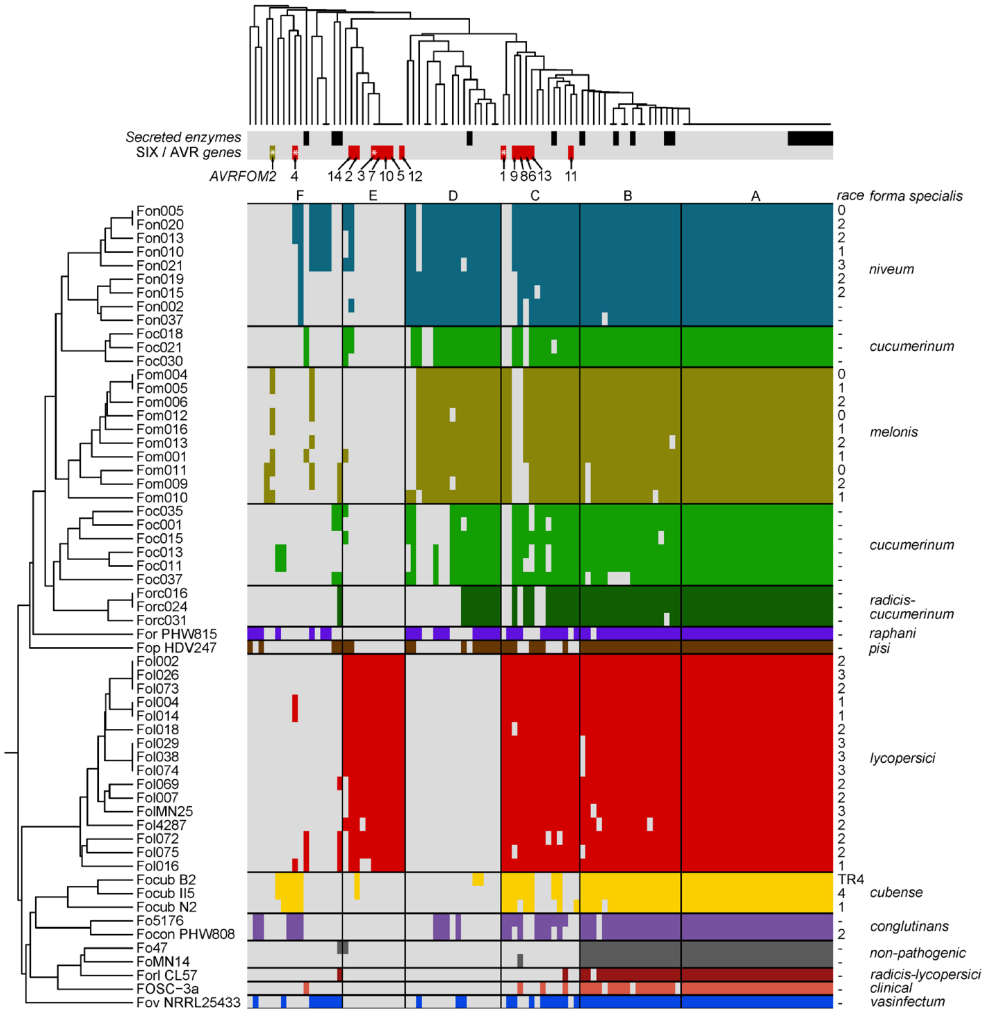


Fig. 3: *Formae speciales* of *F. oxysporum* cluster based on presence/absence of 104 candidate ORFs close to a *mimp*, identified in 59 genomes.
Presence was defined as detection with BLASTN with *e*-value ≤ 1*e*-3 and identity score (number of identical nucleotides divided by the query length) ≥ 30%. A color indicates presence, grey indicates absence. Both the genomes (rows) and effector candidates (columns, 104x) in the resulting table were clustered using a Jaccard binary distance matrix and average linkage. Top: secreted enzymes are marked with black, *SIX* genes and *AVRFOM2* with red, and all four described *AVR* genes are tagged with an asterisk. From left to right: *AVRFOM2*, SIX4* (AVR1), 14, 2, 3* (AVR2), 7, 10, 5, 12, 1* (AVR3), 9, 8, 6, 13, 11.*

the exception of Foc that split into two clades. We conclude that simple scoring presence/absence patterns of candidate effector genes reflects which host can be infected.

Candidate effector genes were then grouped based on their absence/presence in *formae speciales*. Genes in groups A and B are present in virtually all Fo genomes and consists mainly of secreted enzymes and hypothetical proteins, or genes without a BLASTX annotation.

Group C is composed of genes that are present in most pathogenic, but not non-pathogenic strains, including some *SIX* genes (marked in red at the top of Fig. 3). *SIX8* forms an exception in this group, it is also found as an intact ORF in FoMN14 while it is incomplete in the FOSC-3a genome assembly, situated at the end of a contig. Group D contains genes that are mostly present in cucurbit-infecting strains and group E represents Fol genes, including previously described Fol-specific effector genes *SIX2*, *3* and *5* (Lievens *et al.*, 2009).

The putative effector repertoires for strains infecting the same host are remarkably similar to each other, suggesting a highly related (at least on the level of presence-absence) shared accessory genome. Interestingly, the cucurbit-infecting strains form a supercluster distinct from the other *formae speciales*, implying that a significantly overlapping set of effectors is associated with the ability to infect cucurbits. This might explain why there is some degree of cross-pathogenicity of Foc towards e.g. melon plants but not from Fol to any of the cucurbits tested. Between cucurbit-infecting *formae speciales*, different sets of effectors are present (particularly in groups C and D) that may be involved in specific pathogenicity towards their host plant.

While still grouping with the other cucurbit-infecting *formae speciales*, the three (highly similar) Forc strains lack a number of putative effectors present in the other cucurbit-infecting strains, predominantly in group D. Their effector pattern is most similar to that of Foc strains not belonging to VCG0180.

Both Fol *AVR1* (Houterman *et al.*, 2008) and *AVRFOM2* (Schmidt *et al.*, 2016) show the expected presence-absence pattern (absence *of AVR1* in race 2 and 3 Fol strains and absence of *AVRFOM2* in race 2 Fom strains). Recognition of Avr2 in Fol is evaded by sequence difference rather than deletion of the gene and is present in all Fol strains in the figure.

Strikingly, the hierarchical clustering of genomes based on putative effectors also grouped strains belonging to the same clonal lineage within each polyphyletic *forma specialis,* irrespective of race designation (Fig. 3). This is a signature of vertical inheritance, indicating that the common ancestor of such a clonal lineage at a relatively recent point in time obtained the genetic information needed to infect a new host. Similar selection pressures on different clonal lineages likely resulted in the emergence of the same races in different lineages. This is most clear in the cases of Fol and Fom, where strains belonging to races 1, 2, 3 and races 0, 1, 2, respectively, do not cluster according to race, even though the core genome (Fig. 2), as well as the accessory genome (Fig. 3) are strongly similar within each clonal lineage.

Both Foc and Fon are clearly subdivided into two groups, whose accessory genomes are more different from each other compared to other polyphyletic *formae speciales*. Fon strains 005, 010, 013, 020, 021 form one clonal lineage (Fig. 2) and their putative effector profiles are also nearly identical, while the profiles of strains Fon002, 037, 015, 019 are also highly similar, but lack several effector candidates in group F. The same was found in the case of Foc, where Foc018, 021 and 030, belonging to one clonal lineage, are highly similar but quite different from the other Foc strains. This suggests relatively large differences in the accessory genomes

between clonal lineages in Fon and Foc.

Two non-pathogenic strains were included in this study: Fo47, a biocontrol strain (Aimé *et al.*, 2013) and FoMN14 that was isolated from diseased tomato plants from the same field as MN25 but turned out to be non-pathogenic towards tomato (Gale *et al.*, 2003). These two strains group close to each other and contain relatively few candidate effector genes (47 in Fo47 and 46 in FoMN14, compared to an average of 69 in pathogenic strains). The clinical strain FOSC-3a as well as Fo f. sp. *radicis-lycopersici* CL57 were found in the same branch. These strains also have a relatively low number of candidate effector genes. Based on core-genome sequence (Fig. 2), as well as effector content, Focub strains B2 (race 4) and II5 (tropical race 4; TR4) are nearly identical to each other with only two differentials. This is in line with Fo f. sp. *cubense* TR4 being a single clonal lineage (Ordonez *et al.*, 2015). Since these two strains do not differ that much from strain N2 in their effector pattern (race 1; six presence/absence polymorphisms), the Cavendish banana resistance breaking TR4 may have evolved from race 1 (or race 2, for which a the time of this study no genome sequence was available).

To see whether clustering of strains belonging to the same *forma specialis* depends on effector gene sequences and not DNA sequences close to *mimps per se*, we scored presence / absence of 2.5 kb regions downstream of a *mimp* TIR. The resulting figure (Fig. S4) shows that although some *formae speciales* form a single group, much more fragmentation is found and the clustering generally follows the core genome tree (Fig. 2). For example, Fom001 is found among the Fol strains and Foc011 and 013 are close to Focub strains.

For comparison, the pipeline was also run on 29 other *Fusarium* (non-Fo) genome assemblies of strains for which a whole-genome assembly is available. In most cases, 0 to 4 *mimps* or *mimp*-TIRs could be identified and only in the species *F. avenaceum, F. fujikuroi* and *F. nygamai*, 1-4 candidate ORFs with a secretion signal downstream of a *mimp* TIR were found (Table S4). This illustrates the specificity of this method for the FOSC and suggests that specifically within this species complex, *mimps* developed their contextual association with effector genes.

Sequence comparison of *SIX* genes shows evidence of horizontal transfer

Strong grouping of host-specificity can already be seen based on presence/absence clustering. We next wanted to find out if a higher resolution could be achieved by comparing sequence types of (candidate) effectors. We aligned the homologs of *SIX1-14* and looked at their phylogeny. Remarkably, different clonal lines within the same f. sp. possess identical or highly similar sequence variants of each of the 14 described *SIX* genes.

The *SIX1* coding sequence (Rep *et al.*, 2004; van der Does *et al.*, 2008) (Fig. 4A) was detected in Fol, Fom, Focub, Focon and Fop, with strains within each f. sp. possessing a (nearly) identical sequence for this gene. All Fon strains except Fon002 and 037 possess a *SIX1* homolog that is interrupted by a Hornet1-TE at the same position, resulting in a pseudogene. Fom009,

010 and 011 have a second copy of *SIX1* that seems to have been duplicated in the ancestor of this clonal lineage, indicative of vertical inheritance. For *SIX6* (Gawehns *et al.*, 2014), five sequence variants were found: Fol, Foc, Forc-Fon-Fom, Fon015/019 and Focub (Fig. 4B). This suggests that the cucurbit-infecting *formae speciales* Forc (single lineage), Fom (strains from multiple lineages) and Fon (one lineage) have acquired this sequence from the same source. *SIX13* (Schmidt *et al.*, 2013) harbours more sequence variation, but again all Fom, Fol and Focub strains have identical sequences within the respective f. sp. (Fig. 4C). Interestingly, two different *SIX13* sequence variants were found in Foc and Fon, with one of the two variants

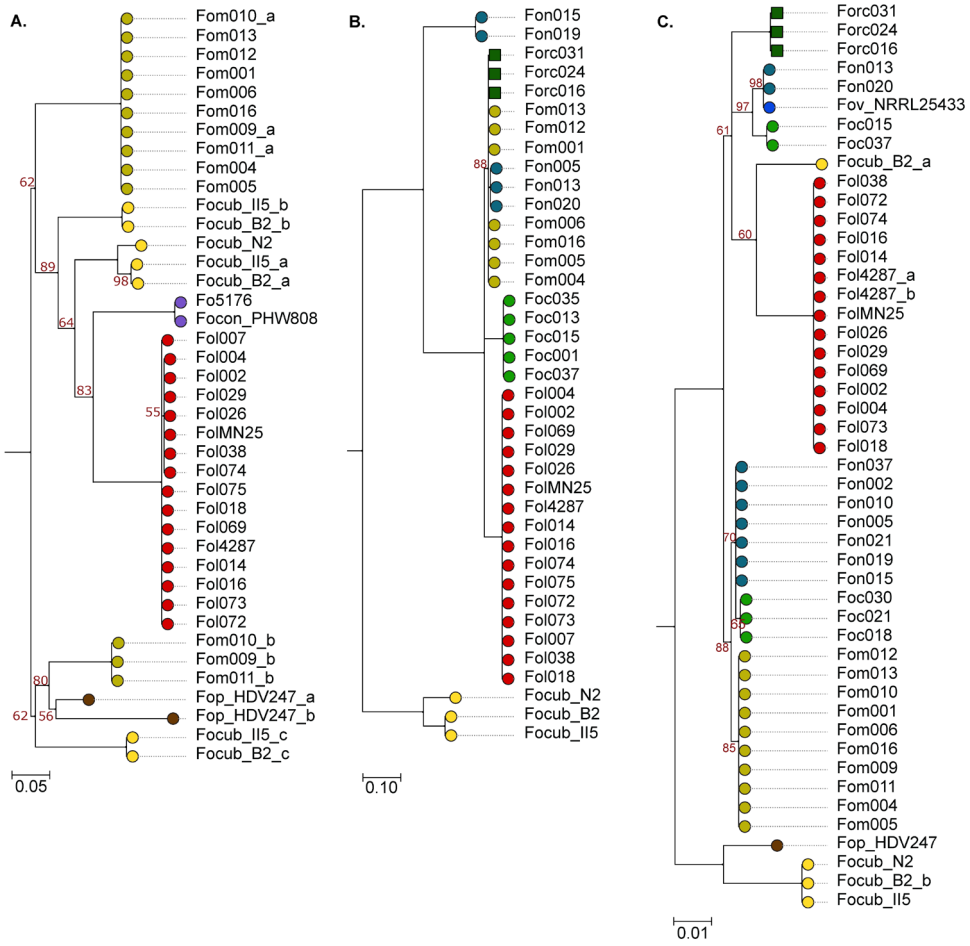


Fig. 4: Identical *SIX* gene sequence types are found in strains belonging to polyphyletic *formae speciales*, suggesting a combination of vertical and horizontal inheritance of these genes.
A MUSCLE alignment was made of the nucleotide sequence of (a) *SIX1* (881 nt), (b) *SIX6* (727 nt) and (c) *SIX13* (943 nt). Phylogeny was inferred using PhyML with 100 bootstrap iterations and plotted with mid-point rooting. Branches with most parsimonious bootstrap partitions below 50% were collapsed; values ≥ 50% and < 100% are indicated in red; 100% are not indicated. A coloured circle (wilting) or square (root and shoot rot) representing the strain's *forma specialis* was plotted on the leaves of the dendrogram.

being similar to the *ForcSIX13* homolog and the other most closely related to *FomSIX13*. *SIX5, 7, 10* and *12* were only found in Fol, with no sequence variation. Earlier studies have shown that some of these genes can be used as Fol-markers (Lievens *et al.*, 2009). *SIX3 (AVR2)* was also only encountered in Fol, albeit with some sequence variation in strains of Fol that evade *I-2* recognition, thus becoming race 3 (V41→M, R45→H or R46→P) (Takken and Rep, 2010). *SIX2, 4, 8, 9, 11* and *14* (de Sain and Rep, 2015) show similar distributions of sequence types as described for *SIX1, 6* and *13*. Phylogenetic trees of these genes can be found in Fig. S5. *SIX2* and *SIX14* were also found in *F. verticillioides* 7600 (not shown).
The *SIX* gene sequences evaluated here do not show the same phylogeny as conserved (core) genes (Fig. 2), which strongly suggests horizontal transfer of these genes. However, Foc, Fom and Fon strains sometimes also have very different *SIX* gene sequence types, suggesting a highly dynamic origin of host-specificity to cucurbits, possibly with multiple horizontal transfer events.

Most effector genes are expressed during invasive growth

To assess whether the candidates identified were indeed expressed during plant infection, we performed RNA sequencing on RNA extracted from infected plant roots ten days after inoculation and five-day-old mycelium from an *in vitro* liquid KNO_3 medium culture. We assessed eight strains belonging to the ff. spp. Foc, Fom, Fon, Forc, Fol. On average, 60.6% (± 3.9% S.D.) of the candidates showed evidence for *in planta* transcription, and 36.7% (± 10.7% S.D.) of the candidates qualified as being clearly expressed *in planta*, but with no or little expression *in vitro* (Fig. S6). In Fol4287, 12 out of 13 SIX genes fall inside the latter category (Table S5*).

Discussion

Using a bioinformatics pipeline, putative effectors were identified in 45 newly sequenced strains and 14 publicly available Fo genomes. Effector candidates were identified by searching for a small terminal inverted repeat sequence of miniature Impala transposable elements. In this way, 104 effector candidates were identified across all 59 *F. oxysporum* genomes. Hierarchical clustering of the presence/absence patterns of these sequences led to grouping of strains that largely coincided with host specificity. This indicates that the accessory genome and specifically the effector genes residing in these regions can be used to identify *formae speciales* in the FOSC. Clustering of ff. spp. was observed even without taking sequence divergence or copy number variation into consideration, underlining the robustness of the method.

* Table S5 is not included in this thesis, but can be accessed online via http://onlinelibrary.wiley.com/doi/10.1111/1462-2920.13445/abstract

We have improved an existing method to identify putative and real effector genes in *F. oxysporum* (Schmidt *et al.*, 2013, 2016) by incorporating a gene prediction module into the pipeline. This at least partially solved the issue of effector genes that were not recognized due to a short first exon, so that *SIX10* was now identified, too. The remaining known false negatives were *SIX5* and *SIX12*, which were not identified due to a low SignalP score or the absence of a signal peptide, respectively. False positives in the list, like duplicate records, very short gene fragments, transcription factors, transposable elements, membrane proteins and hypothetical proteins with an unlikely signal peptide were identified and removed from the list. The effect of fragmentation of the genome assembly on the number of effectors identified was limited, and by combining the effector candidates from multiple short-read Illumina genomes into one large list, a sufficient resolution was achieved. The level of fragmentation in the Illumina assemblies, did, however, prevent an analysis of the degree to which putative effector genes form clusters in the same genomic regions and possible synteny of such regions between strains.

Other methods of computationally recognizing effector genes include the selection of (predicted) proteins that meet some or all of the following requirements: small size, secreted, high number of cysteines, traces of diversifying selection and *in planta* transcriptional induction (Sperschneider, Dodds, *et al.*, 2015; Sperschneider, Gardiner, *et al.*, 2015). The advantage of the method presented here is that genome annotation is not necessary and that relevant contextual genome location is taken into account. However, while the association between effector genes and *mimps* has been demonstrated to hold true in many cases (e.g. *SIX* genes, *Avr-Fom2*) (Schmidt *et al.*, 2013, 2016), it is possible that several effector candidates have been missed.

For Fom, Fon, Forc and Fol a clear and unambiguous clustering could be observed. Foc separated into two clades. Previously, RAPD has been applied as a way of distinguishing between cucurbit-infecting *formae speciales* (Vakalounakis and Fragkiadakis, 1999; Wang *et al.*, 2001; Vakalounakis *et al.*, 2004; Lievens *et al.*, 2007). Aside from being laborious, difficult to replicate and time-consuming, the technique focuses on both the core as well as the accessory genome, which results in different RAPD profiles for each clonal lineage. Vakalounakis & Fragkiadakis (1999) showed that by transforming RAPD patterns into a binary data matrix and calculating the genetic distance value based on presence or absence of RAPD bands, a similar dendrogram could be generated as shown in our study (Fig. 3), again with VCG 0180 of Foc (here represented by Foc018, 021 and 030) clustering in a different group than the other Foc clonal lines. Both Forc (68 strains) and Fom (three related strains from Spain) cluster as single clades in the RAPD study. The fact that clustering based on all DNA sequences associated with *mimps* resulted in a more fragmented clustering showing more traces of the core phylogeny (Fig. S4) illustrates the importance and relevance of clustering based on ORFs for secreted proteins close to a *mimp* in *F. oxysporum*.

For about 60% of the identified effector genes, evidence for transcription was found *in planta*

ten days post inoculation (Fig. S5), indicating that many of these genes might play a role during infection. Some candidate genes might be expressed earlier in the infection process and are therefore not found to be expressed at this timepoint.

Focub strains II5, N2 and B2 and *Brassicaceae*-infecting strains PHW808 (f. sp. *conglutinans*) and Fo5176 (no f. sp. reported, but likely f. sp. *conglutinans*) also clustered together, but these *formae speciales* were not sampled extensively enough in this dataset to cover the complete variation present. In both Fon and Foc, and more subtly also in the other *formae speciales*, footprints of vertical inheritance could be seen. Strains from a single clonal lineage usually contain (near) identical putative effector presence patterns (Fig. 3) as well as *SIX* gene sequences (Figs 4 and S5).

Foc018, 021, 030 as well as Foc011 (all lacking *SIX6*) were pathogenic on *C. sativus* cv. Paraiso, but not on 'Foc-resistant' *C. sativus* cv. Melen plants; the other Foc strains also caused symptoms in cv. Melen. Since they are not in the same core genome clade (Fig. 2), the cause of this race differentiation must have developed at least twice independently, or passed on horizontally. Foc013, which is in the same phylogenetic species as Foc011 (clade 1), was aggressive on both cucumber cultivars.

Because strains belonging to the same *forma specialis* cluster together, the presence of certain putative effectors can indicate with high confidence to which *forma specialis* a strain belongs. For instance, a distinct cluster of putative effector genes is associated with cucurbit-infecting strains (Fig. 3, cluster D). Furthermore, all of the ff. spp. represented by multiple strains share sets of differentiating effector genes. In the case of Fol it was shown in previous studies that presence of effector genes can be indicative of the host; a PCR screen with *SIX1, SIX2, SIX3* and *SIX5* primers resulted in a 100% success rate in identifying Fol strains in a large collection containing fifteen other Fo ff. spp. (Van Der Does *et al.*, 2008; Lievens *et al.*, 2009). From the clustering shown in Fig. 3 we conclude that the suite of putative effector genes present in the genomes of cucurbit-infecting strains, as well as other mono- or polyphyletic *formae speciales* can be applied diagnostically if the genome sequence of a novel unknown strain is available. Members of the same *forma specialis* might share parts of their (accessory) genome involved in specific pathogenic interaction with their host. An important requirement is the completeness of sampling in order to be able to make a well-supported choice of *forma specialis*-specific marker loci (e.g. group E for Fol, genes in groups C and D for Fom and Foc, cluster F for Fon). Such marker loci are essentially the smallest possible set of effectors that is shared by all strains of a f. sp. and absent (at least as a set) in all other strains. This would greatly improve host range identification, something that so far has proved to be a difficult objective to achieve. Applicability to other *formae speciales* will need to be investigated, for example by sequencing genomes of multiple differential strains of the same *forma specialis* in existing culture collections. A similar clustering result would be expected for other host ranges besides tomato and cucurbits. This would allow *forma specialis*-identification for crops where Fusarium wilt and root and shoot rot is a pressing problem.

Materials and Methods

Plant lines and fungal strains

The following plant cultivars were used: *Cucumis sativus* cv. Paraiso (susceptible to Foc), *C. sativus* cv. Melen (resistant to Foc), *C. melo* cv. Cha-T (susceptible to Fom), *Citrullus lanatus* cv. Black Diamond (susceptible to Fon) and *Solanum lycopersicum* C32 (susceptible to Fol). Strains of Fo were selected based on core genome divergence, pathogenicity testing and geographical origin (Table 1). *F. oxysporum* was grown at 25°C in the dark on Czepek Dox agar (CDA, Difco) plates containing 100 mg/L penicillin and 200 mg/L streptomycin.

Pathogenicity testing

Pathogenicity testing was performed using the root dip method (Wellman, 1939). In short, conidia were isolated from five-day-old cultures in NO3-medium (0.17% yeast nitrogen base, 3% sucrose, 100 mM KNO_3) by filtering through miracloth (Merck; pore size of 22–25 μm). Spores were centrifuged, resuspended in sterile MilliQ water, counted and brought to a final concentration of 10^7 spores/mL. When the first true leaves were emerging, seedlings were uprooted, inoculated, individually potted and kept at 25°C in the greenhouse. Two weeks (cucurbits) or three weeks (tomato) after inoculation, disease was scored using a disease index ranging from 0-4 based on published methods (Rep *et al.*, 2004; Vakalounakis *et al.*, 2004; Pavlou and Vakalounakis, 2005). In the case of Fusarium wilt caused by Fol, Fom, Fon and Foc, the following scoring system was used: 0, no symptoms; 1, plant stunted, slightly swollen and/or bent hypocotyl; 2, one or two brown vascular bundles at height of cotyledons; 3, at least two brown vascular bundles and growth distortion and wilting / clear root rot symptoms; 4, plant either dead or very small and wilted). Root and shoot rot caused by Forc infection was scored as follows: 0, no symptoms; 1, slight root rot symptoms, only at tip of main root; 2, root rot symptoms and stem lesions visible aboveground; 3, very clear root rot symptoms of the entire root system, often with a large lesion extending above the cotyledons; 4, plant either dead or very small and wilted.

From each bioassay combination, pieces of hypocotyl were collected randomly from five plants, surface sterilized with 96% ethanol and a slice was cut of each piece to be placed on CDA plates containing 100 mg/L penicillin and 200 mg/L streptomycin. After four days of incubation at 25°C in the dark, *Fusarium* outgrowth was assessed and scored.

DNA isolation, genome sequencing and assembly

F. oxysporum genomic DNA was isolated from freeze-dried mycelium that was harvested from five-day-old NO_3-medium (0.17% yeast nitrogen base, 3% sucrose, 100 mM KNO3) cultures. In the case of all Foc, Forc, Fom, Fon strains as well as Fol007 and Fol4287, DNA was isolated as described by Michielse *et al.* (2009). Briefly, mycelial powder was suspended in 2 ml DNA-extraction buffer (0.2 M TrisHCL pH 8.5, 0.25 M NaCl, 0.05 M

Table 1: Strains of which the genomes were sequenced and assembled in this study.

Strain [a]	Original designation	VCG [b]	Race [b]	Origin of strain	Reference
Foc001	Foc-1	0183	-	Japan	BL[c]
Foc011	9903-1	0186	-	China	Lievens *et al.*, 2007
Foc013	9904-1	0186	-	China	Lievens *et al.*, 2007
Foc015	9906-3	0184	-	China	Lievens *et al.*, 2007
Foc018	Afu-50(B)	0180	-	Crete, Greece	Lievens *et al.*, 2007
Foc021	ATCC 16416	0180	-	USA, Florida	Lievens *et al.*, 2007
Foc030	FOCU-22P	0180	-	Israel	Lievens *et al.*, 2007
Foc035	NETH 11179	0181	-	Netherlands	Lievens *et al.*, 2007
Foc037	Tf-213	0185	-	Japan	Lievens *et al.*, 2007
Forc016	33	0260	-	Canada	Lievens *et al.*, 2007
Forc024	Afu-11(A)	0260	-	Crete, Greece	Lievens *et al.*, 2007
Forc031	AK-2	0261	-	Crete, Greece	Lievens *et al.*, 2007
Fon002	CBS 418.90	-	-	Israel	Lievens *et al.*, 2007
Fon005	TX-471-1	0080	0	USA, Texas	Zhou and Everts, 2007
Fon010	F-016-1	0082	1	USA, Maryland	Zhou and Everts, 2007
Fon013	F-014-2	0082	2	USA, Maryland	Zhou and Everts, 2007
Fon015	F-063-1	0082	2	USA, Maryland	Zhou and Everts, 2007
Fon019	TX-X1D	0082	2	USA, Texas	Zhou and Everts, 2007
Fon020	F-099-1	0083	2	USA, (Delaware)	Zhou and Everts, 2007
Fon021	MD-ZE622	-	3	USA, Maryland	Zhou and Everts, 2007
Fon037	NRRL 38539	-	-	Israel	Hadar and Katan, 1989
Fom004	Fom 0122	0134	0	Spain	Schmidt *et al.*, 2016
Fom005	Fom 0123	0134	1	Spain	Schmidt *et al.*, 2016
Fom006	Fom 0124	0134	2	Spain	Schmidt *et al.*, 2016
Fom009	-	0135	2	Israel	Schmidt *et al.*, 2016
Fom010	-	-	1	Israel	Schmidt *et al.*, 2016
Fom011	-	-	0	Israel	Schmidt *et al.*, 2016
Fom012	ML2	0134	0	-	Schmidt *et al.*, 2016
Fom013	-	0134	2	Spain	Schmidt *et al.*, 2016
Fom016	Fom26	0134	1	-	Schmidt *et al.*, 2016
Fol002	WCS862 / E241	0030	2	Netherlands	Mes *et al.*, 1999
Fol004	IPO1530 / B1	0030	1	Netherlands	Mes *et al.*, 1999
Fol007	D2	0030	2	France	Mes *et al.*, 1999
Fol014	LSU-3	0030	1	USA, Louisiana	Mes *et al.*, 1999
Fol016	BFOL-51	0031	1	USA, Louisiana	Mes *et al.*, 1999
Fol018	LSU-7	0030	2	USA, Louisiana	Mes *et al.*, 1999
Fol026	BRIP 14844 (M1943)	0030	3	Australia	Mes *et al.*, 1999
Fol029	5397	0030	3	USA, Florida	Mes *et al.*, 1999
Fol038	CA92/95	0030	3	USA, California	Lievens *et al.*, 2009
Fol069	DF0-23	0035	2	USA, California	Cai *et al.*, 2003
Fol072	DF0-38	0031	2	USA, California	Cai *et al.*, 2003
Fol073	DF0-40	0030	2	USA, California	Cai *et al.*, 2003
Fol074	DF0-41	0030	3	USA, California	Cai *et al.*, 2003
Fol075	DF0-62	0031	2[d]	USA, California	Cai *et al.*, 2003
FoMN14	MN-14	-	N.P.	USA, California	Gale *et al.*, 2003

[a] Foc: Fo f. sp. *cucumerinum*, Forc: Fo f. sp. *radicis-cucumerinum*, Fon: Fo f. sp.*niveum*, Fom: Fo f. sp. *melonis*, Fol: Fo f. sp. *lycopersici*.

[b] Race designation and VCG was taken from the corresponding reference. In the case of Fol, inoculation on differential cultivars (differing in their I (immunity)-gene genotype: C32/KG52201/PV2002MM (i-i2-i3; susceptible); GCR161 (I-i2); OT264 (KG324)/341F (i-I2); C295 (I-I2-i3); E779 (i-i2-I3)) was used to confirm the race of each strain (N.P.: non-pathogenic on tomato).

[c] Foc001 (Foc-1) was obtained from Bart Lievens, Scientia Terrae Research Institute, Belgium.

[d] Fol075 (DF0-62) was reported to be non-pathogenic. Based on retesting we now designate it as a (weak) race 2 strain.

EDTA pH 8.0, 48 mg/ml sodium 4-aminosalicylate dihydrate (PAS, Sigma-Aldrich), 8 mg/ml Triisopropylnaphthalenesulfonic acid sodium (TIPS, Sigma-Aldrich). 2 ml buffer-saturated phenol:chloroform:isoamylalcohol (25:24:1) was added, mixed and centrifuged for 30 minutes at 3,500 rpm and 4°C. DNA was precipitated from the aqueous phase with 0.7 volume isopropanol and 0.1 volume 4 M NaCl and centrifugation for 30 minutes at 3,500 rpm and 4°C. The pellet was resuspended in 500 µl TE buffer (10 mM TrisHCL pH7.5, 1 mM EDTA pH 8.0) and three consecutive rounds of phenol:chloroform:isoamylalcohol isolation were performed. DNA was precipitated from the aqueous phase with 2.5 volume 96% ethanol and 0.1 volume 4 M NaCl. The pellet was resuspended in 100 µl Milli-Q and treated with RNase A (Roche) and proteinase K (Fermentas, PCR grade). 500 µl Milli-Q was added and one more round of phenol:chloroform:isoamylalcohol was executed. DNA was precipitated with 2.5 volume 96% ethanol and 0.1 volume 4 M NaCl, washed with ethanol, airdried and finally dissolved in 50 µl TE buffer.

DNA isolation of the remaining Fol strains was performed using the Omniprep™ for Fungus (G-Biosciences) kit, followed by an additional purification step by phenol:chloroform extraction and precipitation with 2.5 volume 96% ethanol and 0.1 volume 4 M NaAc.

Library preparation of insert size 550 bp and Illumina HiSeq 2000 (Foc001, 011, 013, 015, 021, 035, 037, Forc016, 024, 031) and Illumina HiSeq 2500 (Fon and Foc018, 030) paired-end sequencing was performed at Keygene N.V. (Wageningen, the Netherlands). All Fom genomes, 'Fol4287-illumina' and Fol007 were sequenced and assembled as described in Schmidt *et al.* (2016) at the Beijing Genome Institute (BGI, Hong Kong), using multiple insert libraries. 'Fol4287-ionproton' was sequenced at the University of Amsterdam (MAD Dutch Genomics Service & Support Provider) using an Ion Proton™ Sequencer (Thermo Fisher Scientific). Fol paired-end genome sequencing of the remaining strains was performed using Illumina HiSeq 2000 180 bp insert libraries at the Broad Institute of Harvard and MIT.

Sequencing reads were trimmed for quality and adapter sequences with FastqMcf v1.04.676 (http://code.google.com/p/ea-utils, quality threshold=20). A preliminary assembly was made and checked for presence of contaminant contigs with blobology (Kumar *et al.*, 2013) (manual parameters were set to extract these contigs based on best BLAST hit, GC-content and coverage). In the case of Foc030, Fon005, Fon013, Fon015, Fon019 and Fon021, bacterial contamination with *Achromobacter xylosoxidans* was determined. Read pairs that could not be mapped with bowtie2 v2.2.5 against contaminated contigs were *de novo* assembled, mapped and checked for remaining contaminations with blobology three times, followed by a *de novo* assembly using CLC-workbench 8.0. Default settings were used, except "minimum contig length=500". Then, finally, any remaining contaminated contigs were removed manually from the CLC assembly. All other genomes were assembled directly in CLC workbench as described. The completeness of the assemblies was assessed with CEGMA v 2.5 (Parra *et al.*, 2009). We determined the repeat content of all genomes using RepeatMasker (Smit *et al.*, 1996) with '-species ascomycota' and repeat libraries from RepBase (version 20140131) (Jurka

et al., 2005).

We used nucmer (with --maxmatch) from the MUMmer package (Delcher *et al.*, 2002) to align all genome sequences to the reference genome, that of Fol4287. We designated scaffolds as 'core' if its best match in terms of bp that could be aligned was a scaffold that is part of a core chromosome in Fol4287 (with a minimum overlap of either query or the reference scaffold of 30%). All scaffolds that did not meet this criterium were designated 'accessory'.

In our assemblies, repeats that are longer than the length of our reads will typically be collapsed into a single sequence, the cumulative length of individual contigs in our assemblies is likely to be an underestimate of true genome size. To improve upon this estimate we determine the coverage by taking the median read depth of all contigs > 100kb, assuming that these large contigs contain relatively few repeats and divide the number of reads that are mapped to our assembly by the coverage to obtain a putative genome size (Table S1).

RNA isolation and transcriptome sequence analysis

For transcriptome sequencing, 10-day-old melon Cha-T (Fom), cucumber Paraiso (Foc, Forc), watermelon Black Diamond (Fon) and tomato C32 (Fol) seedlings were inoculated with conidia of strain Fom001 by dipping the roots in the spore suspension for 5 min; roots of infected plants were harvested ten days after inoculation and flash-frozen in liquid nitrogen. Additionally, mycelium from five-day-old in vitro KNO_3 cultures was harvested. Total RNA was extracted as described previously (Schmidt *et al.*, 2013). cDNA synthesis, library preparation (200 bp inserts) and Illumina sequencing was performed at BGI (Foc013, Fom001, Forc016, Forc031, Fol4287) and Keygene N.V. (Fon019, Fon020, Fom006).

The obtained reads were mapped against the set of candidate effector sequences retrieved from the genome sequence of each of the strains (only sequences starting with 'ATG' were used) with CLC workbench v8.0 with both length and similarity settings set to 0.9. The number of unique mapped reads was divided by the gene length, then multiplied by 1000 to find the number of reads per kb (RPK). This value was then used to calculate relative expression (RE) compared to EF1 alpha expression. Genes with five or more reads were distinguished as showing evidence for transcription. If a RE of >2% was found, the gene was considered to be strongly expressed (Table S5).

Data access

The Whole-Genome Shotgun projects for the newly sequenced strains of Foc, Forc, Fom and Fon have been deposited at Genbank under the BioProject PRJNA306247. Raw sequence data have been deposited into the Sequence Read Archive under the accession numbers SRP067515 (Foc, Fon, Forc, Fol007, Fol4287), SRP042982 (Fom) and SRP002087 (Fol). All publically available genome sequences that were used were obtained from Genbank and the Broad Institute of Harvard and MIT (http://www.broadinstitute.org). Detailed information on these strains and their respective accession numbers are listed in Table 2.

Table 2: Genome assemblies collected from GenBank and the Broad Institute of Harvard and MIT used in this study.

Strain	NRRL #	VCG	Race	f. sp.	Host	GenBank accession
4287	34936	0030	2	*lycopersici*	*Solanum lycopersicum*	GCA_000149955.2
MN25	54003	0033	3	*lycopersici*	*Solanum lycopersicum*	GCA_000259975.2
CL57	26381	0094	-	*radicis-lycopersici*	*Solanum lycopersicum*	GCA_000260155.3
HDV247	37622	-	-	*pisi*	*Pisum sp.*	GCA_000260075.2
PHW808	54008	0101	2	*conglutinans*	*Brassica sp.*	GCA_000260215.2
PHW815	54005	0102	-	*raphani*	*Raphanus sp.*	GCA_000260235.2
Fo5176	-	-	-	*(Brassica)*	*Brassica sp.*	GCA_000222805.1
Fom001	26406	0136	1	*melonis*	*Cucumis melo*	GCA_000260495.2
Fov	25433	-	-	*vasinfectum*	*Gossypium sp.*	GCA_000260175.2
II-5	54006	01213	TR4	*cubense*	*Musa sp.*	GCA_000260195.2
N2	-		1	*cubense*	*Musa sp.*	GCA_000350345.1
B2	-	-	4	*cubense*	*Musa sp.*	GCA_000350365.1
FOSC 3-a	32931	-	-	(human)	*Homo sapiens*	GCA_000271745.2
Fo47	54002	-	-	(biocontrol)	Soil	GCA_000271705.2

Phylogenetic analysis

We searched for homologs of 15,956 Fol4287 core genes (including introns) in the sequences of the other genomes using BLASTN with default parameters. We selected all sequences that overlap > 70% with the query sequence and more than 80% identity to the query. We then selected query genes for which we find only a single hit in each species, leaving us with 1194 genes. We used ClustalO (Sievers *et al.*, 2014) to construct a multiple sequence alignment for each query and a custom python script to concatenate these alignments. This alignment was trimmed using trimAl -strictplus. We used RaxML (with -T2 -N 100 -m GTRGAMMAIX -x 1234567 -p 123 -f a) to infer a phylogeny with 100 bootstrap replicates (Stamatakis, 2014).
For the phylogenetic analysis of the genes encoding Six proteins, we extracted the sequences by BLASTN (with -evalue .001 -task='blastn'), then manually curated them in order to extract complete coding sequences. We used MUSCLE (default values) to construct a multiple sequence alignment for each query, followed by PhyML to infer a phylogeny with 100 bootstrap replicates (Guindon and Gascuel, 2003).

Putative effector identification

Prediction of candidate effectors was performed as described (Schmidt *et al.*, 2013) on each genome using a custom Python script. Briefly, the genome was searched for a consensus sequence of the *mimp* inverted repeats (IRs), 'TT[TA]TTGCNNCCCACTGNN'. Subsequently, two methods were used for open reading frame (ORF) identification: i) the sequence 2500 bp downstream of the *mimp* IR was translated in the three possible ORFs and ORFs bigger than

25 codons were extracted; ii) AUGUSTUS 3.1 gene prediction software was run on the 5000 bp downstream of the *mimp* IR with options '--species=fusarium --singlestrand=true' (Stanke *et al.*, 2006). In both cases, the threshold for distance from TIR to ATG was set to 2000bp. Method ii allowed for the prediction of putative effectors with a short first exon, like *SIX10*. The identified putative ORFs were submitted to a local instance of SignalP4.1 with the option '-u 0.550' for stringent signal peptide prediction (Petersen *et al.*, 2011). All ORFs that met the criteria were collected. After concatenating the putative effectors of the 59 genomes, they were BLASTed against themselves in order to identify and remove redundant ORFs. For each set of homologous putative effectors (criteria: *e*-value: 1*e*-03; percent identity: >60%; alignment length: >60%), the longest entry was selected. A total of 201 candidates were saved and used for the downstream hierarchical clustering analysis. For functional annotation of the found candidate genes, we performed BLASTX and InterProScan using BLAST2GO v3.1.3 with standard settings (Conesa *et al.*, 2005).

Hierarchical clustering

Screening for presence of the putative effectors collected from multiple genomes was done by conducting a BLASTN search (-evalue 0.001, -task='blastn') on each genome assembly. Presence of a candidate effector gene in a genome was defined as having at least one blast hit with an *e*-value ≤ 1*e*-03 and an identity score (number of identical nucleotides in the correct position in the alignment divided by the query length) of at least 30%. A binary datamatrix was generated containing presence ('1') or absence ('0') of each candidate in each genome. This table was used as input for hierarchical clustering performed in R, using a Jaccard binary distance matrix and average linkage.

Acknowledgments

We are grateful to K. Broz for Fol DNA isolations and P.M. Houterman for Fol DNA isolations and retesting of Fol races. We thank B. Lievens, D.J. Vakalounakis and K. Everts for providing their cucurbit-infecting *F. oxysporum* strains. Furthermore, we are grateful to ENZA seeds and Rijk Zwaan for kindly providing cucumber, melon and watermelon seeds. This work was supported by the Horizon programme (project 93512007) of the Netherlands Genomics Initiative (NGI) and the Innovational Research Incentives Scheme Vici of The Netherlands Organisation for Scientific Research (NWO, project 865.10.002) through grants to M.R.

References

Aimé, S., Alabouvette, C., Steinberg, C., and Olivain, C. (2013) The endophytic strain *Fusarium oxysporum* Fo47: a good candidate for priming the defense responses in tomato roots. *Mol. Plant-Microbe Interact.* **26**: 918–26.

Armstrong, G.M. and Armstrong, J.K. (1981) *Formae speciales* and races of *Fusarium oxysporum* causing wilt diseases. In *Fusarium: diseases, biology, and taxonomy*. Pennsylvania State University Press, University Park, PA, pp. 391–399.

Baayen, R.P., O'Donnell, K., Bonants, P.J., Cigelnik, E., Kroon, L.P., Roebroeck, E.J., and Waalwijk, C. (2000) Gene Genealogies and AFLP Analyses in the *Fusarium oxysporum* Complex Identify Monophyletic and Nonmonophyletic *Formae Speciales* Causing Wilt and Rot Disease. *Phytopathology* **90**: 891–900.

Bergemann, M., Lespinet, O., M'Barek, S. Ben, Daboussi, M.J., and Dufresne, M. (2008) Genome-wide analysis of the *Fusarium oxysporum mimp* family of MITEs and mobilization of both native and *de novo* created *mimps*. *J. Mol. Evol.* **67**: 631–642.

Cafri, D., Katan, J., and Katan, T. (2005) Cross-pathogenicity between *formae speciales* of *Fusarium oxysporum*, the pathogens of cucumber and melon. *J. Phytopathol.* **153**: 615–622.

Cai, G., Gale, L.R., Schneider, R.W., Kistler, H.C., Davis, R.M., Elias, K.S., and Miyao, E.M. (2003) Origin of Race 3 of *Fusarium oxysporum* f. sp. *lycopersici* at a Single Site in California. *Phytopathology* **93**: 1014–1022.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.

Covey, P. a, Kuwitzky, B., Hanson, M., and Webb, K.M. (2014) Multilocus analysis using putative fungal effectors to describe a population of *Fusarium oxysporum* from sugar beet. *Phytopathology* **104**: 886–896.

Delcher, A.L., Phillippy, A., Carlton, J., and Salzberg, S.L. (2002) Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* **30**: 2478–2483.

Dodds, P.N. and Rathjen, J.P. (2010) Plant immunity: towards an integrated view of plant‚Äìpathogen interactions. *Nat. Rev. Genet.* **11**: 539–548.

van der Does, H.C., Duyvesteijn, R.G.E., Goltstein, P.M., van Schie, C.C.N., Manders, E.M.M., Cornelissen, B.J.C., and Rep, M. (2008) Expression of effector gene *SIX1* of *Fusarium oxysporum* requires living plant cells. *Fungal Genet. Biol.* **45**: 1257–1264.

Van Der Does, H.C., Lievens, B., Claes, L., Houterman, P.M., Cornelissen, B.J.C., and Rep, M. (2008) The presence of a virulence locus discriminates *Fusarium oxysporum* isolates causing tomato wilt from other isolates. *Environ. Microbiol.* **10**: 1475–1485.

Feschotte, C. and Pritham, E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**: 331–368.

Gale, L.R., Katan, T., and Kistler, H.C. (2003) The probable center of origin of *Fusarium oxysporum* f. sp *lycopersici* VCG 0033. *Plant Dis.* **87**: 1433–1438.

Gawehns, F., Houterman, P.M., Ichou, F.A., Michielse, C.B., Hijdra, M., Cornelissen, B.J.C., *et al.* (2014) The *Fusarium oxysporum* Effector Six6 Contributes to Virulence and Suppresses I-2-Mediated Cell Death. *Mol. Plant-Microbe Interact.* **27**: 336–348.

Giraldo, M.C. and Valent, B. (2013) Filamentous plant pathogen effectors in action. *Nat. Rev. Microbiol.* **11**: 800–14.

Guindon, S. and Gascuel, O. (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst. Biol.* **52**: 696–704.

Hadar, E. and Katan, J. (1989) The Use of Nitrate-Nonutilizing Mutants and a Selective Medium for Studies of Pathogenic Strains of *Fusarium oxysporum*. *Plant Dis.* **73**: 800.

Houterman, P.M., Cornelissen, B.J.C., and Rep, M. (2008) Suppression of plant resistance gene-based immunity by a fungal effector. *PLoS Pathog.* **4**: e1000061.

Houterman, P.M., Speijer, D., Dekker, H.L., De Koster, C.G., Cornelissen, B.J.C., and Rep, M. (2007) The mixed xylem sap proteome of *Fusarium oxysporum*-infected tomato plants. *Mol. Plant Pathol.* **8**: 215–221.

Jacobson, D.J. and Gordon, T.R. (1990) Variability of mitochondrial DNA as an indicator of relationships between populations of *Fusarium oxysporum* f. sp. *melonis*. *Mycol. Res.* **94**: 734–744.

Jiang, C., He, B., Huang, R., Huang, M., and Xu, L. (2014) Identification and functional analysis of LysM effectors from *Marssonina brunnea*. *Australas. Plant Pathol.* **43**: 615–622.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**: 462–467.

de Jonge, R. and Thomma, B.P.H.J. (2009) Fungal LysM effectors: extinguishers of host immunity? *Trends Microbiol.* **17**: 151–157.

Katan, T. (1999) Current status of vegetative compatibility groups in *Fusarium oxysporum*. *Phytoparasitica* **27**: 51–64.

Kim, D.H., Martyn, R.D., and Magill, C.W. (1993a) Chromosomal polymorphism in *Fusarium oxysporum* f. sp. *niveum*. *Phytopathology* **83**: 1209–1216.

Kim, D.H., Martyn, R.D., and Magill, C.W. (1993b) Mitochondrial DNA (mtDNA)-Relatedness among *formae speciales* of *Fusarium oxysporum* in the Cucurbitaceae. *Phytopathology* **83**: 91–97.

Kim, D.H., Martyn, R.D., and Magill, C.W. (1992) Restriction fragment length polymorphism groups and physical map of mitochondrial DNA from *Fusarium oxysporum* f. sp. *niveum*. *Phytopathology* **82**: 53.

Kistler, H.C. (1997) Genetic Diversity in the Plant-Pathogenic Fungus *Fusarium oxysporum*. *Phytopathology* **87**: 474–479.

Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M., and Blaxter, M. (2013) Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front. Genet.* **4**: 237.

Laurence, M.H., Summerell, B.A., Burgess, L.W., and Liew, E.C.Y. (2014) Genealogical concordance phylogenetic species recognition in the *Fusarium oxysporum* species complex. *Fungal Biol.* **118**: 374–384.

Lievens, B., Claes, L., Vakalounakis, D.J., Vanachter, A.C.R.C., and Thomma, B.P.H.J. (2007) A robust identification and detection assay to discriminate the cucumber pathogens *Fusarium oxysporum* f. sp. *cucumerinum* and f. sp. *radicis-cucumerinum*. *Environ. Microbiol.* **9**: 2145–2161.

Lievens, B., Houterman, P.M., and Rep, M. (2009) Effector gene screening allows unambiguous identification of *Fusarium oxysporum* f. sp. *lycopersici* races and discrimination from other *formae speciales*. *FEMS Microbiol. Lett.* **300**: 201–215.

Lu, C., Chen, J., Zhang, Y., Hu, Q., Su, W., and Kuang, H. (2012) Miniature Inverted-Repeat Transposable Elements (MITEs) Have Been Accumulated through Amplification Bursts and Play Important Roles in Gene Expression and Species Diversity in *Oryza sativa*. *Mol. Biol. Evol.* **29**: 1005–1017.

Ma, L.-J., van der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.-J., Di Pietro, A., *et al.* (2010) Comparative analysis reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367–373.

Ma, L.-J., Geiser, D.M., Proctor, R.H., Rooney, A.P., O'Donnell, K., Trail, F., *et al.* (2013) *Fusarium* pathogenomics. *Annu. Rev. Microbiol.* **67**: 399–416.

Ma, L., Houterman, P.M., Gawehns, F., Cao, L., Sillo, F., Richter, H., *et al.* (2015) The *AVR2-SIX5* gene pair is required to activate *I-2*-mediated immunity in tomato. *New Phytol.* **208**: 507–518.

Martyn, R.D. (2014) *Fusarium* Wilt of Watermelon: 120 Years of Research. *Hortic. Rev. Vol. 42* **42**: 349–442.

Mes, J.J., Weststeijn, E.A., Herlaar, F., Lambalk, J.J.M., Wijbrandi, J., Haring, M.A., and Cornelissen, B.J.C. (1999) Biological and molecular characterization of *Fusarium oxysporum* f. sp. *lycopersici* divides race 1 isolates into separate virulence groups. *Phytopathology* **89**: 156–160.

Michielse, C.B. and Rep, M. (2009) Pathogen profile update: *Fusarium oxysporum*. *Mol. Plant Pathol.* **10**: 311–324.

O'Donnell, K., Kistler, H.C., Cigelnik, E., and Ploetz, R.C. (1998) Multiple evolutionary origins of the fungus causing Panama disease of banana: concordant evidence from nuclear and mitochondrial gene genealogies. *Proc. Natl. Acad. Sci. U. S. A.* **95**: 2044–2049.

Ordonez, N., Seidl, M.F., Waalwijk, C., Drenth, A., Kilian, A., Thomma, B.P.H.J., *et al.* (2015) Worse Comes to Worst: Bananas and Panama Disease—When Plant and Pathogen Clones Meet. *PLOS Pathog.* **11**: e1005197.

Parra, G., Bradnam, K., Ning, Z., Keane, T., and Korf, I. (2009) Assessing the gene space in draft genomes. *Nucleic Acids Res.* **37**: 289–297.

Pavlou, G.C. and Vakalounakis, D.J. (2005) Biological control of root and stem rot of greenhouse cucumber, caused by *Fusarium oxysporum* f. sp. *radicis-cucumerinum*, by lettuce soil amendment. *Crop Prot.* **24**: 135–140.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**: 785–786.

Recorbet, G., Steinberg, C., Olivain, C., Edel, V., Trouvelot, S., Dumas-Gaudot, E., *et al.* (2003) Wanted: pathogenesis-related marker molecules for *Fusarium oxysporum*. *New Phytol.* **159**: 73–92.

Rep, M., Van Der Does, H.C., Meijer, M., Van Wijk, R., Houterman, P.M., Dekker, H.L., *et al.* (2004) 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.

de Sain, M. and Rep, M. (2015) The Role of Pathogen-Secreted Proteins in Fungal Vascular Wilt Diseases. *Int. J. Mol. Sci.* **16**: 23970–23993.

Schmidt, S.M., Houterman, P.M., Schreiver, I., Ma, L., Amyotte, S., Chellappan, B., *et al.* (2013) MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genomics* **14**: 119.

Schmidt, S.M., Lukasiewicz, J., Farrer, R., van Dam, P., Bertoldo, C., and Rep, M. (2016) Comparative genomics of *Fusarium oxysporum* f. sp. *melonis* reveals the secreted protein recognized by the *Fom-2* resistance gene in melon. *New Phytol.* **209**: 307–318.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., *et al.* (2014) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**: 539–539.

Smit, A., Hubley, R., and Green, P. (1996) RepeatMasker Open-3.0. url: http://www.repeatmasker.org/*RepeatMasker Open-3.0*.

Sperschneider, J., Dodds, P.N., Gardiner, D.M., Manners, J.M., Singh, K.B., and Taylor, J.M. (2015) Advances and Challenges in Computational Prediction of Effectors from Plant Pathogenic Fungi. *PLOS Pathog.* **11**: e1004806.

Sperschneider, J., Gardiner, D.M., Thatcher, L.F., Lyons, R., Singh, K.B., Manners, J.M., and Taylor, J.M. (2015) Genome-Wide Analysis in Three *Fusarium* Pathogens Identifies Rapidly Evolving Chromosomes and Genes Associated with Pathogenicity. *Genome Biol. Evol.* **7**: 1613–1627.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**: 62.

Stergiopoulos, I. and de Wit, P.J.G.M. (2009) Fungal effector proteins. *Annu. Rev. Phytopathol.* **47**: 233–263.

Takken, F. and Rep, M. (2010) The arms race between tomato and *Fusarium oxysporum*. *Mol. Plant Pathol.* **11**: 309–314.

Vakalounakis, D.J. (1996) Root and Stem Rot of Cucumber Caused by *Fusarium oxysporum* f. sp. *radicis-cucumerinum* f. sp. nov. *Plant Dis.* **80**: 313–316.

Vakalounakis, D.J., Doulis, A.G., and Klironomou, E. (2005) Characterization of *Fusarium oxysporum* f. sp. *radicis-cucumerinum* attacking melon under natural conditions in Greece. *Plant Pathol.* **54**: 339–346.

Vakalounakis, D.J. and Fragkiadakis, G.A. (1999) Genetic Diversity of *Fusarium oxysporum* Isolates from Cucumber: Differentiation by Pathogenicity, Vegetative Compatibility, and RAPD Fingerprinting. *Phytopathology* **89**: 161–168.

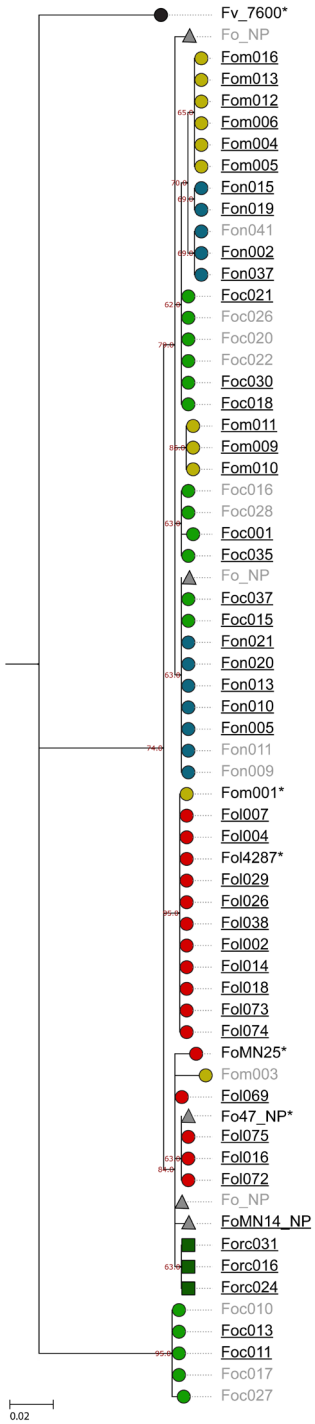
Vakalounakis, D.J., Wang, Z., Fragkiadakis, G.A., Skaracis, G.N., and Li, D.-B. (2004) Isolates Obtained from Cucumber in China by Pathogenicity, VCG, and RAPD. *Plant Dis.* **88**: 645–649.

Wang, P.H., Lo, H.S., and Yeh, Y. (2001) Identification of *F. o. cucumerinum* and *F. o. luffae* by RAPD-generated DNA probes. *Lett. Appl. Microbiol.* **33**: 397–401.

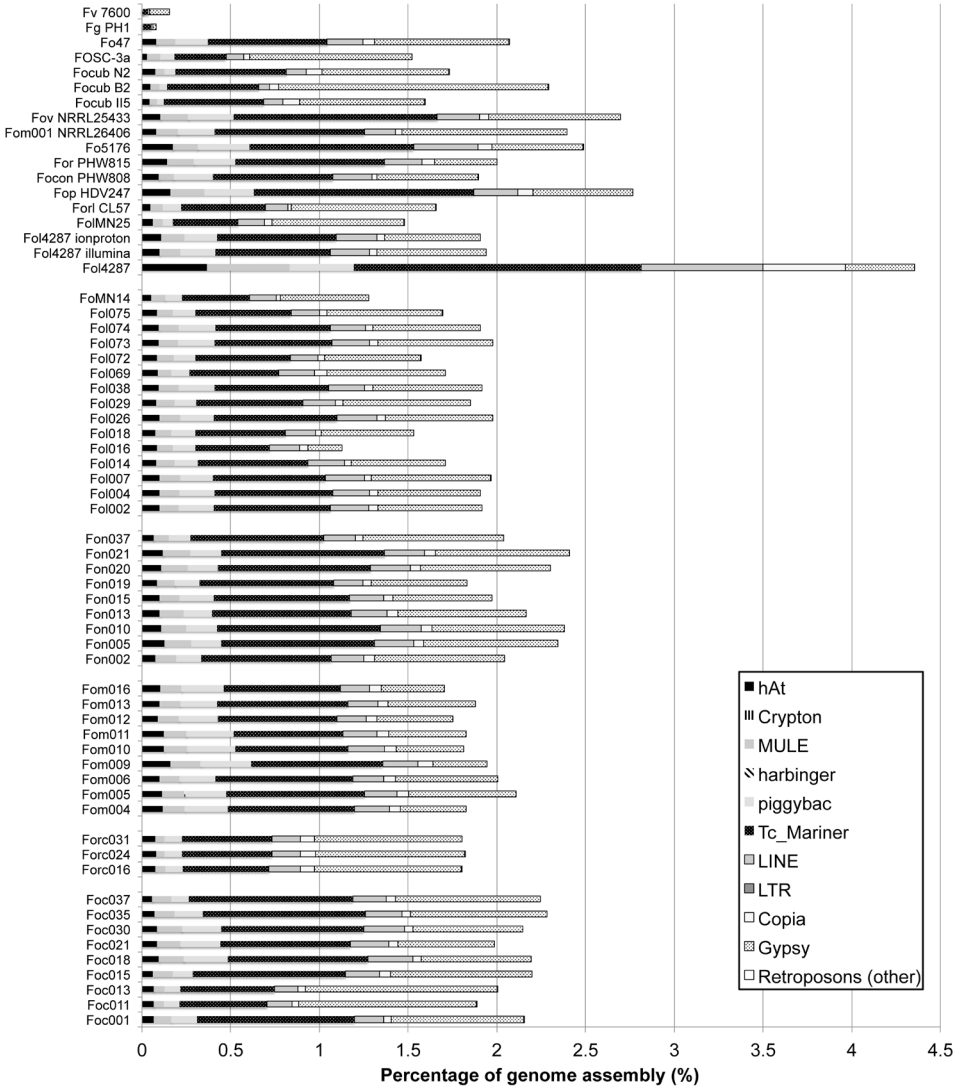
Wellman, F.L. (1939) A technique for studying host resistance and pathogenicity in tomato Fusarium wilt. *Phytopathology* **29**: 945–956.

Zhou, X.G. and Everts, K.L. (2007) Characterization of a Regional Population of *Fusarium oxysporum* f. sp. *niveum* by Race, Cross Pathogenicity, and Vegetative Compatibility. *Phytopathology* **97**: 461–469.

Supplementary data

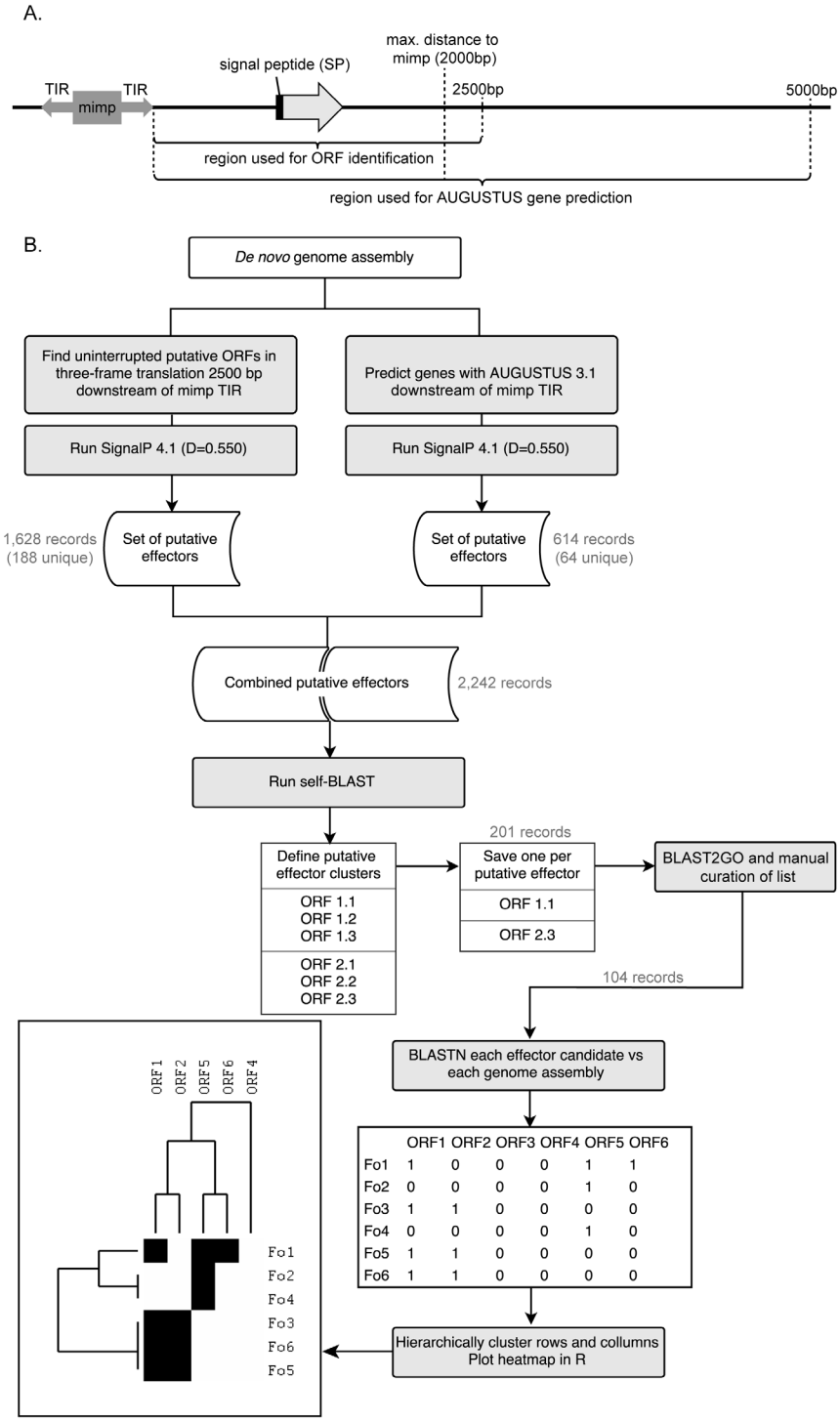


◀ **Fig. S1: Phylogenetic relationships of strains considered for whole genome sequencing inferred from the *EF-1α* sequence.**
Part of the *EF-1α* gene was PCR-amplified using primers FP889 (tcgtcgtcatcggccacgtc) and FP1614 (ggaagtaccagtgatcatgtt) (Van Der Does *et al.*, 2008) and sequenced using an ABI3730XL DNA Analyzer (Macrogen, the Netherlands). A MUSCLE nucleotide sequence alignment was made using 569 nt after trimming. Phylogeny was inferred using PhyML with 100 bootstrap iterations and plotted with *F. verticillioides* as an outgroup. Branches with most parsimonious bootstrap partitions below 50% were collapsed; values ≥ 50% and < 100% are indicated in red; 100% are not indicated. A coloured circle (wilting), square (root and shoot rot) or triangle (non-pathogenic ["_NP"] / other) representing the strain's *forma specialis* was plotted on the leaves of the dendrogram. Strains that are coloured grey were not selected for whole genome sequencing, those that are black and underlined were and those with an asterisk (*) have a publically available genome assembly.



▲Fig. S2: Repeat-density, RNA and DNA transposon abundance is largely similar both amongst the *de novo* assemblies as well as compared to the reference assemblies.
Simple repeats and low complexity comprise *c.* 0.61% ± 0.03% S.D. of the sequence of all of the *Fusarium* genome assemblies evaluated, including *F. verticillioides* 7600 (0.70%) and *F. graminearum* PH-1 (0.69%), while both RNA transposons (1.08% ± 0.35% S.D.) and DNA transposons (0.91% ±0.20% S.D.) are much more prevalent in *F. oxysporum* than the other two species.

▶Fig. S3: Method used for identifying candidate effector genes in a whole genome shotgun assembly.
(A) Putative effectors were identified by looking for uninterrupted ORFs in a region of i) 2500 bp or ii) 5000 bp downstream of a *miniature impala (mimp)* terminal inverted repeat (TIR). Maximum distance between the mimp terminal inverted repeat (TIR) and the start codon (ATG) of the putative ORF was set in both cases to 2000 bp. **(B)** Workflow diagram of mimp-associated putative effector discory pipeline from *de novo* genome assembly to identification of effector candidates, clustering them and identifying presence/absence patterns in each genome. In blue, the number of records found for the assessed Fo genomes is represented.


A.
max. distance to mimp (2000bp)
signal peptide (SP)
TIR
TIR
mimp
2500bp
5000bp
region used for ORF identification
region used for AUGUSTUS gene prediction
B.
De novo genome assembly
Find uninterrupted putative ORFs in three-frame translation 2500 bp downstream of mimp TIR
Predict genes with AUGUSTUS 3.1 downstream of mimp TIR
Run SignalP 4.1 (D=0.550)
Run SignalP 4.1 (D=0.550)
1,628 records (188 unique)
Set of putative effectors
Set of putative effectors
614 records (64 unique)
Combined putative effectors
2,242 records
Run self-BLAST
Define putative effector clusters
ORF 1.1
ORF 1.2
ORF 1.3
ORF 2.1
ORF 2.2
ORF 2.3
201 records
Save one per putative effector
ORF 1.1
ORF 2.3
BLAST2GO and manual curation of list
104 records
BLASTN each effector candidate vs each genome assembly
ORF1 ORF2 ORF3 ORF4 ORF5 ORF6
Fo1 1 0 0 0 1 1
Fo2 0 0 0 0 1 0
Fo3 1 1 0 0 0 0
Fo4 0 0 0 0 1 0
Fo5 1 1 0 0 0 0
Fo6 1 1 0 0 0 0
Hierarchically cluster rows and collumns
Plot heatmap in R
ORF1
ORF2
ORF5
ORF6
ORF4
Fo1
Fo2
Fo4
Fo3
Fo6
Fo5

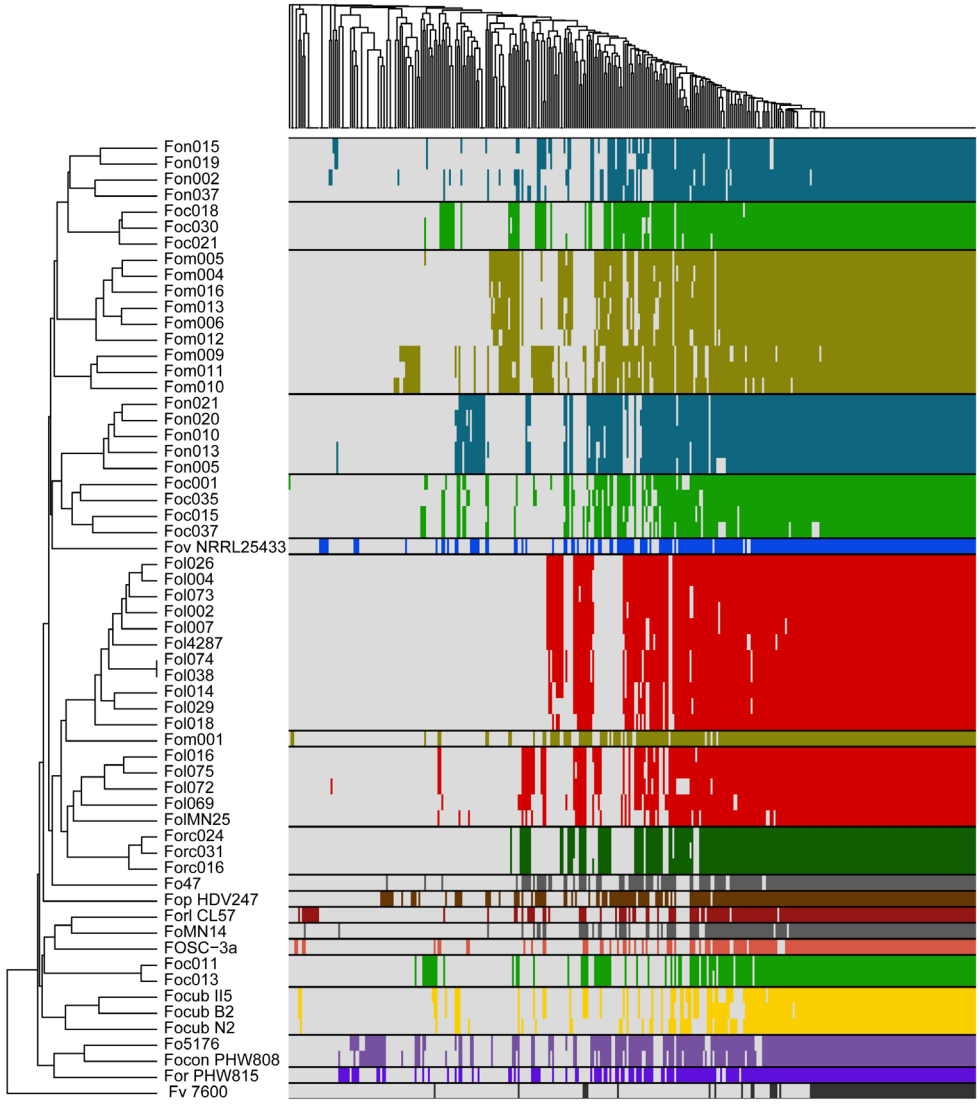


Fig. S4: Clustering of 2.5kb regions downstream of a mimp inverted repeat shows a more fragmented clustering compared to clustering based on effector candidates.
2.5 kb windows were extracted downstream of each occurrence of a mimp inverted repeats (IRs, 'TT[TA] TTGCNNCCCACTGNN'). Terminal ambiguous bases ('N's) were trimmed from the sequence and redundancy was reduced by self-BLASTN. Presence of 360 regions was detected using BLASTN with the same thresholds as candidate clustering. Traces of the core genome are more clearly visible, for example in the splitting of Fon into two groups, the position of Foc011 and Foc013 (clade 1 isolates, Fig. 2) close to Focub and Fom001 (NRRL26406) in between Fol isolates with a highly similar core genome.

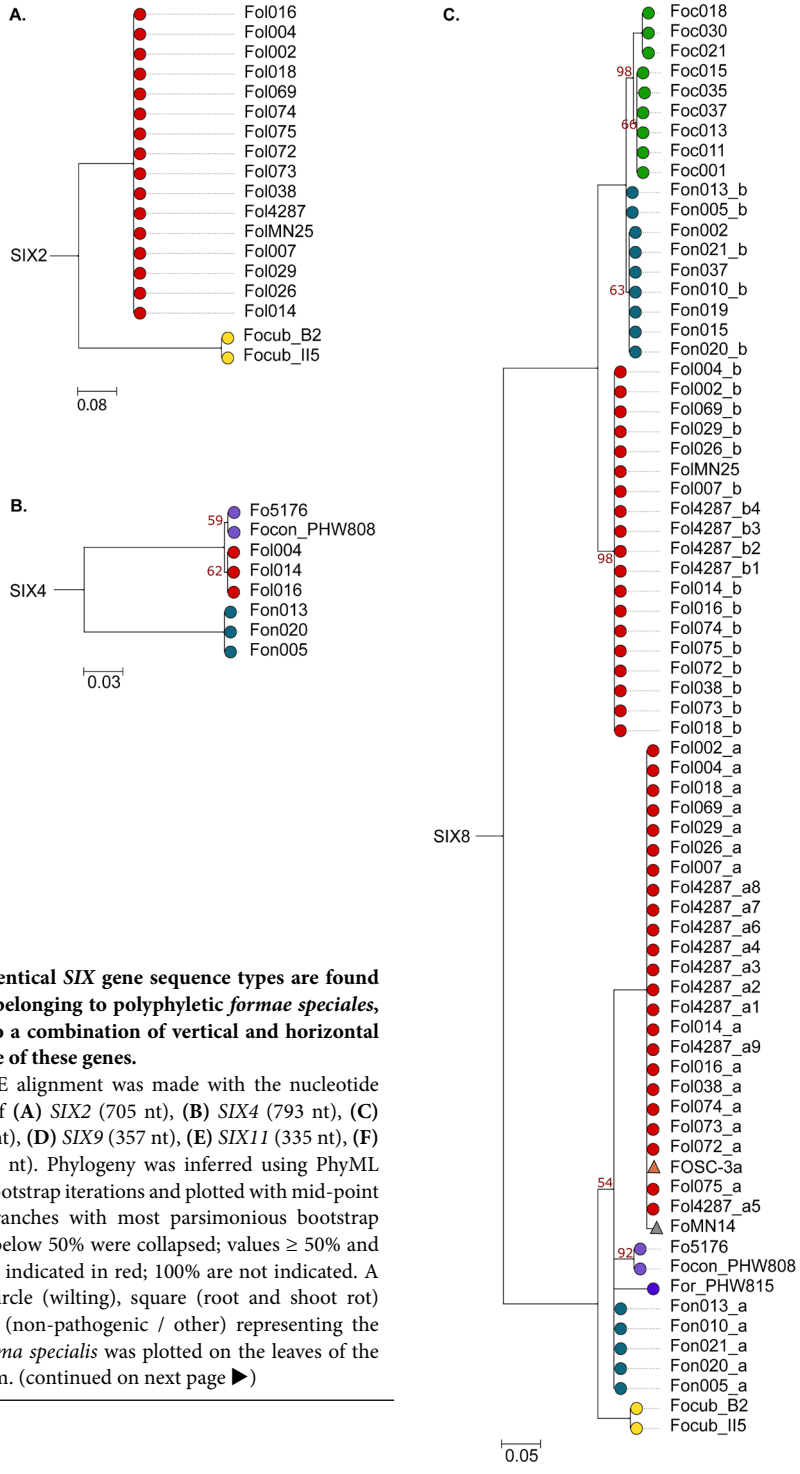
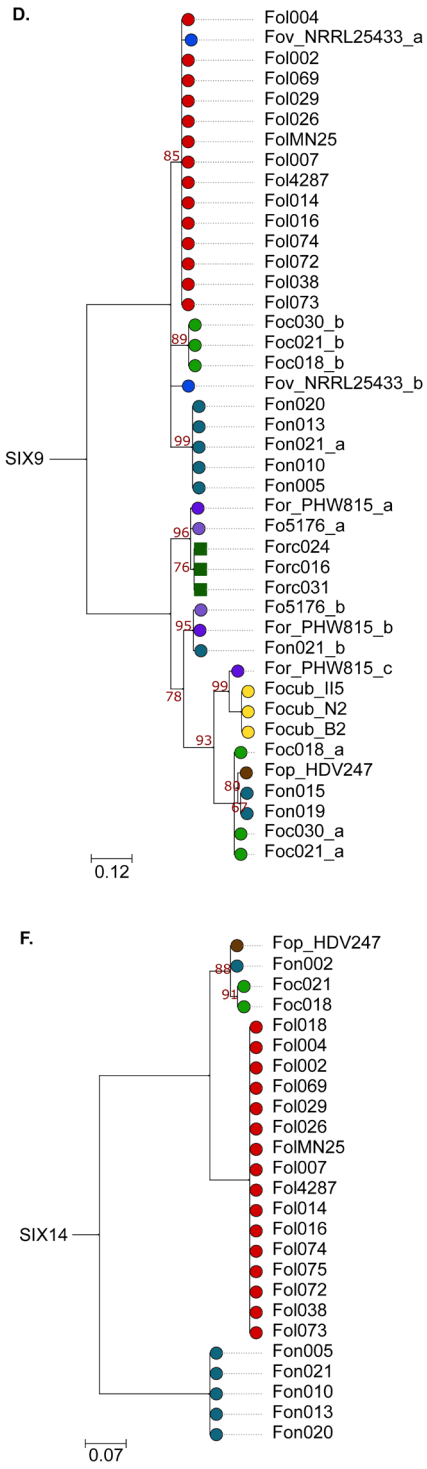


Fig. S5: Identical *SIX* gene sequence types are found in strains belonging to polyphyletic *formae speciales*, pointing to a combination of vertical and horizontal inheritance of these genes.
A MUSCLE alignment was made with the nucleotide sequence of **(A)** *SIX2* (705 nt), **(B)** *SIX4* (793 nt), **(C)** *SIX8* (525 nt), **(D)** *SIX9* (357 nt), **(E)** *SIX11* (335 nt), **(F)** *SIX14* (317 nt). Phylogeny was inferred using PhyML with 100 bootstrap iterations and plotted with mid-point rooting. Branches with most parsimonious bootstrap partitions below 50% were collapsed; values ≥ 50% and < 100% are indicated in red; 100% are not indicated. A coloured circle (wilting), square (root and shoot rot) or triangle (non-pathogenic / other) representing the isolate's *forma specialis* was plotted on the leaves of the dendrogram. (continued on next page ▶)



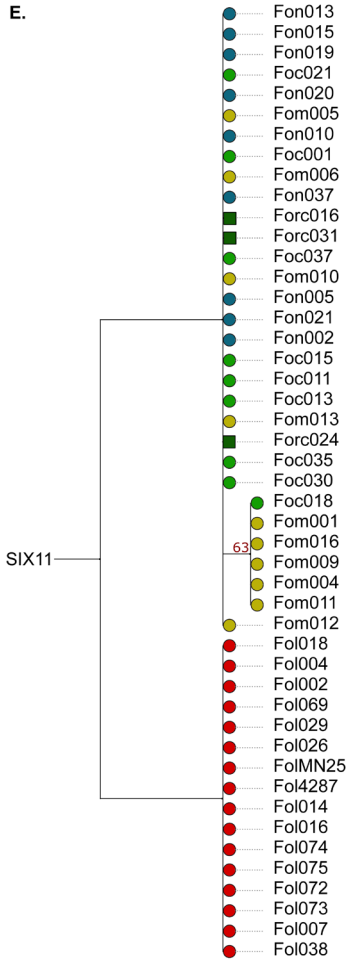


Fig. S5: Identical *SIX* gene sequence types are found in strains belonging to polyphyletic *formae speciales*, pointing to a combination of vertical and horizontal inheritance of these genes.
(continued from previous page)

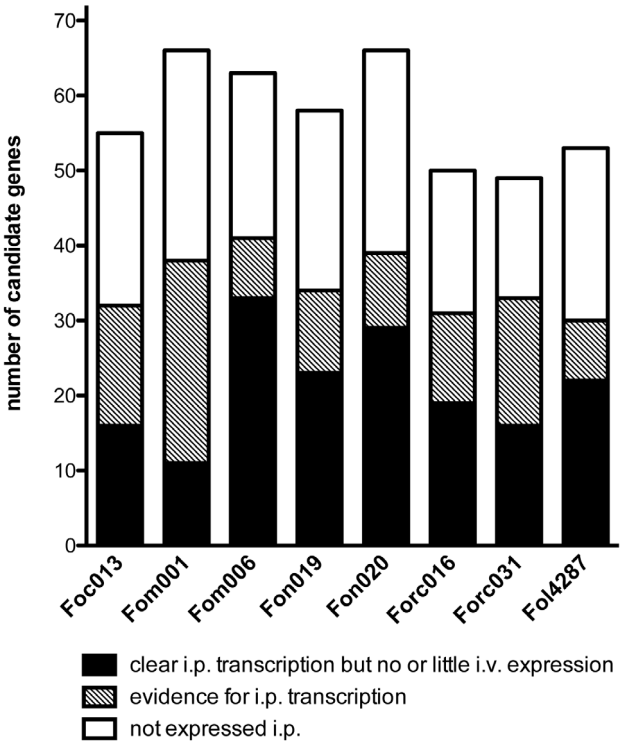


Fig. S6: RNA sequencing analysis of candidate effectors shows that many candidates show strong expression *in planta* (i.p.) and little or no expression during *in vitro* (i.v.) conditions.

Table S1. Statistics of *de novo* genome assemblies

Genome	Contigs or scaffolds	GC-content (%)	Average coverage (X)	L50	Longest sequence	Assembly size	Unknown bases (Ns)	Estimated genome size	Simple repeats and low complexity (%)	CEGMA completeness partial (%)	Nr ofcomplete *mimps* [a]	Nr of single TIRs	Genbankaccession
Foc001	1,329	47.32	84	339,401	1,018,461	51,402,461	114,042	61,439,107	0.63	98	42	50	MAKZ01000000
Foc011	1,136	47.60	96	398,155	2,000,702	48,668,302	110,980	55,459,228	0.62	98	41	36	MABT01000000
Foc013	1,131	47.59	86	410,395	2,001,134	48,900,174	102,823	56,216,292	0.61	96	38	44	MABJ01000000
Foc015	1,745	47.23	78	228,060	1,267,461	51,990,628	183,004	58,918,855	0.63	94	39	52	MABK01000000
Foc018	3,867	47.21	70	224,128	4,224,935	57,442,040	167,519	74,418,196	0.61	98	38	62	MABM01000000
Foc021	3,636	47.10	81	232,516	1,938,854	56,865,952	379,201	67,229,901	0.62	98	36	67	MABL01000000
Foc030	3,558	47.29	68	194,535	3,464,760	56,804,043	98,061	75,200,905	0.61	96	34	66	MABN01000000
Foc035	1,502	47.37	75	200,766	941,764	51,616,628	128,889	56,778,486	0.62	98	50	44	MABO01000000
Foc037	1,316	47.34	84	228,807	987,741	48,404,699	144,104	56,227,525	0.63	98	46	49	MABP01000000
Forc016	845	47.55	91	575,042	2,409,929	50,061,337	85,353	54,229,680	0.62	98	24	27	MABQ01000000
Forc024	826	47.56	109	467,406	2,476,658	49,353,483	75,983	52,995,981	0.62	98	26	29	MABR01000000
Forc031	881	47.56	100	662,141	3,357,292	49,347,207	92,603	52,993,821	0.61	98	25	31	MABS01000000
Fom004	1,302	47.42	122	607,560	2,769,734	57,732,739	2,892,117	59,723,502	0.57	98	34	62	MALX01000000
Fom005	2,472	47.45	82	212,447	983,939	54,498,897	97,561	61,915,135	0.61	98	35	58	MALY01000000
Fom006	2,225	47.50	91	179,765	1,364,775	52,761,534	56,853	59,659,447	0.62	98	32	53	MALZ01000000
Fom009	1,729	47.52	120	549,419	2,828,477	57,743,631	2,910,125	63,292,078	0.56	98	43	58	MAMA01000000
Fom010	3,384	47.75	86	130,950	753,059	53,674,587	144,941	65,057,011	0.60	98	33	66	MAMB01000000
Fom011	2,590	47.79	88	155,940	1,419,445	52,928,865	90,741	61,476,865	0.60	98	31	50	MAMC01000000
Fom012	2,283	47.69	86	190,689	1,022,864	52,601,880	112,928	57,466,117	0.61	98	30	57	MAMD01000000
Fom013	2,256	47.59	91	187,479	1,350,666	51,897,170	122,194	56,623,747	0.62	98	36	50	MAME01000000
Fom016	2,257	47.79	91	196,576	959,539	52,942,496	136,841	58,533,514	0.61	98	32	59	MAMF01000000
Fon002	2,192	47.41	97	219,468	2,247,237	51,981,004	84,780	63,510,600	0.63	98	33	33	MALA01000000
Fon005	3,514	47.41	107	171,813	988,797	54,725,717	81,875	76,222,549	0.61	98	30	90	MAKY01000000
Fon010	3,379	47.26	92	188,118	1,543,545	55,478,542	202,011	67,808,920	0.61	98	37	79	MALB01000000
Fon013	3,009	47.38	71	208,446	1,308,757	54,217,484	177,483	66,000,848	0.60	98	33	92	MALC01000000
Fon015	2,383	47.48	70	219,973	2,415,299	52,059,009	107,128	63,774,023	0.61	98	24	44	MALD01000000
Fon019	1,759	47.59	59	222,569	2,079,380	49,726,215	71,808	62,488,851	0.61	98	24	44	MAMH01000000
Fon020	3,606	47.33	104	185,956	1,132,243	55,134,009	135,587	67,113,570	0.62	98	48	82	MALE01000000
Fon021	3,084	47.33	106	187,895	989,948	54,806,891	135,085	65,917,299	0.61	98	28	91	MALG01000000
Fon037	1,738	47.24	77	268,898	2,513,951	51,296,626	156,256	62,356,805	0.63	96	31	40	MALF01000000
Fol002	3,451	47.82	180	158,838	878,799	51,731,767	13,748	54,799,659	0.61	98	27	60	MAMG01000000

Fol004	3,433	47.83	177	152,287	878,909	51,882,962	13,617	55,121,458	0.60	96	27	65	MALH01000000
Fol007	1,999	47.66	88	137,832	951,960	51,322,760	152,360	55,042,842	0.60	98	33	48	MALI01000000
Fol014	2,590	47.95	161	163,895	1,021,840	50,065,645	8,545	52,957,669	0.59	98	26	61	MALJ01000000
Fol016	2,340	48.33	85	182,560	855,033	49,577,770	9,846	52,916,179	0.53	98	24	56	MALM01000000
Fol018	2,141	48.00	216	176,036	1,012,475	48,953,637	8,523	51,220,636	0.60	98	20	47	MALL01000000
Fol026	3,302	47.79	196	165,262	858,320	51,805,189	13,305	55,181,438	0.60	98	25	64	MALK01000000
Fol029	2,270	47.78	189	181,501	1,184,915	50,011,956	9,185	52,275,182	0.60	98	25	56	MALN01000000
Fol038	3,188	47.83	217	158,709	835,465	51,297,012	13,246	53,666,231	0.59	98	28	59	MALO01000000
Fol069	1,888	47.75	197	329,247	1,253,049	49,123,713	5,957	50,447,870	0.61	98	25	52	MALP01000000
Fol072	1,859	47.81	192	323,703	2,228,049	50,095,049	7,826	52,323,815	0.61	98	23	59	MALQ01000000
Fol073	3,317	47.79	163	158,918	809,029	51,785,856	12,534	55,011,902	0.59	98	25	54	MALR01000000
Fol074	3,410	47.83	196	155,194	1,022,034	51,538,220	13,035	53,761,525	0.60	98	27	61	MALS01000000
Fol075	2,010	47.73	168	322,909	1,517,907	50,653,633	6,643	53,025,329	0.60	98	26	47	MALT01000000
FoMN14	1,978	47.87	185	311,933	1,535,963	48,997,843	5,724	50,408,134	0.60	98	4	8	MALU01000000
Fol4287	114	47.28	-	1,976,106	4,358,238	61,357,934	1,421,151		0.54	97	40	78	GCA_000149955.2
Fol4287 illum.	1,955	47.75	115	165,457	1,364,783	51,024,763	165,327		0.60	98	49	31	MALW01000000
Fol4287 ion pr.	4,290	47.85	72	75,432	464,937	51,156,164	-		0.59	98	21	67	MALV01000000
Fol MN25	388	47.59	-	2,091,760	4,599,047	48,637,398	166,737		0.62	97	29	41	GCA_000259975.2
Forl CL57	418	47.52	-	2,471,847	5,064,356	49,359,289	98,712		0.62	98	10	10	GCA_000260155.3
Fop HDV247	472	47.04	-	2,838,763	5,895,957	55,188,216	658,483		0.63	98	24	36	GCA_000260075.2
Focon PHW808	2,552	47.39	-	161,291	937,530	53,575,352	378,842		0.60	98	39	53	GCA_000260215.2
For PHW815	1,218	47.29	-	449,381	3,043,968	53,499,362	603,813		0.62	98	53	70	GCA_000260235.2
Fo5176	9,086	47.81	-	59,995	415,898	54,948,534	33		0.59	98	44	59	GCA_000222805.1
Fom001 NRRL26406	1,146	47.29	-	2,206,590	6,345,831	54,034,280	247,345		0.63	98	42	51	GCA_000260495.2
Fov NRRL25433	985	47.26	-	423,970	3,361,134	52,914,414	455,576		0.63	96	48	49	GCA_000260175.2
Focub II5	418	47.38	-	1,130,391	4,544,391	46,553,780	128,093		0.63	97	11	19	GCA_000260195.2
Focub B2	840	44.37	-	1,947,538	6,307,302	52,926,277	4,118,234		0.59	98	2	7	GCA_000350365.1
Focub N2	1,341	47.28	-	653,140	3,690,067	47,657,417	747,135		0.59	97	0	3	GCA_000350345.1
FOSC-3a	168	46.93	-	4,457,292	5,466,126	47,906,303	697,973		0.62	99	4	8	GCA_000350365.1
Fo47	124	47.47	-	3,844,136	6,198,950	49,664,628	225,164		0.62	98	4	17	GCA_000271705.2
Fg PH1	31	48.04	-	5,350,016	8,931,406	36,446,046	222,405		0.69	97	0	0	GCA_000240135.1
Fv 7600	36	48.62	-	1,959,799	4,627,337	41,776,161	75,816		0.70	98	1	0	GCA_000149555.1

[a] Two complementary mimp TIRs identified within a 400 nt window.

Table S3. Comparison of candidate ORFs identified in three genome assemblies of the reference strain Fol4287 shows that each individual assembly yields a few unique candidate genes.

Genome assemblies were made from Illumina (three libraries of 170 bp paired-end, 500 bp paired-end and 1500 bp mate-pair combined) and IonProton read data and the output of our method on these assemblies was compared to that of the Fol4287 reference genome assembly (Broad). Each gene is represented by its signal peptide sequence in this table. When compared one-to-one, the Broad reference assembly yielded one completely unique candidate (SIX13) that was not picked up in either the Illumina or the Ion Proton assembly. The other SIX genes were found in all three assemblies. This comparison shows that different assemblies and thus sequence breakpoints can have some influence on which effectors are found by a genomic context identification approach. Unlikely effector candidates are in grey (see Table S2).

BROAD	ILLUMINA	IONPROTON	gene
MAPYSMVLLGALSILGFGAYA	MAPYSMVLLGALSILGFGAYA	MAPYSMVLLGALSILGFGAYA	*SIX1*
MFNHFCWLLASCLSSLARA		MFNHFCWLLASCLSSLARA	-
MFSKAIPISLLISTSHA	MFSKAIPISLLISTSHA	MFSKAIPISLLISTSHA	*SIX11*
MKLLAVVATALAVFSTAEA	MKLLAVVATALAVFSTAEA	MKLLAVVATALAVFSTAEA	*SIX9*
MHSLSLLSLLMFPADALC	MHSLSLLSLLMFLRTLYA	MHSLSLLSLLMFLRTLYA	-
MHTEYLFLLLIPMGAVS	MHTEYLFLLLIPMGAVS	MHTEYLFLLLIPMGAVS	*SIX14*
MHVQSLLLASGLVPFVAS		MHVQSLLLASGLVPFVAS	-
MIALYFLPFVLWACLGQA	MIALYFLPFVLWACLGQA		***metallopeptidase***
MINRCRYFRL	MINRCRYFRL	MINRCRYFRL	-
MKIFILIFSFIGLIAA	MKIFILIFSFIGLIAA	MKIFILIFSFIGLIAA	-
MKKQTSTGSLCILFSKKVILLSLSLL SKSSYA	MKKQTSTGSLCILFSKKVILLSLSLL SKSSYA	MKKQTSTGSLCILFSKKVILLSLSLL SKSSYA	***glycosyltransferase***
MKLALIASILAAGCVA	MKLALIASILAAGCVA	MKLALIASILAAGCVA	*SIX6*
MKLLALLALASPLVSA			*eg45-like domain containing protein*
MKLLWLIPVVAS	MKLLWLIPVVAS	MKLLWLIPVVAS	*SIX10*
MKLSALLGFAVCVPSTMA	MKLSALLGFAVCVPSTMA	MKLSALLGFAVCVPSTMA	*hypothetical protein FOXG_22828*
MKPTLVSLGSMVLLASA	MKPTLVSLGSMVLLASA	MKPTLVSLGSMVLLASA	*hypothetical protein FOTG_18813*
MLFKIAWVSLFTTWAISVAA	MLFKIAWVSLFTTWAISVAA	MLFKIAWVSLFTTWAISVAA	*SIX2*
	MKSILIYLGNMVLLASA		-
	MKTIGLAITLSFISCVSA	MKTIGLAITLSFISCVSA	*hypothetical protein FOWG_18246*
MLPISILVPLLLLGTTTR	MLPISILVPLLLLGTTTR	MLPISILVPLLLLGTTTR	*c6 transcription factor*
MPKLQALTLAALSCLHATTASA	MPKLQALTLAALSCLHATTASA	MPKLQALTLAALSCLHATTASA	***putative oxidoreductase***
MPSSKGYEASWPLILLLCIATATPGLS	MPSSKGYEASWPLILLLCIATATPGLS	MPSSKGYEASWPLILLLCIATATPGLS	*integral membrane protein*
MRFLLLIAMSMTWVCSIAG	MRFLLLIAMSMTWVCSIAG	MRFLLLIAMSMTWVCSIAG	*SIX3*
MRFLNIAIGVITLLQGADA	MRFLNIAIGVITLLQGADA		***conserved secreted protein FOXG_14254***
MTRFHLILLPLLFSWFSYCLG			*SIX13*
total: 23	**total: 21**	**total: 20**	

Table S4. Effector identification based on mimp-proximity is applicable only to F. oxysporum.
Running the candidate identification pipeline on all currently (December 2015) available non-Fo Fusarium genome assemblies from Genbank reveals only few mimp terminal inverted repeats (TIRs) and in total 14 (11 unique) candidate ORFs in 29 genomes. The low abundance of mimp TIRs in most other Fusarium species indicates that this method is not usable for species outside the FOSC.

Strain	Genbank accession	Complete *mimps* [a]	single TIRs	nr of candidates	nr of candidates Augustus
F. acuminatum CS5907	CBMG000000000.1	0	4	0	0
F. avenaceum Fa05001	JPYM00000000.1	1	2	1	1
F. avenaceum FaLH03	JQGD00000000.1	1	6	0	0
F. avenaceum FaLH27	JQGE00000000.1	1	4	0	0
F. circinata FSP 34	AYJV00000000.1	0	1	1	0
F. circinatum GL1327	JRVE00000000.1	0	0	0	0
F. culmorum CS7071	CBMH000000000.1	0	1	0	0
F. fujikuroi B14	ANFV00000000.1	0	0	0	0
F. fujikuroi FGSC 8932	JRVF00000000.1	0	18	4	0
F. fujikuroi KSU X-10626	JRVG00000000.1	0	0	0	0
F. fujikuroi KSU 3368	JRVH00000000.1	0	15	4	1
F. graminearum 233423	LAJZ00000000.1	0	0	0	0
F. graminearum 241165	LAKA00000000.1	0	0	0	0
F. graminearum CS3005	JATU00000000.1	0	1	0	0
F. graminearum PH1	AACM00000000.2	0	0	0	0
F. langsethiae Fl201059	JXCE00000000.1	1	2	0	0
F. nygamai MRC8546	LBNR00000000.1	2	7	2	0
F. pseudograminearum CS3096	AFNW00000000.1	0	3	0	0
F. pseudograminearum CS3220	CBMC000000000.1	0	3	0	0
F. pseudograminearum CS3270	JTGB00000000.1	0	3	0	0
F. pseudograminearum CS3427	CBMD000000000.1	0	3	0	0
F. pseudograminearum CS3487	CBME000000000.1	0	3	0	0
F. pseudograminearum CS5834	CBMF000000000.1	0	2	0	0
F. verticillioides 7600	AAIM00000000.2	1	0	0	0
F. virguliforme Mont 1	AEYB00000000.1	0	0	0	0
Fusarium. sp. CS3069	CBMI000000000.1	0	0	0	0
Fusarium. sp. JS1030	JWIW00000000.1	0	0	0	0
Fusarium. sp. JS626	JWIV00000000.1	0	0	0	0
N. haematococca mpVI77134	ACJF00000000.1	0	0	0	0

[a] Two complementary mimp TIRs identified within a 400 nt window.

Due to size restrictions, Supplementary Tables S2 and S5 can be found online via http://onlinelibrary.wiley.com/doi/10.1111/1462-2920.13445/abstract (files "emi13445-sup-0002-suppinfo2.xlsx" and "emi13445-sup-0003-suppinfo3.xlsx")

Chapter 3

A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of multiple cucurbit species

This chapter has been published (excluding Appendices) as:
Peter van Dam, Like Fokkens, Yu Ayukawa, Michelle van der Gragt, Anneliek ter Horst, Balázs Brankovics, Petra M. Houterman, Tsutomu Arie, Martijn Rep (2017) A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of multiple cucurbit species. *Scientific Reports* **7**: 9042. doi:10.1038/s41598-017-07995-y

A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of multiple cucurbit species

Abstract

The genome of *Fusarium oxysporum* (Fo) consists of a set of eleven 'core' chromosomes, shared by most strains and responsible for housekeeping, and one or several accessory chromosomes. We sequenced a strain of Fo f.sp. *radicis-cucumerinum* (Forc) using PacBio SMRT sequencing. All but one of the core chromosomes were assembled into single contigs, and a chromosome that shows all the hallmarks of a pathogenicity chromosome comprised two contigs. A central part of this chromosome contains all identified candidate effector genes, including homologs of *SIX6, SIX9, SIX11* and *SIX13*. We show that *SIX6* contributes to virulence of Forc. Through horizontal chromosome transfer (HCT) to a non-pathogenic strain, we also show that the accessory chromosome containing the *SIX* gene homologs is indeed a pathogenicity chromosome for cucurbit infection. Conversely, complete loss of virulence was observed in Forc016 strains that lost this chromosome. We conclude that also a non-wilt-inducing Fo pathogen relies on effector proteins for successful infection and that the Forc pathogenicity chromosome contains all the information necessary for causing root rot of cucurbits. Three out of nine HCT strains investigated have undergone large-scale chromosome alterations, reflecting the remarkable plasticity of Fo genomes.

Introduction

Fusarium oxysporum Schlechtend.: Fr. f.sp. *radicis-cucumerinum* Vakalounakis (Forc) is the causal agent of root and stem rot in cucurbits resulting in severe damage, particularly in greenhouse cucumber (*Cucumis sativus*) and muskmelon (*C. melo*). The disease was first described in Greece in 1989 by Vakalounakis, who identified the pathogen as a new *forma specialis* of *F. oxysporum* (Fo) (Vakalounakis, 1996). Forc has since been recorded in several other countries including Canada, France, Spain, China, Turkey and Israel (Punja and Parker, 2000; Moreno *et al.*, 2001; Vakalounakis *et al.*, 2004; Karaca and Kahveci, 2010; Cohen *et al.*, 2015). Unlike *Fusarium* wilt caused by Fo f.sp. *cucumerinum* (Foc) or Fo f.sp. *melonis* (Fom), the main symptoms caused by Forc are external rotting of the root and stem and profuse sporulation in the rotted tissue (Vakalounakis, 1996; Cohen *et al.*, 2015). Still, the infection mechanism appears to be the same: the fungus invades the roots and colonizes the xylem vessels of the plant (Video S1).

Formae speciales of Fo typically have a very narrow host range, often restricted to a single plant species (Armstrong and Armstrong, 1981; Kistler *et al.*, 1998). Forc is exceptional because its host range includes not only cucumber and melon, but also additional *Cucurbitaceae* species such as watermelon *(Citrullus lanatus),* squash *(Cucurbita pepo)* and gourd *(Luffa aegyptiaca)* (Vakalounakis, 1996; Punja and Parker, 2000; Cohen *et al.*, 2015). The genetic mechanism underlying the difference in disease symptoms (root rot versus wilt) as well as the extended host range of this pathogen are unknown, but may be associated with the suite of effector genes present in the genome of this *forma specialis* (van Dam *et al.*, 2016). We found that Forc strains possess four Secreted In Xylem *(SIX)* gene homologs: *SIX6, SIX9, SIX11* and *SIX13*, which encode small secreted proteins originally identified in tomato-infecting strains (Houterman *et al.*, 2007; Schmidt *et al.*, 2013; Gawehns *et al.*, 2014). Additionally, we found several other genes encoding candidate effectors based on small size, predicted secretion signal, and vicinity to a "miniature impala" (*mimp*) transposable element, including a secreted astacin-like metalloprotease (van Dam *et al.*, 2016).

The genome of Fo is typically divided into a set of eleven 'core' chromosomes, with sequences generally conserved in all *Fusarium* species, and responsible for housekeeping, and one or several transposon-rich and gene-poor 'accessory' chromosomes (Ma, 2014). In Fo f.sp. *lycopersici* (Fol), one of these accessory chromosomes was shown to be required for pathogenicity towards tomato (Vlaardingerbroek, Beerens, Schmidt, *et al.*, 2016). Moreover, it can be horizontally transferred to the non-pathogenic strain Fo47, thereby transforming this strain into a tomato pathogen (Ma *et al.*, 2010; Vlaardingerbroek, Beerens, Rose, *et al.*, 2016). In *de novo* Illumina assemblies, accessory chromosomes are typically dispersed over many contigs or scaffolds due to their high repeat-content, making it impossible to determine how many accessory chromosomes are present in a strain. Three Forc strains have been sequenced so far, each resulting in assemblies of several hundred scaffolds (van Dam *et al.*, 2016). A

solution to the high level of fragmentation of Fo assemblies could be long-read sequencing technology, such as PacBio Single Molecule Real-Time (SMRT) sequencing. This would allow the multiple kb-sized repetitive elements to be spanned by individual reads, leading to much larger contigs.
The aims of this study were to (i) determine the genome structure of Forc, (ii) investigate whether Forc, like wilt-causing strains of Fo, relies on effector proteins for successful colonization and (iii) identify which part(s) of the Forc genome are necessary for the root- and shoot-rot phenotype as well as the extended host range of Forc. To reach these aims, we applied SMRT sequencing of a representative strain of Forc (strain Forc016) as well as Fom (Fom001; NRRL26406) as a step towards answering the question what differentiates Forc from strains causing wilt.

Results

A corrected SMRT assembly of Forc contains 33 sequences including 12 (near) full-length chromosomes

In order to obtain a better understanding of the genome composition of Forc, an HGAP.3 de novo assembly was generated for Forc016, a strain previously sequenced by Illumina (van Dam *et al.*, 2016). The initial SMRT assembly consisted of 41 contigs, including seven contigs that contained ribosomal DNA (rDNA) repeats. Two of these show rDNA sequences at one end and telomeric repeats (CCCTAA) on the other end, indicating that they together form chromosome 2 (Ma, 2014). The rDNA copy number was estimated through Illumina read coverage (~98 copies), and the two contigs were joined to reconstruct chromosome 2 (N.B. numbering of core chromosomes follows the Fol4287 reference genome). Three rDNA repeats of each contig were kept. The 91 copies in between were filled with the first rDNA repeat of the first contig.
Chromosome 13 was also assembled into two contigs, but an overlap of 13,396 nucleotides and synteny to the SMRT assemblies of Fom001 (Fig. S1), as well as a related *Fusarium* species, *F. subglutinans*, were found (B. Brankovics, personal communication). This allowed us to merge these sequences into chromosome 13.
Contour-clamped homogeneous electric field (CHEF) electrophoretic karyotyping followed by Southern blotting and hybridization with a radioactive probe generated from a Fol-*SIX6* Polymerase Chain Reaction (PCR) product revealed that the *SIX6* sequence is present on a ~2.5 megabase (Mb)-sized chromosome in Forc016 (Fig. 1, Fig. S2). This chromosome is present in the SMRT assembly as two separate contigs (13 and 17) of which the ends have an overlap of 586 nucleotides. Comparison to the Fom001 SMRT assembly revealed that this chromosome is largely syntenic to contig 127 of Fom001, but with a large (1.448 Mb) inversion between inverted, highly similar regions of about 200 kb (Fig. S3). Either end of this region

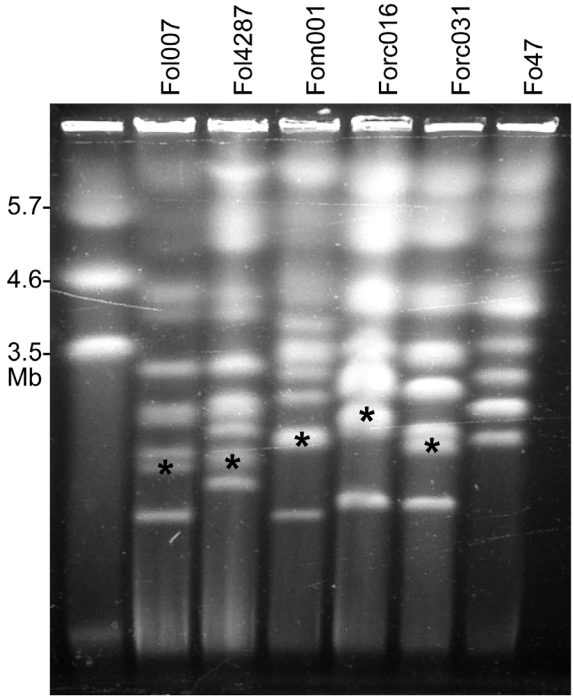


Fig. 1: Electrophoretic karyotypes of strains belonging to *F. oxysporum* f.sp. *lycopersici* (Fol007, Fol4287), *melonis* (Fom001) or *radicis-cucumerinum* (Forc016, Forc031) and non-pathogenic strain Fo47.
A black asterisk indicates the location of the radioactive Fol-*SIX6* probe hybridization signal, identifying the chromosomes potentially involved in pathogenicity. Fo47, a non-pathogenic strain, does not have a *SIX6* homolog. The left lane shows the karyotype of *Schizosaccharomyces pombe*, applied as a marker.

matched the end of Forc016 contig 17. However, when the 1.448 Mb region was manually inverted, not a single nucleotide polymorphism (SNP) was found in the pairwise sequence alignment, whereas in the original assembly three single nucleotide InDel mismatches were identified (data not shown). We therefore conclude that it is more likely that the 1.448 Mb region is in the reverse orientation and we adjusted this manually. One contig containing the mitochondrial DNA (mtDNA) sequence was identified by BLAST, removed from the SMRT assembly and the 47,541 nucleotide-long (annotated) mitogenome generated through Illumina reads by the GRAbB program (Brankovics *et al.*, 2016) was added (Brankovics *et al.*, submitted).

The final, manually corrected assembly of Forc016 is composed of eleven core chromosomes, one pathogenicity chromosome, twenty unpositioned sequences with a cumulative size of 2.572 Mb and the mitochondrial genome (Table 1). Nine of the unpositioned sequences end in telomeric repeats, indicating that they should probably be attached to the ends of chromosome-sized contigs that lack a telomere. The assembly is of a very high quality, with the L90 being reached with only eleven sequences.

Table 1: Comparison of the Forc016 genome assembly generated with Illumina HiSeq 2500 reads with the manually corrected SMRT HGAP.3 assembly and the raw SMRT HGAP.3 assembly of Fom001.

	Forc016 Illumina HiSeq	Forc016 SMRT	Fom001 SMRT
Assembly size	50,061,337	52,860,752	60,704,002
Ambiguous bases (Ns)	85,353	0	0
Gaps	902	0	0
GC (%)	47.63	47.69	47.65
Mean sequence length [a]	59,244	1,651,898	632,333
Shortest sequence length [a]	505	4587	5,714
Longest sequence length [a]	2,409,929	6,470,671	6,402,286
L50 (kbp)	575.0 (n=21)	4,490.1 (n=5)	4,357.5 (n=6)
L70 (kbp)	237.5 (n=49)	3,661.0 (n=8)	2,962.0 (n=10)
L90 (kbp)	37.7 (n=149)	2,466.0 (n=11)	761.3 (n=17)
Coverage	91x	72x	59x
# Sequences [a]	845	32 + mtDNA	96
# Sequences having telomeric repeats on both ends [a]	0	5	0
# Sequences having telomeric repeats on one end [a]	0	15	19
# Sequences having no telomeric repeats [a]	845	12	77

[a] 'Sequences' refers to scaffolds (Illumina assembly) or contigs (SMRT assembly)

Two large sequence duplications are present on contig 53 of the SMRT assembly (Fig. S4E). Because of its size, this contig is likely a large part of one of the small (±1-1.5Mb) accessory chromosome shown in Fig. 1. This is supported by the fact that it contains a GC-content drop typical of a centromeric region (Fig. S4B).

Comparison of the Forc016 SMRT assembly to that of Fol4287, the reference genome of *F. oxysporum*, revealed that (i) the eleven core chromosomes are highly syntenic between the strains (with 98.9% sequence identity), (ii) the Forc016 assembly has six contigs that contain sequences that align to known Fol accessory regions – likely due to the presence of similar transposable elements (TEs) in both and (iii) one of these six contigs is a putative pathogenicity chromosome on which the *SIX6* sequence was identified earlier (Fig. 1) with a high number of repeats and effector candidates, that we named chr^{RC} (Fig. 2). Gene ontology (GO) terms related to metabolism, protein ADP-ribosylation and DNA integrity were found to be overrepresented among the predicted genes on chr^{RC} (Fig. S5, Table S1). We further focused on this chromosome.

Most candidate effector genes reside in a subregion of chr^{RC}

The putative pathogenicity chromosome of Forc016, chr^{RC}, is highly similar (99.8%) to sequences present in the two other previously sequenced Forc strains (van Dam *et al.*, 2016). Surprisingly, high similarity (>99%) was also observed with sequences in the genomes of Fom001, Fom004, Fom005, Fom006, Fom012, Fom13, Fom016 but not Fom009, Fom010,

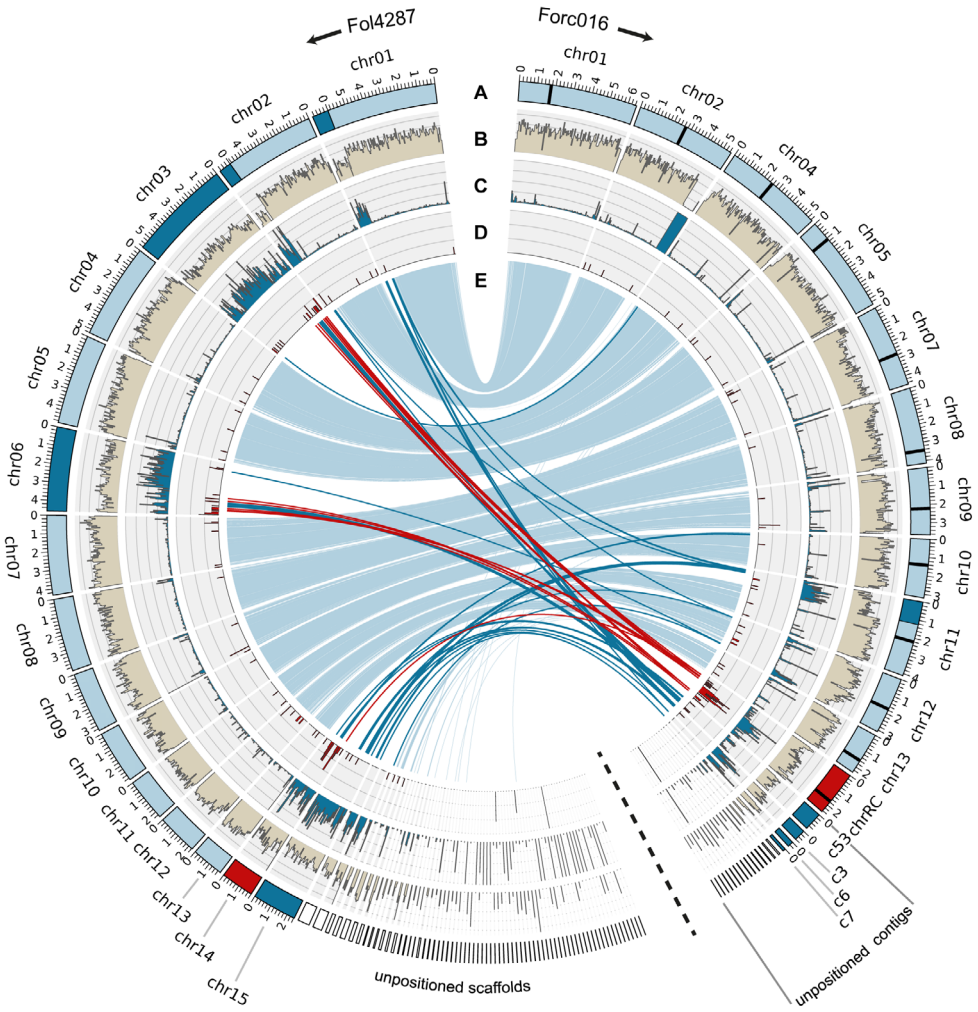


Fig. 2: The SMRT genome assembly of Forc016 includes eleven core chromosomes, several repeat-rich, gene-poor accessory regions and one chromosome enriched in candidate effector genes.
Comparison of the Forc016 assembly to that of Fol4287 reveals **(A)** eleven conserved core chromosomes (light blue), one putative pathogenicity chromosome (red) and several other accessory sequences (dark blue). Accessory regions typically have **(B)** low gene density and **(C)** high repeat density, both calculated here in 50 kb windows. The putative pathogenicity chromosome is marked by **(D)** the presence of many candidate effector genes. **(E)** indicates nucmer alignments with the Fol4287 reference assembly: in red alignments to the putative pathogenicity chromosome of Forc016, in dark blue alignments from known accessory regions in Fol4287 (chr1B; chr2B; chr3; chr6; chr14; chr15) and in light blue the remaining alignments, mostly between core regions in both genomes.

Fom011 (data not shown). A *de novo* HGAP.3 assembly for Fom001 was generated and we found that synteny is mostly preserved between chr^{RC} and Fom001 contig 127 (Fig. 3A). A notable exception is a central, ~700 kb region of the chromosome, which is exceptionally repeat-rich and of which ~300 kb is absent in Fom001 contig 127, flanked by several multi-kb inversions (Fig. 3B). Within the 700kb region, 195 genes were predicted in the Forc016 SMRT

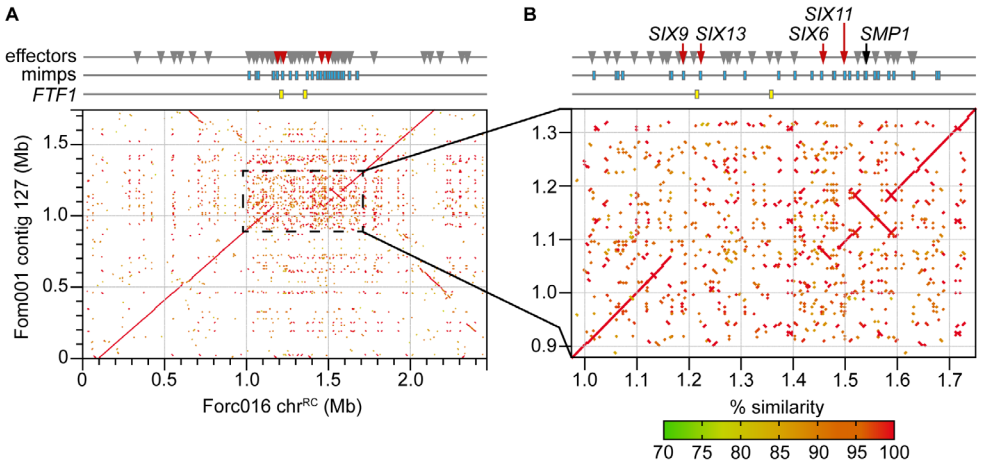


Fig. 3: Comparison of chr^{RC} with Fom001 contig 127 reveals a highly dynamic central region containing the majority of the miniature impala (*mimp*) TEs and candidate effector genes of the Forc genome.
(A) The chromosomes of the Fom and Forc strains are mostly syntenic, with large stretches showing 100% identity in this nucmer alignment. **(B)** The dynamic central region of the chromosome, about 700 kb in size, has the highest repeat density, 30 of the 35 *mimps* of the Forc genome and the majority of the Forc candidate effector genes, including *SIX9*, *13*, *6* and *11* and *SMP1*. Additionally, the two *FTF1* homologs present in the Forc genome are found here.

assembly. 185 of these have a highly similar homolog in Fom001 (average nucleotide sequence identity is 99.1%). 135 of these genes (including *SIX6* and *SIX11*) are almost identical between the two strains, returning a BLAST hit percentage of 99.8% or higher, suggesting that they have been reshuffled recently. Only ten genes could not be identified with BLASTN (e-value < 1e-20; perc_identity > 90%; query coverage > 70%) in Fom001: *SIX9* (g15834), three beta-lactamases (g15883, g15832, g15833), three hypothetical proteins (g15854, g15957, g15835), a cytochrome p450 (g15902), a putative lysine decarboxylase (g15903) and an NADH-flavin oxidoreductase (g15826). One or several of these genes may contribute to the ability of Forc to cause root rot in several cucurbit species.

Interestingly, this is exactly the region where 30 of the 35 full-size miniature impala (*mimp*) elements were found in the Forc genome. Two other *mimps* were found on chr12, one on chr11 and two more on contig 14, a contig of only 22.4 kb. *Mimps* are contextually associated with effector genes in *F. oxysporum* (Schmidt *et al.*, 2013, 2016). Indeed, the majority (51 out of 98) of candidate effector genes identified by BLAST from the list of 104 candidates that we identified earlier (van Dam *et al.*, 2016) are localized in this region (Fig. 3B). Among these are the four *SIX* homologs that are present in Forc: *SIX6* (g15909), *SIX9* (g15834), *SIX11* (g16807) and *SIX13* (g15844). These were previously shown to be expressed during plant infection (van Dam *et al.*, 2016). Additionally, two homologs of the *FTF1* transcription factor, associated with effector gene expression (Niño-Sánchez *et al.*, 2016; van der Does *et al.*, 2016), are found here (g15884 and g15843).

In Fom001, homologs of *SIX1* (contig 22), *SIX6* (contig 127), *SIX11* (two copies; contig 10

and 127) and *SIX13* (contig 10) are present. Fom contigs 10 and 22 are 2.962 and 1.268 Mb in size, respectively, suggesting that Fom001 may have more than one chromosome associated with pathogenicity. The sequences of *SIX6* and *SIX11* that are located on Fom contig 127 are identical between Forc016 and Fom001. From the list of candidate Fom effectors in Schmidt *et al.* (2016), only candidate 1A is present in the Forc016 assembly on chr^{RC} (100% identical between Fom and Forc). None of the other candidates, including *AVRFOM2,* is present in Forc.

Six6 contributes to virulence of Forc

In order to investigate the role candidate effectors play in Forc pathogenicity, knockout strains were generated through homologous recombination with a hygromycin resistance marker. Although this process is very inefficient in regions with many repeats, such as the region depicted in Fig. 4B, successful knockout was achieved for *SIX6, SIX9* and an astacin-like Secreted MetalloProtease gene (which we named *SMP1* (g15931); Fig. 3). All three genes are single copy in the Forc016 genome. Bioassays were conducted with cucumber, melon and watermelon plants to evaluate whether the fungus had become less pathogenic to one or several host plants upon loss of these genes.

When inoculated with 10^6 spores/ml at an ambient temperature of 25°C, the three independent *SIX6* knockout strains caused reduced disease symptoms in cucumber compared to a transformant with an ectopic integration of the T-DNA, as well as to the wild type Forc016 strain and the two other knockout strains (Fig. 4; pictures shown in Fig. S6). When tested at a lower ambient temperature of 18-20°C, all strains caused quick and uniform death to all plants of the three tested species, indicating that these conditions are more favourable for Forc disease development and do not require Six6 (data not shown). Disruption of *SIX9* or *SMP1* did not significantly affect virulence under the tested conditions (Fig. 4). The absence of *SIX9* in Fom001 is therefore not responsible for the phenotypic difference between Fom and Forc.

Cell wall degrading enzymes in rot symptom development

In a comparison between Forc and Foc strains, it was found that isolates of Forc have more pectolytic enzyme activity (Punja and Parker, 2000). This may, in part, account for the crown rot and tissue maceration seen in root and shoot rot disease caused by Forc. Production of cell wall degrading enzymes (CWDEs) by Fo is well documented (reviewed in Di Pietro *et al.* (2003); Michielse and Rep (2009)) and was shown to be positively correlated with virulence in Fo f.sp. *dianthi* (Baayen *et al.*, 1997). Individual knockout of CWDE- or protease-encoding genes, however, usually does not result in a detectable reduction in virulence in Fo (Di Pietro *et al.*, 2003; Michielse and Rep, 2009). Site-directed mutagenesis of three amino acid residues located at the putative active site of an endopolygalacturonase from *F. verticillioides* (formerly

F. moniliforme) did result in reduced macerating activity on potato medullary tissues (Caprari *et al.*, 1996). This led Reignault *et al.* to hypothesize that pectinases are important for necrosis and maceration (e.g. by Forc), but are less important for vascular wilt disease (Reignault *et al.*, 2008).

In total, 179 gene products are predicted to have proteolytic activity (ontology term GO:0006508) in the Forc genome, of which four are encoded on chr^{RC}. Two of these possess a predicted signal peptide: *SMP1* and a subtilase gene. Polygalacturonase activity (GO:0004650) was predicted for 11 genes, none of which resides on chr^{RC}. Likewise, none of six pectinesterase-encoding genes in the Forc genome (GO:0030599) resides on chr^{RC}. Since knockout of *SMP1* did not result in reduced virulence, there may be functional redundancy with other proteases. Despite these observations, protease or CWDE activity may still be important during plant colonization and rot symptom development.

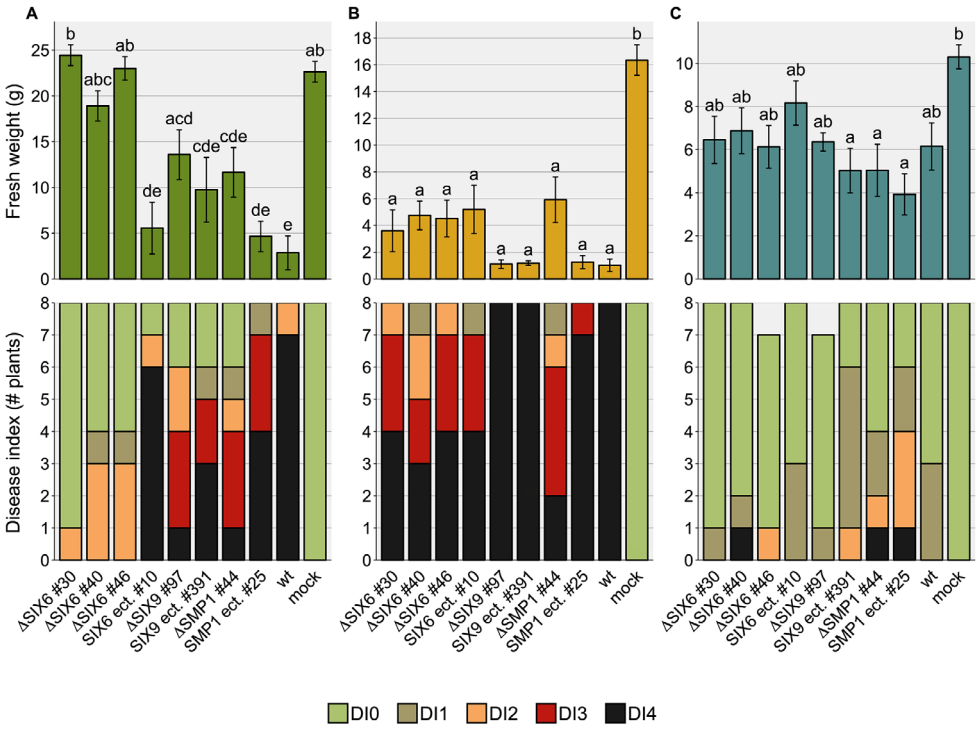


Fig. 4: Three independent *SIX6* deletion strains cause less symptom development in cucumber.
Fresh weight (± S.E.) and disease index (DI) of **(A)** cucumber, **(B)** melon and **(C)** watermelon plants were scored 14 days post inoculation. An ANOVA followed by a Tukey HSD test ($p<0.05$) was performed to determine significance of the differences in the fresh weight measurements (significance categories shown with letters above the bars). Under the tested conditions (10^6 spores/ml, 25°C), three independent *SIX6* deletion strains caused reduced symptoms in cucumber compared to wildtype (wt) and an ectopic transformant (*SIX6* ect. #10). Knockout of the two other candidate virulence genes *SIX9* and *SMP1* did not have a significant effect on virulence. Under these conditions (25°C), Forc causes only mild symptoms on watermelon.

Chr^{RC} is a mobile chromosome

Chr^{RC} is similar to the mobile Fol pathogenicity chromosome (Ma *et al.*, 2010) in that it is repeat-rich, gene-poor and contains most candidate effector genes, of which at least one *(SIX6)* contributes to virulence towards cucumber. In order to assess whether this chromosome could be horizontally transferred to other strains, a co-cultivation experiment was performed. Forc016Δ*SIX6*#46 was chosen as the potential chromosome donor strain, since it carries the *HPH* hygromycin-resistance marker on chr^{RC}. Spores from this strain were mixed with spores from three different 'recipient' strains: Fo47, Fol4287 and Fom001, all tagged by random insertion of the *BLE* zeocin-resistance gene. Double-resistant colonies were recovered only in the combination with Fo47. Nine such strains were saved and used for further analysis. All were shown by PCR to contain both *HPH* and *BLE* genes.

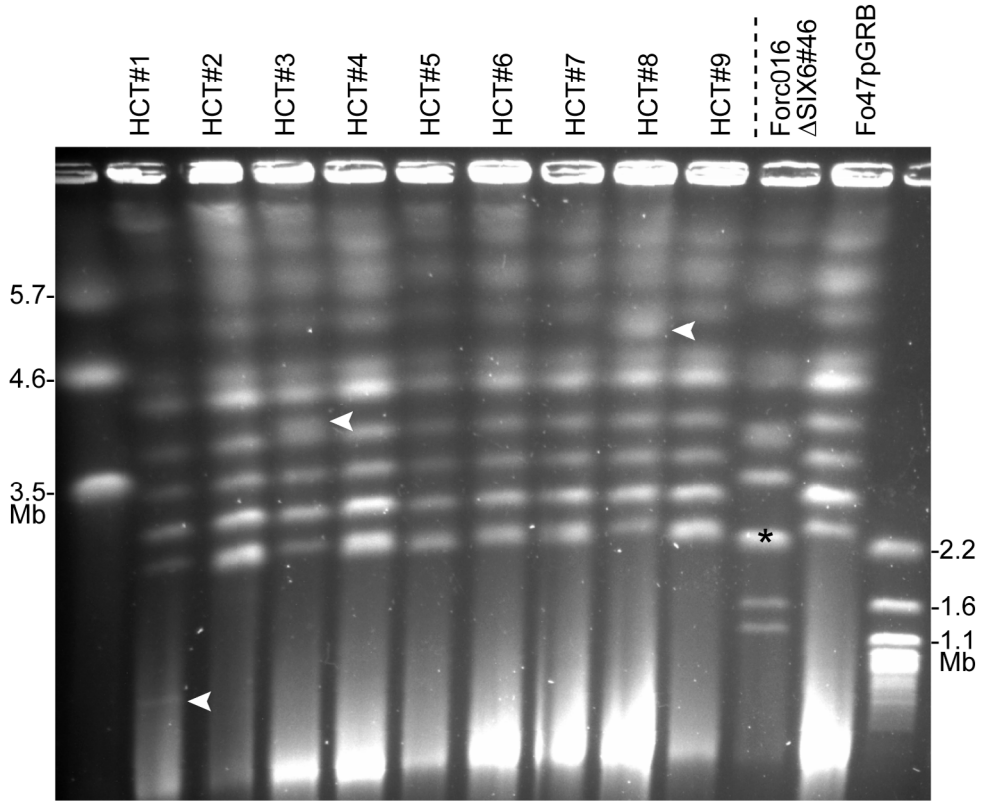
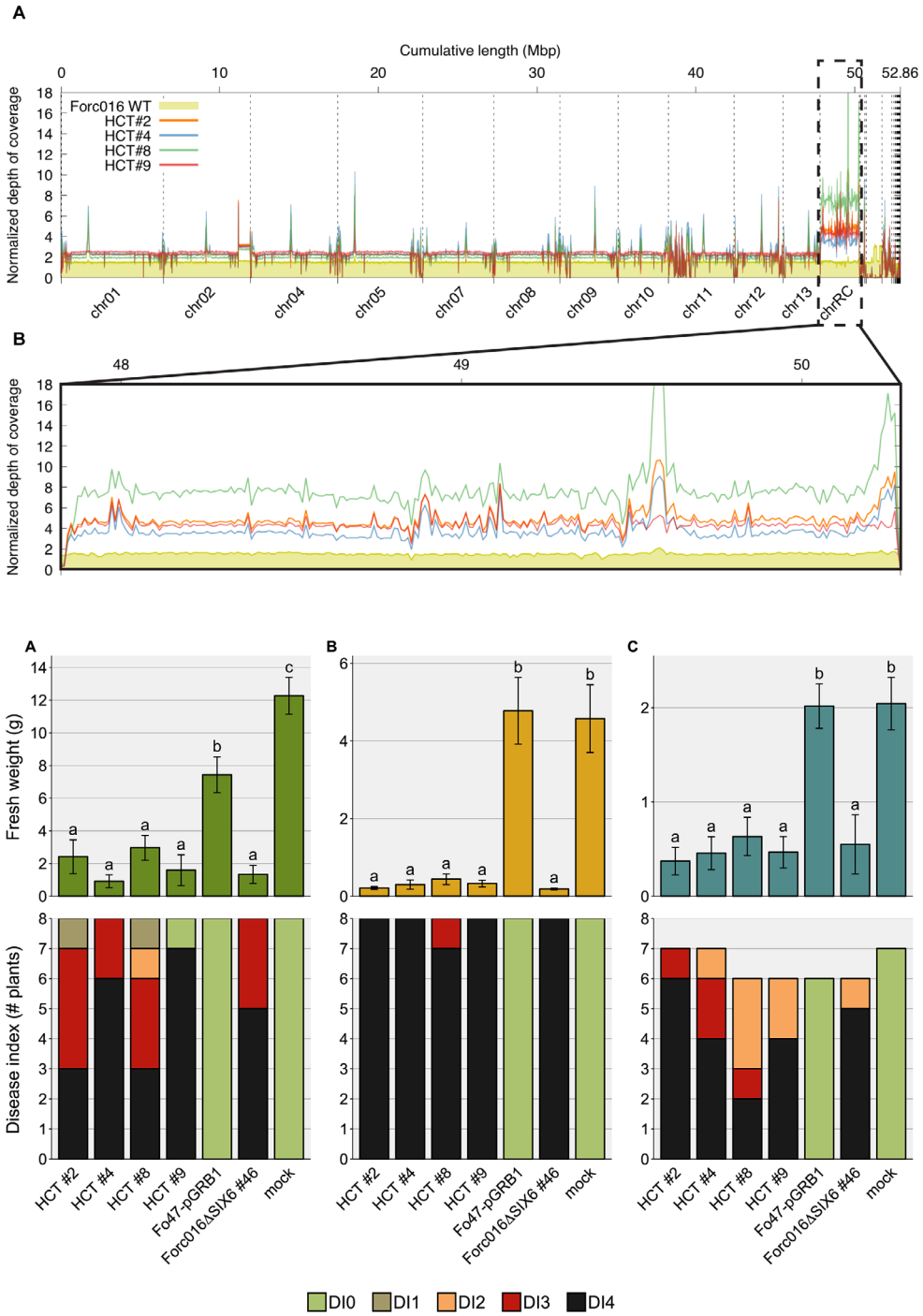


Fig. 5: Nine strains derived from a HCT experiment between Forc016Δ*SIX6*#46 and Fo47pGRB contain new chromosomes in the Fo47 background.
Lanes 1 – 9 show the karyotype of HCT-derived strains, resembling that of Fo47pGRB (lane 11). Most of these strains have a double band at the size of chr^{RC} (~2.5Mb), marked in Forc016Δ*SIX6*#46 with a black asterisk (lane 10). Strains #1, #3 and #8 do not have this double band, but instead have at least one novel chromosome that is not found in either parental strain (white arrowheads). The left and right lanes show the karyotypes of *S. pombe* and *S. cerevisiae,* respectively, applied as markers.


A
Cumulative length (Mbp)
0
10
20
30
40
50
52.86
Normalized depth of coverage
Forc016 WT
HCT#2
HCT#4
HCT#8
HCT#9
chr01
chr02
chr04
chr05
chr07
chr08
chr09
chr10
chr11
chr12
chr13
chrRC
B
48
49
50
Normalized depth of coverage
A
B
C
Fresh weight (g)
Disease index (# plants)
HCT #2
HCT #4
HCT #8
HCT #9
Fo47-pGRB1
Forc016ΔSIX6 #46
mock
DI0
DI1
DI2
DI3
DI4

◀ Fig. 6: Normalized Illumina read mapping to the SMRT assembly of Forc016 confirms horizontal transfer of chr^{RC} in a Fo47 background.
(A) Reads mapped more abundantly to the transferred chr^{RC} sequence than the rest of the assembly. (B) HCT strains #2, #4, #9 showed a relative (compared to total # mapped reads) depth of coverage of ~4 on chr^{RC}, whereas the relative coverage of HCT#8 sequences was ~8 along the entire chromosome. This indicates a chromosomal duplication, in accordance with the ~5Mb-sized band in the CHEF gel in Fig. 5.

▶ Fig. 7: Horizontal transfer of chr^{RC} transforms the non-pathogenic strain Fo47 into a root and shoot rot pathogen of cucurbits.
Fresh weight (± S.E.) and disease index (DI) of **(A)** cucumber, **(B)** melon and **(C)** watermelon plants were scored 14 days post inoculation. An ANOVA followed by a Tukey HSD test ($p<0.05$) was performed to determine the significance of differences in the fresh weight measurements (significance categories shown with letters above the bars). Under the tested conditions (10^7 sp/ml, 20°C), HCT strains #2, #4, #8 and #9 caused a similar level of disease severity in all three cucurbit species as the chr^{RC} donor strain (Forc016*ΔSIX6*#46).

To assess whether indeed chr^{RC} from Forc016 had been transferred to Fo47, a CHEF gel was run (Fig. 5). This revealed that all nine double-resistant strains displayed the karyotype of Fo47, with an additional chromosome presumably resulting from horizontal chromosome transfer (HCT). In the cases of HCT-derived strains #2, #4, #5, #6, #7 and #9 this chromosome is similar in size to a chromosome in the Forc016 donor strain (~2.5 Mb). However, since it is roughly the same size as the smallest chromosome of Fo47, the two co-migrated through the gel, resulting in a band with double intensity. In the three other cases (HCT #1, #3 and #8), this double band was absent and instead other new chromosomes were observed (Fig. 5, white arrowheads).

Stringent Illumina short-read mapping of HCT strains #2, #4, #8 and #9 on the Forc016 assembly showed that these strains indeed contain the full chr^{RC} sequence but no other Forc016-derived sequences (the core genomes of Forc016 and Fo47 have an average SNP density of 0.4% (more towards the telomeres), reducing mapping of reads to ~85%) (Fig. 6). This confirms that chr^{RC} now resided in a Fo47 core genome background. Remarkably, HCT#8 showed a relative depth of coverage of chr^{RC} about two times higher than the other HCT strains, indicating that the chr^{RC} sequences are present twice. Since a double band at the expected size of chr^{RC} (~2.5 Mb) is missing in this strain but a double-sized band of ~5 Mb is visible (Fig. 5), this duplication appears to have resulted in a single chromosome twice the size of chr^{RC}. From the normalized read coverage of HCT strains #1 and #3 it is clear that large-scale rearrangements also took place along chr^{RC} in these strains (Fig. S7). These strains were found to be less pathogenic towards cucurbits (Appendix A).

Assembly of the mitochondrial genome with GRABb (Brankovics *et al.*, 2016) and comparison to the mitogenomes of Fo47 and Forc016 showed that the mitochondrial DNA of the Fo47 acceptor strain had been retained in all cases (data not shown).

Chr^{RC} is capable of turning Fo47 into a cucurbit root rot pathogen

To assess whether the HCT-strains, carrying chr^{RC} in a Fo47 background, are pathogenic on cucurbits, a bioassay was performed on cucumber, melon and watermelon plants. This time, the assay was done under conditions ideal for Forc infection, with a relatively low ambient temperature of 18-20°C. All four tested strains (#2, #4, #8, #9) caused abundant symptom development in each of the three plant species, comparable to the control strain, Forc016*ΔSIX6*#46 (Fig. 7; Fig. S8). This shows that the biocontrol strain Fo47 can be transformed into a *radicis-cucumerinum* strain, capable of infecting multiple host plants and causing root and shoot rot, by a single chromosome of Forc (Fig. S8).

Chr^{RC} and the two smaller accessory chromosomes are conditionally dispensable

Incubation of the Forc016*ΔSIX9* strain (harboring the *HPH* hygromycin resistance gene on chr^{RC}) in medium containing 12.5 µg/ml benomyl yielded five strains that had become hygromycin sensitive, indicating a loss of the genomic region containing the *HPH* gene. Electrophoretic karyotyping showed that in all five cases, chr^{RC} had been completely lost (Fig. 8, Fig. S9). Chromosome loss strain #2 had additionally lost the two smaller accessory chromosomes of ~1.1 and 1.6 Mb. When tested in a bioassay, none of these strains caused symptoms in any of the three tested host plants (Fig. 9, Fig. S10). No growth rate alteration compared to their parent strain was found when the strains were grown on CDA or PDA

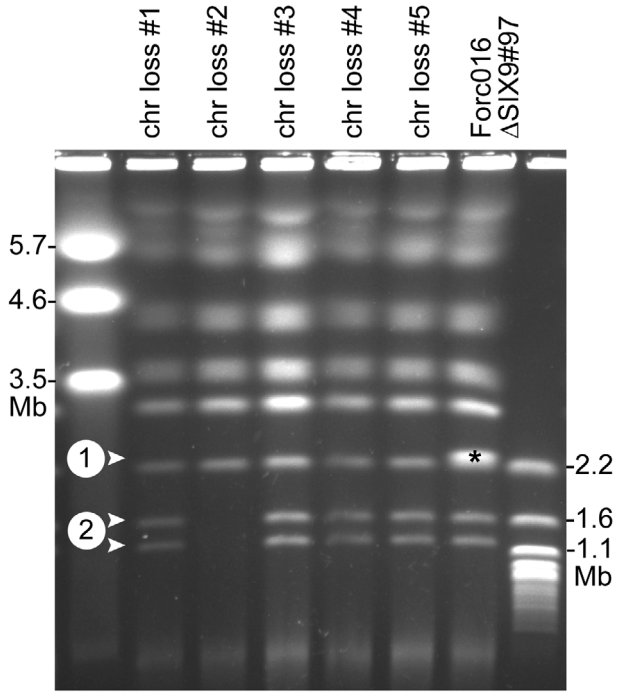


Fig. 8: Chr^{RC} and the two smaller accessory chromosomes of Forc016 are conditionally dispensable. Electrophoretic karyotyping shows complete absence of chr^{RC} (marked with a black asterisk in lane 6) in all five hygromycin sensitive strains (arrowhead 1, lanes 1-5). Additionally, chromosome loss strain #2 displays absence of the two smallest chromosomes (arrowheads 2). The left and right lanes show the karyotypes of *S. pombe* and *S. cerevisiae*, respectively, applied as markers.

medium. This demonstrates that chr^{RC} is required for pathogenicity of Forc016 towards multiple cucurbit species.

Sge1 is required for pathogenicity of Fom and Forc

The transcription factor *SIX Gene Expression 1* (Sge1) is required for pathogenicity in the Fol-tomato pathosystem, because Sge1 regulates effector gene transcription level (van der Does *et al.*, 2016, Michielse *et al.*, 2009). To uncover if this protein has a similar function in Forc and Fom, we made gene disruption mutants in Forc016 and Fom001 using the construct from Michielse *et al.* (2009). The knockout strains (ΔSGE1) resulted in clearly reduced symptom development when tested on cucumber, melon and watermelon (Appendix B). The knockouts of Fom and Forc were complemented with the Fol-*SGE1* gene (99% identical to the Fom/Forc copies), which resulted in 2/3 Forc (Appendix B Fig. A, B, D) and 1/1 Fom (Appendix B Fig. C) complementation strains regaining their pathogenicity.

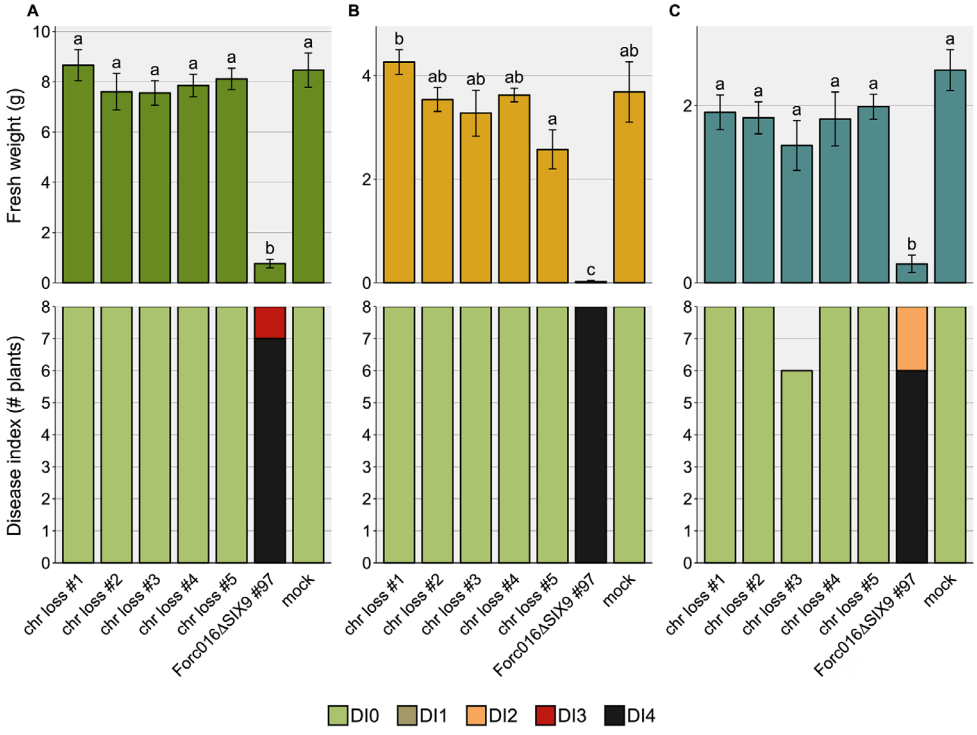


Fig. 9: Forc016 strains without chr^{RC} are completely avirulent.
Fresh weight (± S.E.) and disease index (DI) of **(A)** cucumber, **(B)** melon and **(C)** watermelon plants were scored 14 days post inoculation. An ANOVA followed by a Tukey HSD test ($p<0.05$) was performed to determine the significance of differences in the fresh weight measurements (significance categories shown with letters above the bars). Under the tested conditions (10^7 spores/ml, 20°C), none of the chromosome loss strains were able to cause disease symptoms in cucurbit plants, while their parent strain (Forc016Δ*SIX9*#97) was.

Discussion

We show that Forc016 has 11 core chromosomes, one horizontally transferrable 2.446 Mb pathogenicity chromosome (chr^{RC}) and two smaller accessory chromosomes. We conclude that chr^{RC} is necessary and sufficient for the root- and shoot-rot phenotype in several cucurbit species when infected by Forc. More specifically, the middle region of chr^{RC}, which differs from Fom001's homologous chromosome, might play a decisive role in both the extended host range and rotting symptoms caused by Forc.

Are effectors important for Forc pathogenicity?

Since the identification of *SIX1*, the first effector gene from *F. oxysporum* f.sp. *lycopersici* (Rep *et al.*, 2004), thirteen more *SIX* genes have been described (Takken and Rep, 2010; Schmidt *et al.*, 2013). For several of these, including *SIX6*, a role in virulence has been shown (Rep *et al.*, 2004; Houterman *et al.*, 2009; Thatcher *et al.*, 2012; Ma *et al.*, 2013; Gawehns *et al.*, 2014; Ma *et al.*, 2015). *SIX6* homologs have been found in Fo f.sp. *lycopersici, cucumerinum, radicis-cucumerinum, melonis, niveum, pisi, passiflorae, cubense* and *vasinfectum* (Chakrabarti *et al.*, 2011; Guo *et al.*, 2014; van Dam *et al.*, 2016; Williams *et al.*, 2016), as well as in Fo f.sp. *momordicae, luffae, Fusarium hostae* (van Dam and Rep, 2017) and in *Colletotrichum* spp. (Kleemann *et al.*, 2012). Strains belonging to the same *forma specialis* typically have the same sequence types for effector genes, even when core genes are not identical in sequence (van Dam *et al.*, 2016). This is likely the result of horizontal inheritance of dispensable genomic regions (Williams *et al.*, 2016) and is corroborated by the incongruent phylogeny of *SIX* genes compared to the housekeeping gene *EF1α* reported by Rocha *et al.* (2015). Interestingly, Forc shares its *SIX6* sequence type with strains belonging to Fo f.sp. *melonis* (Fom) as well as some – but not all – Fo f. sp. *niveum* (Fon) strains (van Dam *et al.*, 2016). This may be indicative of a (partially) shared ancestry of cucurbit-infection between these *formae speciales*.

Deletion of *SIX6* in Fol marginally compromises virulence in the Fol-tomato pathosystem. Additionally, Six6 suppresses I-2-mediated cell death upon transient expression in *N. benthamiana*, but does not compromise the activity of other cell-death-inducing genes (Gawehns *et al.*, 2014). Three individual Fol strains with a partial deletion of chromosome 14, thereby losing *SIX6, SIX9* and *SIX11*, as well as *ORX1* encoding an in xylem-secreted oxidoreductase, did not show a significant reduction in disease severity, indicating that these genes are largely dispensable for Fol pathogenicity (Vlaardingerbroek, Beerens, Schmidt, *et al.*, 2016). In the Fon-watermelon pathosystem, however, Six6 has been reported to be involved in virulence (Niu *et al.*, 2016).

We find a clear reduction in virulence of three independent Forc strains in which the *SIX6* locus was disrupted (Fig. 4). However, this phenotype could only be observed at relatively high ambient temperatures (25°C) in cucumber, while at lower temperatures all plants died.

In contrast to most wilt-causing Fo pathogens like Fol, Foc, Fom and Fon, Forc symptoms develop most efficiently at temperatures below 20°C (Vakalounakis and Fragkiadakis, 1999; Punja and Parker, 2000; Tok and Kurt, 2010), particularly during seedling infection, when plants may be under physiological stress (Punja and Parker, 2000). We conclude that Six6 contributes to virulence only under non-optimal conditions and only in cucumber.
The other tested effector candidate knockout strains (Δ*SIX9* and Δ*SMP1*) did not display a reduction in virulence towards cucumber, melon or watermelon compared to Forc-wt and ectopic transformant strains. Six9 and Smp1 are therefore, by themselves, not important for disease development caused by Forc.

HCT of chr^{RC} contributes to genome evolution in Fo

HCT has so far been described for Fol chromosomes 7, 8, 14 and the smallest chromosome of Fol007 (Ma *et al.*, 2010; Vlaardingerbroek, Beerens, Rose, *et al.*, 2016), but was until now not shown for other *formae speciales* of Fo. Non-pathogenic recipient strain Fo47 became pathogenic towards tomato upon receiving Fol chromosome 14, albeit less than the Fol donor strain. A higher aggressiveness of HCT-strains was observed when another accessory chromosome co-migrated, potentially due to the influence of transcription factors located on that chromosome (van der Does *et al.*, 2016). Interestingly, two copies of *FTF1,* a transcription factor associated with effector gene expression (Ramos *et al.*, 2007; Niño-Sánchez *et al.*, 2016; van der Does *et al.*, 2016), are located in the effector-rich central part of chr^{RC} (Fig. 3), potentially indicating a partial transcriptional autonomy of chr^{RC}.
Horizontal transfer of chr^{RC} was accomplished with Forc016Δ*SIX6* as a donor and Fo47 as a recipient strain. Nine double-drug resistant colonies were recovered after co-cultivation and electrophoretic karyotyping of these HCT strains (Fig. 5) showed that six strains gained chr^{RC} while three strains (#1, #3, #8) had undergone chromosome rearrangements. HCT strain #8 had a double relative coverage of chr^{RC} compared to that of the other strains and a band at twice the size of chr^{RC} (~5 Mb, Fig. 5) in its electrophoretic karyotype. The chromosome apparently duplicated but remained present as a single entity, pointing to a high level of genome plasticity.
A recent study by Vlaardingerbroek, Beerens, Rose, *et al.* (2016) also showed chromosomal plasticity in horizontal transfer experiments of the Fol pathogenicity chromosome. Transformation for marker insertion on this chromosome resulted in a larger (estimated 250 kb) pathogenicity chromosome in a donor strain that was used for HCT towards Fo47. Selection for loss of this chromosome in another study (Vlaardingerbroek, Beerens, Schmidt, *et al.*, 2016) resulted in several strains that only partially lost the chromosome. Interestingly, deletions within a chromosome and chromosomal breaks appeared to happen non-randomly at so-called 'deletion hotspots'. It is clear that genomes of *F. oxysporum*, particularly the accessory parts defining host virulence, are highly plastic. This could result in accelerated

genetic diversification, possibly facilitating adaptation to new environments including new host plants.

In Forc, two vegetative compatibility groups (VCGs) have been described: VCG0260 (to which Forc016 belongs) and VCG0261 (Katan, 1999). RAPD fingerprinting analyses and concatenated sequence alignment of 1195 conserved core genes showed that the two VCGs are very similar and appear to be clonally related (Vakalounakis and Fragkiadakis, 1999; van Dam *et al.*, 2016). The other two sequenced Forc strains, Forc031 (VCG0261) and Forc024 (VCG0260), both possess the chr^{RC} sequence and their effector gene content is nearly identical (van Dam *et al.*, 2016). Interestingly, large parts of chr^{RC} were also identified in two out of three sequenced Fom VCGs, including Fom001 (VCG0136, Fig. 3), whose core genome is highly similar to that of the Fol4287 reference strain and other Fol strains in VCG0030 (van Dam *et al.*, 2016). This is a strong indication that the Forc and Fom pathogenicity chromosomes evolved from a shared ancestor. Integration of the highly diverse central region in an ancestral chromosome from an unknown source potentially gave rise to chr^{RC} and the new *forma specialis radicis-cucumerinum*. The suite of candidate effector genes found in Forc, concentrated in the central region of chr^{RC} (Fig. 3), is most similar to that of strains belonging to Fo f.sp. *cucumerinum* (van Dam *et al.*, 2016). Systematic comparative and functional analysis of the accessory genomic regions of multiple cucurbit-infecting *formae speciales* will be necessary to reconstruct the evolutionary paths that led to host-specificity of Fo towards this plant family.

The wider host range of Forc compared to Fom could be caused by the absence of avirulence genes. *SIX1* has been reported as an avirulence gene in the Fol-tomato interaction *(AVR3)* and could potentially be recognized by cucumber and watermelon, triggering a defence response by these plants upon colonization by Fom. This is the only *SIX* gene homolog that is consistently present in the Fom genome but is not found in Forc.

Assembly of highly repetitive genomes benefits from long-read sequencing technology

Repetitive regions, including centromeres of *Fusarium*, are difficult to assemble using short-read sequencing technologies such as Illumina. *F. oxysporum*'s compartmentalized genome is a good example of a genome that can benefit greatly from longer read sequencing techniques, such as the PacBio SMRT sequencing technology employed here (median read length 15kb) as well as the development of novel technologies such as Oxford Nanopore sequencing (Datema *et al.*, 2016). Manual curation of the assembly improved it to the point where five chromosomes are complete (telomere-to-telomere), fifteen contigs have telomeric repeats on only one end and 20 contigs are left with no telomeric repeats on either end. The estimated chromosome count of Forc016 is 14, which is most clearly visible in Fig. 6 where the separation in the 1-1.5Mb region shows that Forc016 possesses two small accessory chromosomes. Considering their size and accessory-like appearance (high TE and low gene content), these

possibly correspond to contig 53 and the non-conserved region that is probably erroneously attached to chromosome 11 in the assembly. This is supported by the read mapping of chromosome loss strain #2, where no coverage was found for the mentioned two regions, as well as contig 3 and several of the smaller unplaced contigs (Fig. S9B). This strain lost these two chromosomes in addition to chr^{RC}, showing that they are conditionally dispensable. Moreover, the pathogenicity of the chr^{RC} chromosome transfer strains demonstrates that the two smallest Forc016 chromosomes are not required for pathogenicity. Comparison of the read coverage from chromosome loss strain #2 to wildtype will allow us to see which contigs belong to these chromosomes.
Even though PacBio SMRT sequencing is a great improvement compared to short-read technologies, it does not as yet allow for completely closed assemblies for *F. oxysporum* if not combined with other techniques like optical mapping (Faino *et al.*, 2015; Van Kan *et al.*, 2016). Nevertheless, the assembly of the core chromosomes as well as the pathogenicity chromosome of Forc016 were of sufficient quality to answer the biological questions addressed here.

Conclusions

We report here the near-complete genome assembly of Fo f.sp. *radicis-cucumerinum* strain Forc016 and horizontal transfer of its pathogenicity chromosome, chr^{RC}, to the non-pathogenic strain Fo47. This is the first time HCT has been accomplished using a donor strain from a *forma specialis* other than *lycopersici*. The virulence of the progeny strains deriving from this experiment is identical to that of the Forc chromosome donor, indicating that chr^{RC} is sufficient for root and shoot rot disease development. Complete loss of virulence of the five strains that lost chr^{RC} shows that chr^{RC} is also required for pathogenicity of Forc016. The experimental evidence presented here provides compelling confirmation that horizontal transfer of genetic material plays a crucial role in the adaptation to new host ranges of pathogenic isolates within the *F. oxysporum* species complex.

Materials and Methods

Fungal strains
F. oxysporum strains Forc016 ('33'; CBS141123) (Lievens *et al.*, 2007; van Dam *et al.*, 2016) and Fom001 (NRRL26406) (Ma *et al.*, 2014) were sequenced with SMRT sequencing technology. Fo47pGRB1 (Vlaardingerbroek, Beerens, Rose, *et al.*, 2016) was used as a chromosome recipient in HCT experiments.

Cloning
pPDh was constructed by introducing a *KpnI-KpnI* fragment containing a multiple cloning site (MCS) and the *eGFP* coding sequence followed by the *SIX1* terminator sequence,

amplified from pPZP200-pSIX1GFP (van der Does *et al.*, 2008), into the binary vector pRW2h (Houterman *et al.*, 2008). Additionally, a *HindIII-HindIII* fragment containing a MCS and the Herpes Simplex Virus thymidine kinase *(HSVtk)* gene under the control of the *C. heterostrophus* glyceraldehyde-3-phosphate dehydrogenase (ChGPD) gene promoter and the *N. crassa* β-tubulin gene terminator was inserted into the vector as a conditional negative selection marker against ectopic transformants (Khang *et al.*, 2005).
For knockout constructs, two ~1 kb fragments flanking the gene of interest were amplified using the primers listed in Table S2. The fragments were digested with *PacI-SpeI* and *AscI-Sbf1 (SIX6; SMP1)* or *PacI-SpeI* and *AscI-BstEII (SIX9)* and subsequently inserted on either side of the *GFP / HPH* cassette of pPDh.

Forc gene knockout
F. oxysporum strain Forc016 was transformed by *Agrobacterium* mediated transformation as described previously (Takken *et al.*, 2004). Following monosporing of hygromycin-resistant colonies, the transformants were grown in 96-well plates containing in each well 150 µl PDB supplemented with hygromycin and 5 µM 5-Fluoro-2-deoxyuridine (Alfa-Aesar) for pre-selection of *in locus* transformation (Khang *et al.*, 2005). Successful knockout of the genes was confirmed by PCR, using primers inside the T-DNA and outside the 1kb flanking region.

Disease assays
Pathogenicity tests were performed using the root dip method (Wellman, 1939). In short, conidia were isolated from five-day-old cultures NO_3-medium (0.17% yeast nitrogen base, 3% sucrose, 100 mM KNO_3) by filtering through miracloth (Merck; pore size of 22–25 µm). Spores were centrifuged, resuspended in sterile MilliQ water, counted and brought to a final concentration of 10^6 (effector KO assay) or 10^7 spores/mL (chromosome transfer and loss assays). When the first true leaves were emerging (after ±10 days), 6-8 seedlings per treatment were uprooted, inoculated, individually potted and kept at 25°C (effector KO assay) or 20°C (HCT assay) in the greenhouse. The following plant cultivars were used: *Cucumis sativus* cv. Paraiso, *Cucumis melo* cv. Cha-T, *Citrullus lanatus* cv. Black Diamond. Two weeks after inoculation, disease was scored using a disease index from 0-4 (0, no symptoms; 1, slight root rot symptoms, only at tip of main root; 2, root rot symptoms and stem lesions visible aboveground; 3, very clear root rot symptoms of the entire root system, often with a large lesion extending above the cotyledons; 4, plant either dead or very small and wilted).

Chromosome transfer and loss
Chromosome transfer from Forc016ΔSIX6#46 to Fo47pGRB (Vlaardingerbroek, Beerens, Rose, *et al.*, 2016) was performed through co-cultivation of the strains (van der Does and Rep, 2012). 1×10^5 microconidia from each of the two strains were mixed and co-incubated on PDA plates for six days. Newly formed spores were washed from the co-incubation plate

using 5 ml sterile MilliQ, filtered through sterile miracloth and pipetted on a double selective PDA plate containing 0.1 M Tris pH 8 supplemented with 100 µg/ml hygromycin (Duchefa) and 100 µg/ml zeocin (InvivoGen). Double drug resistant colonies were selected after six days and monospored by spreading on a fresh plate supplemented with both antibiotics. After two days of growth, single-spore colonies were selected and transferred to fresh plates.
Chromosome loss was induced as previously described (VanEtten *et al.*, 1998) with some modifications. Forc016*ΔSIX9*#97 was grown on PDA supplemented with hygromycin for 4 to 10 days. A Forc016 *ΔSIX9*#97 mycelial agar block was incubated in M100 broth (VanEtten *et al.*, 1998) containing 12.5 µg/ml benomyl (methyl 1-(butylcarbamoyl)-2-benzimidazolecarbamate, Aldrich) for 4 days, 175 rpm at 25°C. The culture was filtered through sterile miracloth. Conidia were collected by centrifugation and resuspended in 5 ml sterile water. 100 µl of a 100-fold dilution of conidia suspension was spread on M100 plates containing 0.04% Triton X-100 (Sigma). The plates were overlaid with a sterile filter paper and plates and conidia were incubated at 25°C for 2 days. The paper was transferred from M100 plates to PDA with hygromycin. After 1-2 days, the paper was removed and the colonies surviving only on M100 were selected and transferred to fresh PDA plates for further analysis.

Electrophoretic karyotyping and Southern analysis
Preparation of protoplasts and running of pulsed-field gel elecrophoresis was performed as described previously (Teunissen *et al.*, 2002; Vlaardingerbroek, Beerens, Rose, *et al.*, 2016). *F. oxysporum* was cultured in 100ml NO3 medium for five days. Next, microconidia were collected by filtration through a double layer of sterile miracloth. 5×10^8 spores were transferred to 40 ml PDB (BD Biosciences) and grown for 13 h at 25°C, followed by incubation at 30°C for 13–16 h in $MgSO_4$ solution (1.2 M $MgSO_4$, 50mM sodium citrate, pH 5.8) supplemented with 50 mg/ml Glucanex (Sigma). Protoplasts were filtered through a double layer of miracloth, collected by centrifugation and cast in InCert agarose (Lonza) plugs at a concentration of 1×10^8 protoplasts per ml. Plugs were treated with Pronase E and chromosomes were separated by running for 260 hours in 1% Seakem Gold agarose (Lonza) at 1.5 V/cm in a CHEF-DRII system (Biorad) in 0.5× TBE at 4°C with switch times between 1200 and 4800 s. The gels were stained with ethidium bromide and de-stained using 0.5× TBE.
DNA was blotted to a Hybond N+ membrane (Amersham Pharmacia) by alkaline transfer with 0.4N NaOH. A 793 bp PCR product containing the Fol*SIX6* open reading frame was generated with primers FP1490 and FP1491 (Table S2). This fragment was radioactively labelled with [α^{32}P]-dATP using the DecaLabel DNA labeling kit (Thermo Scientific). Hybridization was performed overnight at 65°C in Church and Gilbert buffer containing 0.5M phosphate, 7% SDS and 1 mM EDTA at pH 7.2. Blots were washed at 65°C with 0.5× SSC, 0.1% SDS. The position of chromosomal sequences to which the *SIX6* probe hybridized was visualized by phosphoimaging (Storm 840, Molecular Dynamics).

DNA isolation, genome sequencing and assembly

DNA isolation was performed on freeze-dried mycelium ground in liquid nitrogen as starting material, using multiple rounds of phenol-chloroform extraction and precipitation, as well as treatment with RNase A and proteinase K.

SMRT sequencing was performed at Keygene N.V. (Wageningen, the Netherlands). PacBio libraries were prepared and size-selected at ~20Kb using Blue Pippin prep. Sequencing of 5 SMRT cells was performed using the P6-C4 polymerase-chemistry combination, ≥4 hr movie time, stage start. This resulted in a sum of 4772Mb (Forc016) and 4846Mb (Fom001) filtered data. *De novo* assembly was performed with the Hierarchical Genome Assembly Process v3 (HGAP.3, Pacific Biosciences) within the SMRT Portal environment (v1.87.139483). Default values were kept and the expected genome size was set to 60Mb.

The raw assembly was manually improved by removing contigs originating from mtDNA and rDNA repeats. Two contigs that ended in telomeric repeats on one end and rDNA repeats on the other were joined together with in total 97 rDNA repeats in between (based on Illumina read mapping and coverage estimation on 10 rDNA repeats). Chromosome 13 could be reconstructed by joining two contigs that showed conserved synteny in Fom001 and the SMRT assembly of *F. subglutinans.* The two contigs were merged at the position of an overlapping region of 13,396 nt.

The mitochrondrial DNA was assembled from Illumina reads using GRAbB (Brankovics *et al.*, 2016) by specifying the mitochondrial genome of *F. oxysporum* F11 as reference and employing SPAdes as assembler. Annotation of the mitogenome was performed as described in Brankovics *et al.* (submitted) using a combination of MFannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl), tRNAscan-SE (Lowe and Eddy, 1996), NCBI ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder), InterPro (Mitchell *et al.*, 2015) and CD-Search (Marchler-Bauer and Bryant, 2004).

Illumina sequencing (150bp paired-end, insert size ~450bp) of HCT strains was performed on a HiSeq 2500 machine by the Hartwig Medical Foundation (Amsterdam, the Netherlands) at ~100X coverage, resulting in 5.0-5.6 Mb of sequence data per sample.

Genome annotation

Repeats were identified with RepeatMasker v4.0.6 (with -engine ncbi -species "ascomycota") (Smit *et al.*, 2015). Gene prediction was executed on the repeat-masked genome assembly by running BRAKER1 v1.9 (Hoff *et al.*, 2015), using RNA-seq read mappings (both *in vitro* and 10 days post inoculation *in planta* conditions) as additional evidence and supplying the following flags: --fungus --useexisting="fusarium_graminearum". Repeats and genes were counted over 50kb windows along the genome.

InterProScan v5.18-57.0 was used to assign functional annotation (including GO terms) to predicted genes. In order to find overrepresented GO terms on chr^{RC} versus the whole genome, a hypergeometric test was performed on the GO term frequencies using the 'phyper' function

in R. The p values were adjusted for multiple comparisons using 'p.adjust' and selecting the Bonferroni method in R. The results were visualized using REVIGO (Supek *et al.*, 2011).

Read mapping and genome analysis

For coverage plots, reads were trimmed to remove low-quality bases and adapter sequences using fastq-mcf v1.04.807 (-q 20) and mapped against the Forc genome assembly with Bowtie2 v2.2.5 (DNAseq) or Tophat2 v2.1.0 (RNAseq). Optical duplicates were removed using PicardTools MarkDuplicates v2.7.1 and coverage per 10kb (HCT plots) or 50kb (circos plots) windows was calculated with the samtools v1.3.1 mpileup command.

Whole genome or chromosome alignments were performed using nucmer (with --maxmatch) from the MUMmer v3.23 package (Delcher *et al.*, 2002). Comparison to the Fol4287 reference genome was done against an approximate chromosome-level assembly in which we concatenated scaffolds as assigned to chromosomes in (Ma *et al.*, 2010; Schmidt *et al.*, 2013), separated by 1000 Ns. We kept accessory regions of chromosomes 1 and 2 as separate sequences (for visualisation in Fig. 2E).

Identification of candidate effectors was done with BLASTN using the list of 104 candidates from van Dam *et al.* 2016 (van Dam *et al.*, 2016) as a query fasta. *Mimps* were identified by searching the genome for a consensus sequence of the *mimp* inverted repeat (IR), "TT[TA] TTGCNNCCCACTGNN". If two were found within 400 nt from each other in the correct orientation, they were marked as the ends of an intact *mimp* element.

Data availability

The Whole-Genome Shotgun projects for the resequenced strains have been deposited at Genbank under the BioProjects PRJNA389503 and PRJNA389439. The genome assemblies can be found on GenBank under accession numbers MABQ01000000 (Forc016 Illumina assembly), MABQ02000000 (Forc016 SMRT assembly) and NJCY01000000 (Fom001 SMRT assembly). Raw SMRT sequence data, Illumina read data of the HCT and chromosome loss strains and RNAseq reads have been deposited into the Sequence Read Archive under the accession number SRP108975. Illumina paired-end read data for Forc016 is available under accession number SRP067515 (DNAseq).

Acknowledgments

We are grateful to M. Paauw for help in the horizontal chromosome transfer experiments. We wish to thank B. Lievens and D.J. Vakalounakis for providing the Forc016 ('33') strain. Furthermore, we are grateful to ENZA seeds for kindly providing cucumber and melon seeds and Rijk Zwaan for kindly providing watermelon seeds. Finally, we would like to thank L. Tikovsky and H. Lemereis for their help in the greenhouse and H.C. van der Does for help with the CHEF gel. This work was supported by the Horizon programme (project 93512007) of the Netherlands Genomics Initiative (NGI) through a grant to M. Rep.

References

Armstrong, G.M. and Armstrong, J.K. (1981) *Formae speciales* and races of *Fusarium oxysporum* causing wilt diseases. In *Fusarium: diseases, biology, and taxonomy*. Pennsylvania State University Press, University Park, PA, pp. 391–399.

Baayen, R.P., Schoffelmeer, E.A.M., Toet, S., and Elgersma, D.M. (1997) Fungal polygalacturonase activity reflects susceptibility of carnation cultivars to fusarium wilt. *Eur. J. Plant Pathol.* **103**: 15–23.

Brankovics, B., Zhang, H., van Diepeningen, A.D., van der Lee, T.A.J., Waalwijk, C., and de Hoog, G.S. (2016) GRAbB: Selective Assembly of Genomic Regions, a New Niche for Genomic Research. *PLoS Comput. Biol.* **12**: 1–9.

Caprari, C., Mattei, B., Basile, M.L., Salvi, G., Crescenzi, V., De Lorenzo, G., and Cervone, F. (1996) Mutagenesis of endopolygalacturonase from *Fusarium moniliforme*: histidine residue 234 is critical for enzymatic and macerating activities and not for binding to polygalacturonase-inhibiting protein (PGIP). *Mol. Plant. Microbe. Interact.* **9**: 617—624.

Chakrabarti, A., Rep, M., Wang, B., Ashton, A., Dodds, P., and Ellis, J. (2011) Variation in potential effector genes distinguishing Australian and non-Australian isolates of the cotton wilt pathogen *Fusarium oxysporum* f.sp. *vasinfectum*. *Plant Pathol.* **60**: 232–243.

Cohen, R., Orgil, G., Burger, Y., Saar, U., Elkabetz, M., Tadmor, Y., *et al.* (2015) Differences in the responses of melon accessions to fusarium root and stem rot and their colonization by *Fusarium oxysporum* f. sp. *radicis-cucumerinum*. *Plant Pathol.* **64**: 655–663.

van Dam, P., Fokkens, L., Schmidt, S.M., Linmans, J.H.J., Kistler, H.C., Ma, L.-J., and Rep, M. (2016) Effector profiles distinguish formae speciales of Fusarium oxysporum. *Environ. Microbiol.* **18**: 4087–4102.

van Dam, P. and Rep, M. (2017) The Distribution of *Miniature Impala* Elements and *SIX* Genes in the *Fusarium* Genus is Suggestive of Horizontal Gene Transfer. J. Mol. Evol.

Datema, E., Hulzink, R.J.M., Blommers, L., Espejo Valle-Inclan, J., Van Orsouw, N., Wittenberg, A.H.J., and De Vos, M. (2016) The megabase-sized fungal genome of *Rhizoctonia solani* assembled from nanopore reads only. *bioRxiv*.

Delcher, A.L., Phillippy, A., Carlton, J., and Salzberg, S.L. (2002) Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* **30**: 2478–2483.

van der Does, H.C., Duyvesteijn, R.G.E., Goltstein, P.M., van Schie, C.C.N., Manders, E.M.M., Cornelissen, B.J.C., and Rep, M. (2008) Expression of effector gene *SIX1* of *Fusarium oxysporum* requires living plant cells. *Fungal Genet. Biol.* **45**: 1257–1264.

van der Does, H.C., Fokkens, L., Yang, A., Schmidt, S.M., Langereis, L., Lukasiewicz, J.M., *et al.* (2016) Transcription Factors Encoded on Core and Accessory Chromosomes of *Fusarium oxysporum* Induce Expression of Effector Genes. *PLOS Genet.* **12**: e1006401.

van der Does, H.C. and Rep, M. (2012) Horizontal transfer of supernumerary chromosomes in fungi. *Plant Fungal Pathog. Methods Protoc.* 427–437.

Faino, L., Seidl, M.F., Datema, E., van den Berg, G.C.M., Janssen, A., Wittenberg, A.H.J., and Thomma, B.P.H.J. (2015) Single-Molecule Real-Time Sequencing Combined with Optical Mapping Yields Completely Finished Fungal Genome. *MBio* **6**: e00936--15.

Gawehns, F., Houterman, P.M., Ichou, F.A., Michielse, C.B., Hijdra, M., Cornelissen, B.J.C., *et al.* (2014) The *Fusarium oxysporum* Effector Six6 Contributes to Virulence and Suppresses I-2-Mediated Cell Death. *Mol. Plant-Microbe Interact.* **27**: 336–348.

Guo, L., Han, L., Yang, L., Zeng, H., Fan, D., Zhu, Y., *et al.* (2014) Genome and Transcriptome Analysis of the Fungal Pathogen *Fusarium oxysporum* f. sp. *cubense* Causing Banana Vascular Wilt Disease. *PLoS One* **9**: e95543.

Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2015) BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* btv661.

Houterman, P.M., Cornelissen, B.J.C., and Rep, M. (2008) Suppression of plant resistance gene-based immunity by a fungal effector. *PLoS Pathog.* **4**: e1000061.

Houterman, P.M., Ma, L., Van Ooijen, G., De Vroomen, M.J., Cornelissen, B.J.C., Takken, F.L.W., and Rep, M. (2009) The effector protein Avr2 of the xylem-colonizing fungus *Fusarium oxysporum* activates the tomato resistance protein I-2 intracellularly. *Plant J.* **58**: 970–978.

Houterman, P.M., Speijer, D., Dekker, H.L., De Koster, C.G., Cornelissen, B.J.C., and Rep, M. (2007) The mixed

xylem sap proteome of *Fusarium oxysporum*-infected tomato plants. *Mol. Plant Pathol.* **8**: 215–221.
Van Kan, J.A.L., Stassen, J.H.M., Mosbach, A., Van Der Lee, T.A.J., Faino, L., Farmer, A.D., *et al.* (2016) A gapless genome sequence of the fungus *Botrytis cinerea*. *Mol. Plant Pathol.* 1–15.
Karaca, G. and Kahveci, E. (2010) First report of *Fusarium oxysporum* f. sp. *radicis-cucumerinum* on cucumbers in Turkey. *Plant Pathol.* **59**: 1173–1174.
Katan, T. (1999) Current status of vegetative compatibility groups in *Fusarium oxysporum*. *Phytoparasitica* **27**: 51–64.
Khang, C.H., Park, S.-Y., Lee, Y.-H., and Kang, S. (2005) A dual selection based, targeted gene replacement tool for *Magnaporthe grisea* and *Fusarium oxysporum*. *Fungal Genet. Biol.* **42**: 483–492.
Kistler, H.C., Alabouvette, C., Baayen, R.P., Bentley, S., Brayford, D., Coddington, A., *et al.* (1998) Systematic Numbering of Vegetative Compatibility Groups in the Plant Pathogenic Fungus *Fusarium oxysporum*. *Phytopathology* **88**: 30–32.
Kleemann, J., Rincon-Rivera, L.J., Takahara, H., Neumann, U., van Themaat, E.V.L., van der Does, H.C., *et al.* (2012) Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathog.* **8**: e1002643.
Lievens, B., Claes, L., Vakalounakis, D.J., Vanachter, A.C.R.C., and Thomma, B.P.H.J. (2007) A robust identification and detection assay to discriminate the cucumber pathogens *Fusarium oxysporum* f. sp. *cucumerinum* and f. sp. *radicis-cucumerinum*. *Environ. Microbiol.* **9**: 2145–2161.
Lowe, T.M. and Eddy, S.R. (1996) TRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**: 955–964.
Ma, L.-J. (2014) Horizontal chromosome transfer and rational strategies to manage fusarium vascular wilt diseases. *Mol. Plant Pathol.* **15**: 1–8.
Ma, L.-J., van der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.-J., Di Pietro, A., *et al.* (2010) Comparative analysis reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367–373.
Ma, L.-J., Shea, T., Young, S., Zeng, Q., and Kistler, H.C. (2014) Genome sequence of *Fusarium oxysporum* f. sp. *melonis* strain NRRL 26406, a fungus causing wilt disease on melon. *Genome Announc.* **2**: e00730-14.
Ma, L., Cornelissen, B.J.C., and Takken, F.L.W. (2013) A nuclear localization for Avr2 from *Fusarium oxysporum* is required to activate the tomato resistance protein I-2. *Front. Plant Sci.* **4**: 94.
Ma, L., Houterman, P.M., Gawehns, F., Cao, L., Sillo, F., Richter, H., *et al.* (2015) The *AVR2-SIX5* gene pair is required to activate *I-2*-mediated immunity in tomato. *New Phytol.* **208**: 507–518.
Marchler-Bauer, A. and Bryant, S.H. (2004) CD-Search: Protein domain annotations on the fly. *Nucleic Acids Res.* **32**: 327–331.
Michielse, C.B. and Rep, M. (2009) Pathogen profile update: *Fusarium oxysporum*. *Mol. Plant Pathol.* **10**: 311–324.
Michielse, C.B., van Wijk, R., Reijnen, L., Cornelissen, B.J.C., and Rep, M. (2009) Insight into the molecular requirements for pathogenicity of *Fusarium oxysporum* f. sp. *lycopersici* through large-scale insertional mutagenesis. *Genome Biol.* **10**: R4.
Mitchell, A., Chang, H.Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., *et al.* (2015) The InterPro protein families database: The classification resource after 15 years. *Nucleic Acids Res.* **43**: D213–D221.
Moreno, A., Alférez, A., Avilés, M., Diánez, F., Blanco, R., Santos, M., and Tello, J.C. (2001) First report of *Fusarium oxysporum* f. sp. *radicis-cucumerinum* on cucumber in Spain. *Plant Dis.* **85**: 1206–1206.
Niño-Sánchez, J., Casado-Del Castillo, V., Tello, V., de Vega-Bartol, J.J., Ramos, B., Sukno, S.A., *et al.* (2016) The *FTF* gene family regulates virulence and expression of *SIX* effectors in *Fusarium oxysporum*. *Mol. Plant Pathol.* 1–18.
Niu, X., Zhao, X., Ling, K.-S., Levi, A., Sun, Y., and Fan, M. (2016) The FonSIX6 gene acts as an avirulence effector in the *Fusarium oxysporum* f. sp. *niveum* - watermelon pathosystem. *Sci. Rep.* **6**: 28146.
Di Pietro, A., Madrid, M.P., Caracuel, Z., Delgado-Jarana, J., and Roncero, M.I.G. (2003) *Fusarium oxysporum*: exploring the molecular arsenal of a vascular wilt fungus. *Mol. Plant Pathol.* **4**: 315–325.
Punja, Z.K. and Parker, M. (2000) Development of fusarium root and stem rot, a new disease on greenhouse cucumber in British Columbia, caused by *Fusarium oxysporum* f. sp. *radicis-cucumerinum*. *Can. J. Plant Pathol.* **22**: 349–363.
Ramos, B., Alves-Santos, F.M., García-Sánchez, M. a., Martín-Rodrigues, N., Eslava, A.P., and Díaz-Mínguez, J.M. (2007) 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.
Reignault, P., Valette-Collet, O., and Boccara, M. (2008) The importance of fungal pectinolytic enzymes in plant invasion, host adaptability and symptom type. *Eur. J. Plant Pathol.* **120**: 1–11.

Rep, M., Van Der Does, H.C., Meijer, M., Van Wijk, R., Houterman, P.M., Dekker, H.L., *et al.* (2004) 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.

Rocha, L.O., Laurence, M.H., Ludowici, V.A., Puno, V.I., Lim, C.C., Tesoriero, L.A., *et al.* (2015) Putative effector genes detected in *Fusarium oxysporum* from natural ecosystems of Australia. *Plant Pathol.* **65**: 914–929

Schmidt, S.M., Houterman, P.M., Schreiver, I., Ma, L., Amyotte, S., Chellappan, B., *et al.* (2013) MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genomics* **14**: 119.

Schmidt, S.M., Lukasiewicz, J., Farrer, R., van Dam, P., Bertoldo, C., and Rep, M. (2016) Comparative genomics of *Fusarium oxysporum* f. sp. *melonis* reveals the secreted protein recognized by the *Fom-2* resistance gene in melon. *New Phytol.* **209**: 307–318.

Smit, A.F.A., Hubley, R., and Green, P. (2015) RepeatMasker Open-4.0. 2013-2015. *Inst. Syst. Biol.*

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**:.

Takken, F. and Rep, M. (2010) The arms race between tomato and *Fusarium oxysporum*. *Mol. Plant Pathol.* **11**: 309–314.

Takken, F.L.W., van Wijk, R., Michielse, C.B., Houterman, P.M., Ram, A.F.J., and Cornelissen, B.J.C. (2004) A one-step method to convert vectors into binary vectors suited for *Agrobacterium*-mediated transformation. *Curr. Genet.* **45**: 242–248.

Teunissen, H.A.S., Verkooijen, J., Cornelissen, B.J.C., and Haring, M.A. (2002) Genetic exchange of avirulence determinants and extensive karyotype rearrangements in parasexual recombinants of *Fusarium oxysporum*. *Mol. Genet. Genomics* **268**: 298–310.

Thatcher, L.F., Gardiner, D.M., Kazan, K., and Manners, J.M. (2012) A highly conserved effector in *Fusarium oxysporum* is required for full virulence on *Arabidopsis*. *Mol. Plant. Microbe. Interact.* **25**: 180–90.

Tok, F.M. and Kurt, Ş. (2010) Pathogenicity, vegetative compatibility and amplified fragment length polymorphism (AFLP) analysis of *Fusarium oxysporum* f. sp. *radicis-cucumerinum* isolates from Turkish greenhouses. *Phytoparasitica* **38**: 253–260.

Vakalounakis, D.J. (1996) Root and Stem Rot of Cucumber Caused by *Fusarium oxysporum* f.sp. *radicis-cucumerinum* f.sp. nov. *Plant Dis.* **80**: 313–316.

Vakalounakis, D.J. and Fragkiadakis, G.A. (1999) Genetic Diversity of *Fusarium oxysporum* Isolates from Cucumber: Differentiation by Pathogenicity, Vegetative Compatibility, and RAPD Fingerprinting. *Phytopathology* **89**: 161–168.

Vakalounakis, D.J., Wang, Z., Fragkiadakis, G.A., Skaracis, G.N., and Li, D.-B. (2004) Isolates Obtained from Cucumber in China by Pathogenicity, VCG, and RAPD. *Plant Dis.* **88**: 645–649.

VanEtten, H., Jorgensen, S., Enkerli, J., and Covert, S.F. (1998) Inducing the loss of conditionally dispensable chromosomes in *Nectria haematococca* during vegetative growth. *Curr. Genet.* **33**: 299–303.

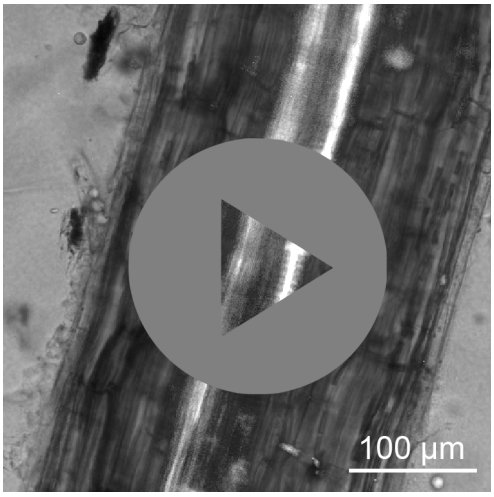
Vlaardingerbroek, I., Beerens, B., Rose, L., Fokkens, L., Cornelissen, B.J.C., and Rep, M. (2016) Exchange of core chromosomes and horizontal transfer of lineage-specific chromosomes in *Fusarium oxysporum*. *Environ. Microbiol.***18:** 3702–3713.

Vlaardingerbroek, I., Beerens, B., Schmidt, S.M., Cornelissen, B.J.C., and Rep, M. (2016) Dispensable chromosomes in *Fusarium oxysporum* f. sp *lycopersici*. *Mol. Plant Pathol.* **17:** 1455-1466

Wellman, F.L. (1939) A technique for studying host resistance and pathogenicity in tomato Fusarium wilt. *Phytopathology* **29**: 945–956.

Williams, A.H., Sharma, M., Thatcher, L.F., Azam, S., Hane, J.K., Sperschneider, J., *et al.* (2016) Comparative genomics and prediction of conditionally dispensable sequences in legume–infecting *Fusarium oxysporum formae speciales* facilitates identification of candidate effectors. *BMC Genomics* **17**: 191.

Supplementary Data



Video S1: Microscopic timelapse movie of GFP-tagged Forc016 shows the fungus colonizing a cucumber plant through the xylem tissue.
Nine day old cucumber seedlings were inoculated with a Forc016 strain that was transformed with the pPK2*hphgfp* construct (*Hyg*^{R}*-GFP* phusion protein under the control of the constitutive *gpdA* promoter) (Michielse *et al.*, 2009). At 9 days post inoculation, the timelapse was recorded over an 8h20m time period with 1 minute-intervals, showing that Forc colonizes the plant like wilt-inducing strains of *F. oxysporum* do: by growing through the xylem vessels of the plant. (Available online via https://www.nature.com/articles/s41598-017-07995-y [Supplementary Data])

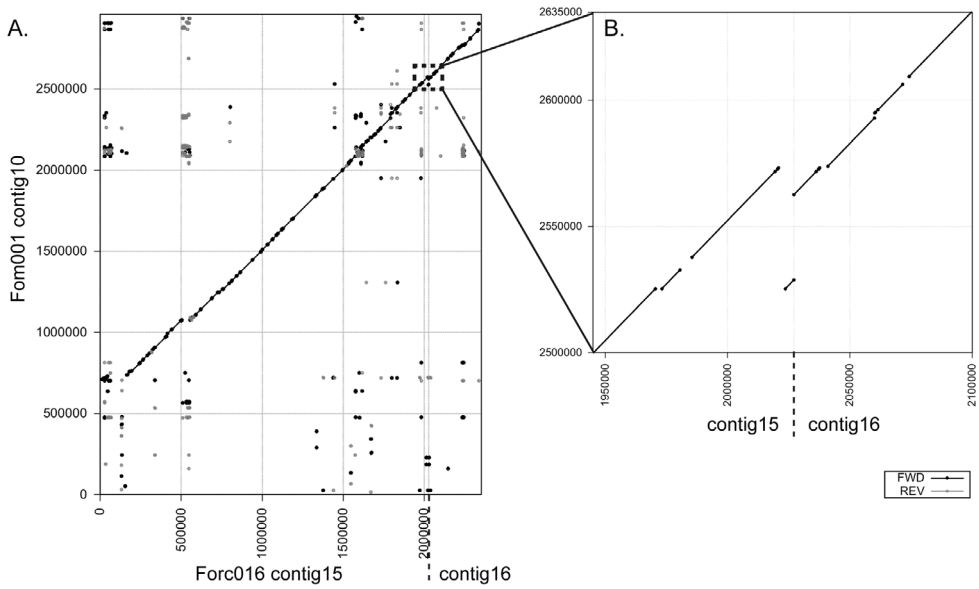


Fig. S1: Two contigs in the Forc016 assembly, 15 and 16, display an overlap of 13,396 nt and are syntenic in Fom001, indicating that it is highly likely that they together form chromosome 13 in Forc016.

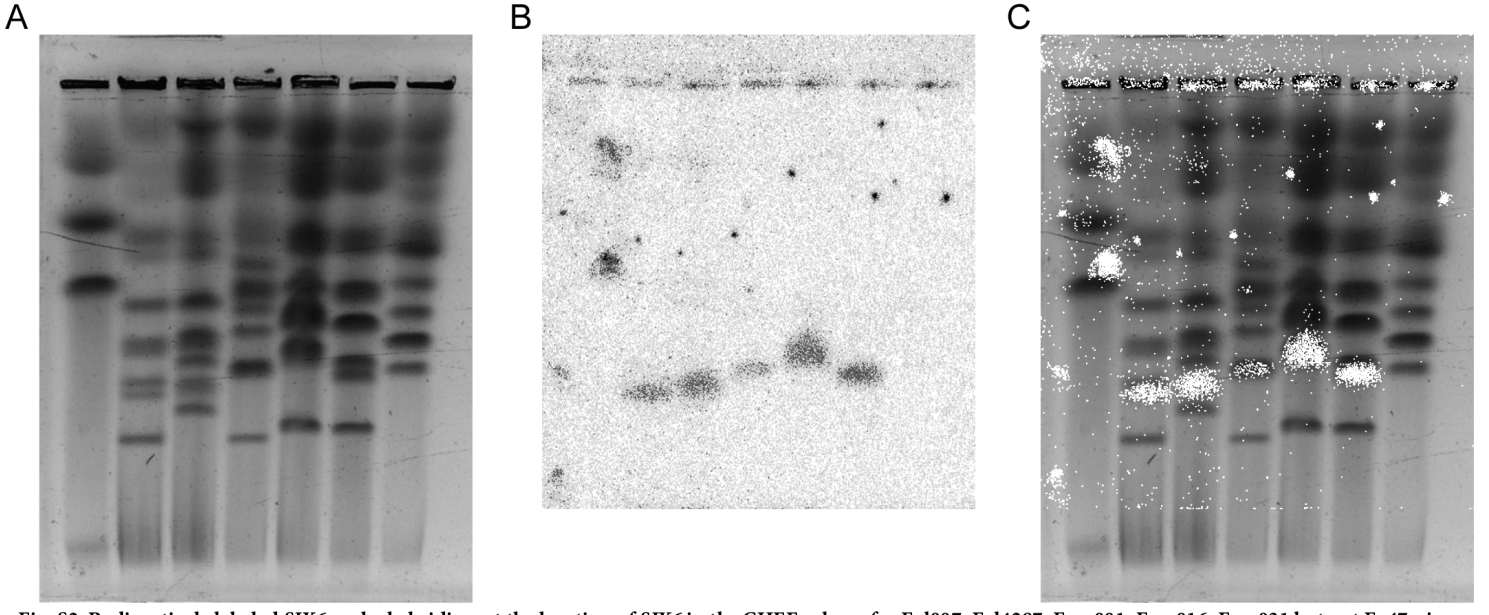


Fig. S2: Radioactively labeled *SIX6* probe hybridizes at the location of *SIX6* in the CHEF gel run for Fol007, Fol4287, Fom001, Forc016, Forc031 but not Fo47, since it does not possess this gene.

(A) CHEF gel separation of the chromosomes of these strains. The left lane shows the marker *(Saccharomyces cerevisiae* chromosomes) with bands indicating 5.7, 4.6 and 3.5 Mb. **(B)** Southern hybridization signal using a *SIX6* probe. **(C)** Overlay of the Southern hybridization signal (in white) over the CHEF gel in figure A.

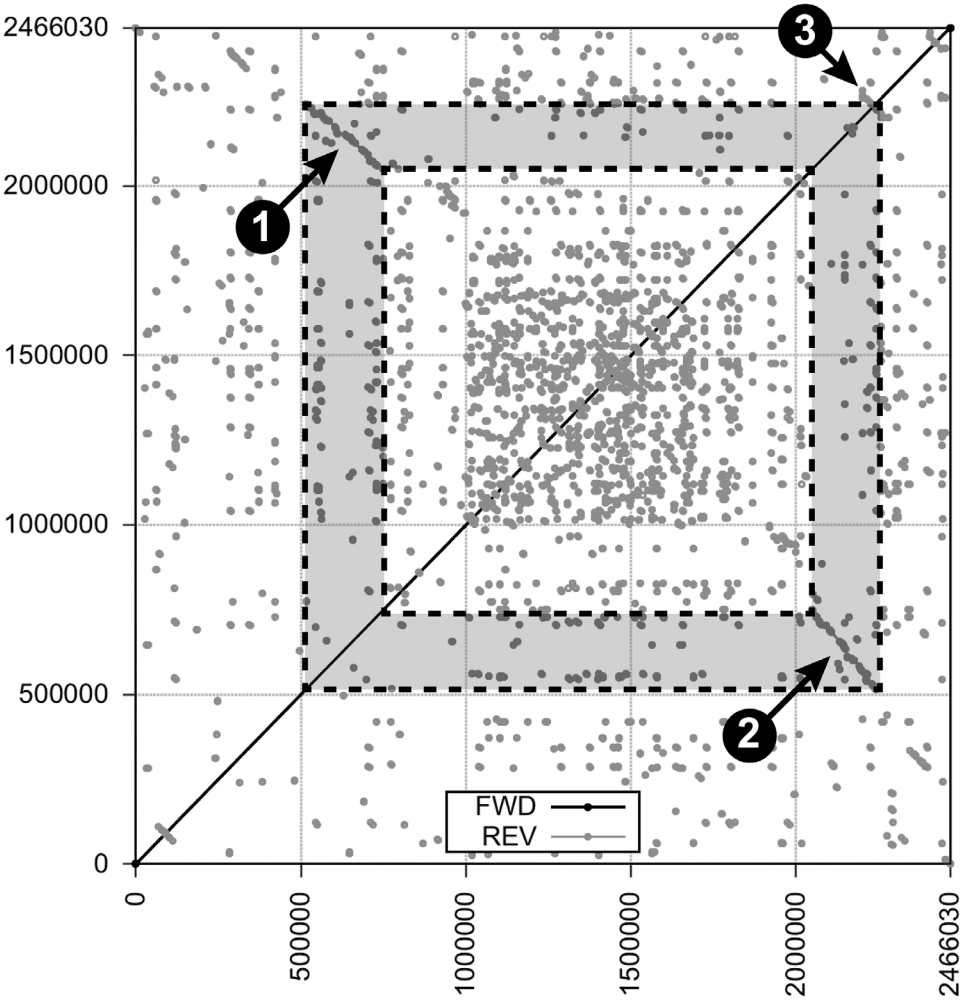


Fig. S3: Nucmer comparison of Forc016's chr^{RC} to itself reveals that large repetitive regions are present around the middle region of the chromosome.
These repetitive regions are located on 500-700kb (arrow 1) and 2000-2200kb (arrow 2) of the chromosome, which resulted in a misassembled inversion of this sequence in the original assembly. This region (marked as a gray box) was manually inverted in contig 13 and merged with contig 17 at the position of arrow 3.

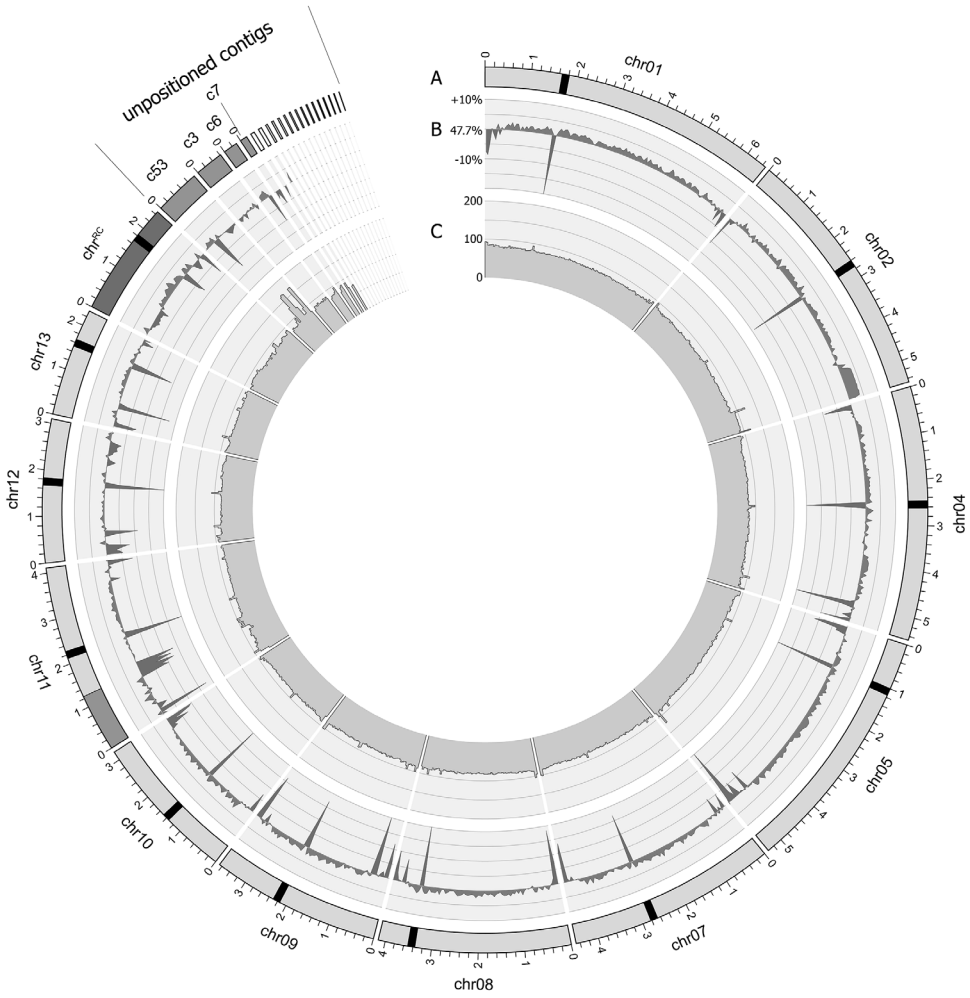


Fig. S4: Visualization of the Forc016 genome assembly reveals centromeres in the assembly and two segmental duplications on contig 53.
The panels in this figure indicate **(A)** the karyotype of the assembly, with core chromosomes (light gray), accessory regions (darker gray) and the pathogenicity chromosome, chr^{RC} (dark gray). Probable centromeres (characterized by low GC content, shown in **(B)**) are indicated with black blocks. **(C)** Read density levels calculated in 50kb windows for Illumina paired-end read mapping shows that almost the complete genome is covered at about 90X, with the notable exception being contig 53 which shows two large segmental duplications, roughly 220kb and 140kb in size.

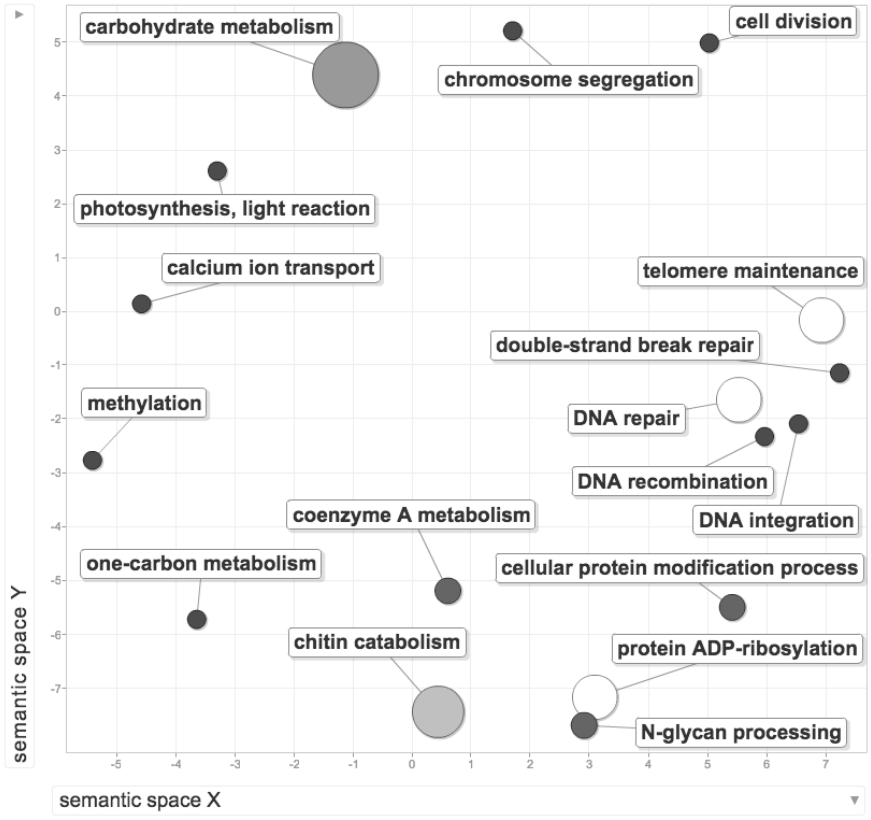


Fig. S5: Overrepresented gene ontology (GO) terms on chr^{RC} include genes related to carbohydrate metabolism, chitin metabolism, protein ADP-ribosylation and several groups related to DNA integrity (DNA repair, telomere maintenance, DNA recombination, DNA integration, chromosome segregation and cell division).

A hypergeometric GO term enrichment analysis ($p<0.05$) was performed to identify which types of genes (other than effectors) are overrepresented on chr^{RC} compared to the rest of the Forc016 genome. Overrepresented GO terms related to Biological Process (P) were visualized using REVIGO (http://revigo.irb.hr/). The axes have no intrinsic meaning - REVIGO uses multi-dimensional scaling to reduce the dimensionality of a matrix of the GO terms' pairwise semantic similarities. Semantically similar GO terms should remain close together in the plot. See Table S2 for further details on the overrepresented genes.

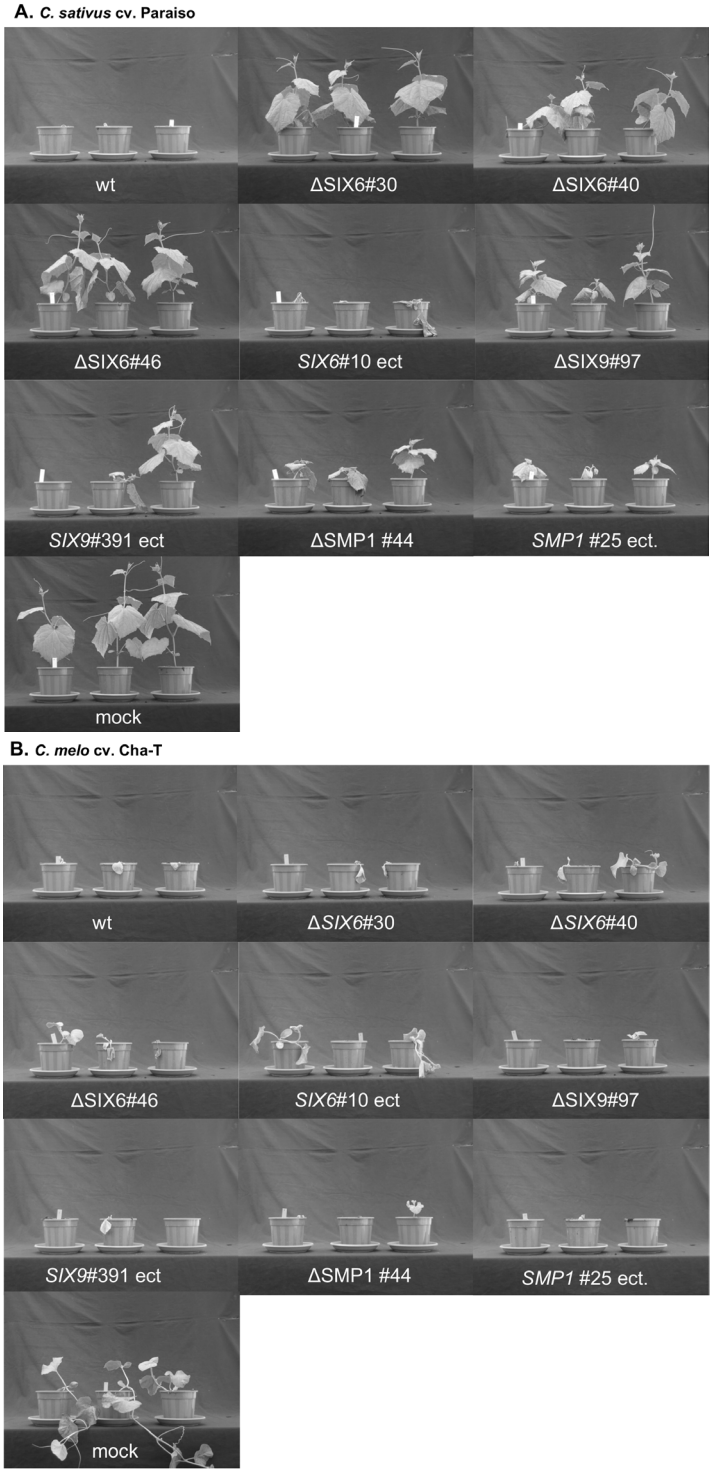
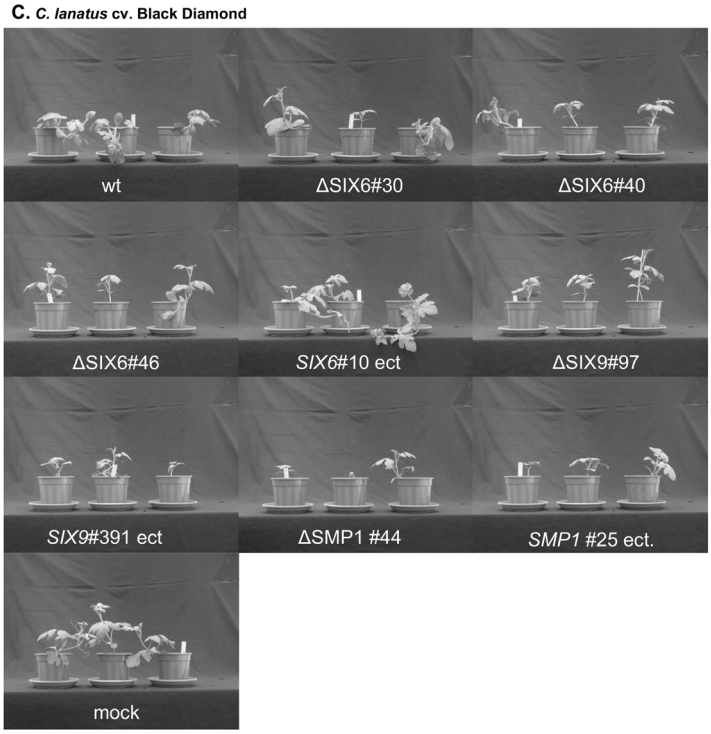


Fig. S6:
Symptom development in Forc016Δ*SIX6* treated plants is less strong than in the control treatments, particularly on cucumber plants.
(A) Cucumber, **(B)** melon and **(C)** watermelon plants two weeks after inoculation with different effector candidate knockout strains. The most notable difference is seen in **(D)** cucumber plants treated with three individual *SIX6* deletion strains (Δ*SIX6* #30, #40 and #46) compared to an ectopic transformant , a wildtype strain and mock.



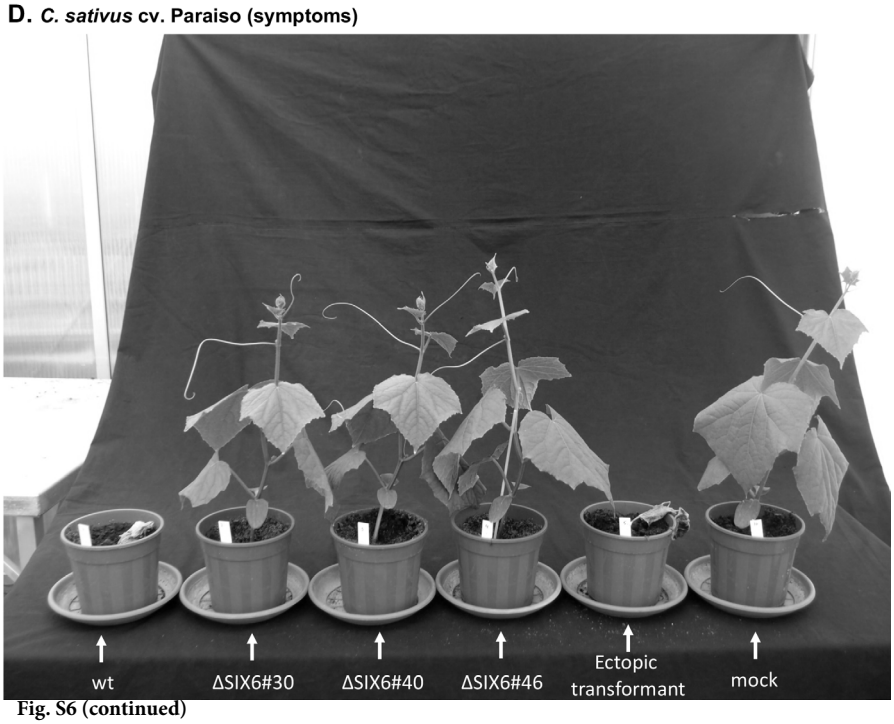
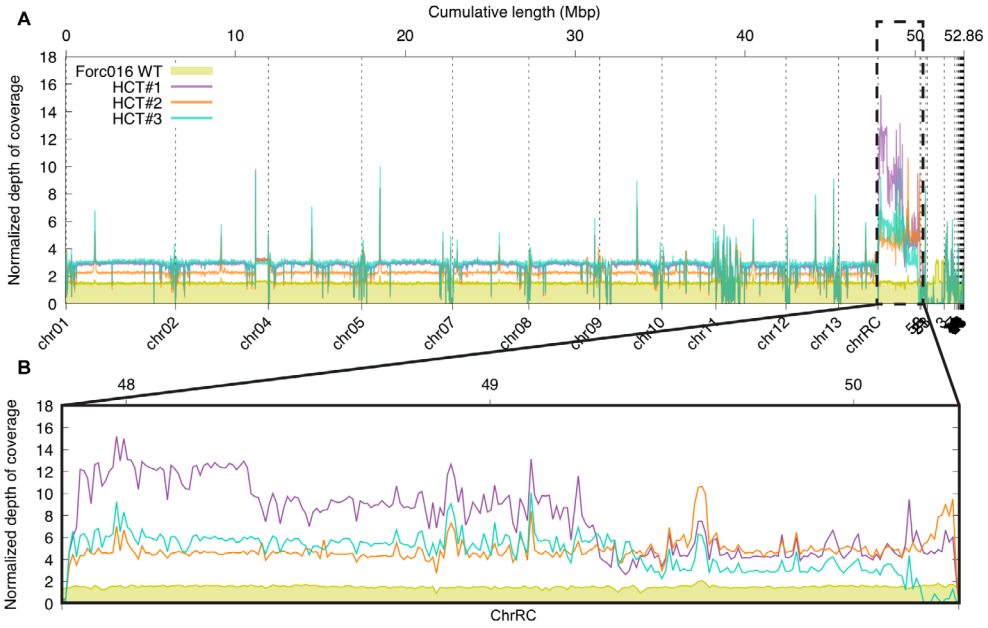


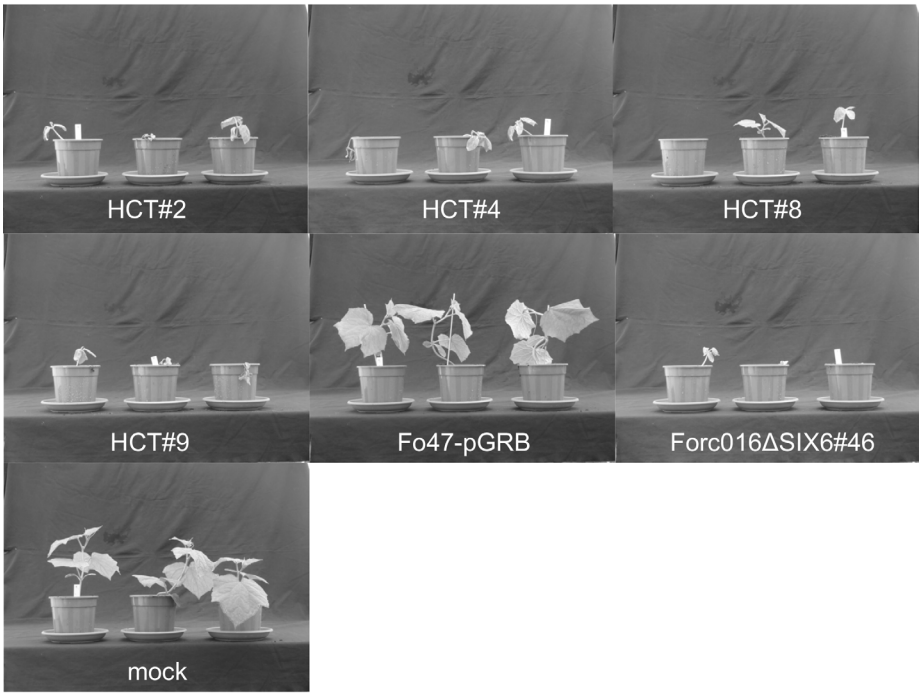
Fig. S6 (continued)



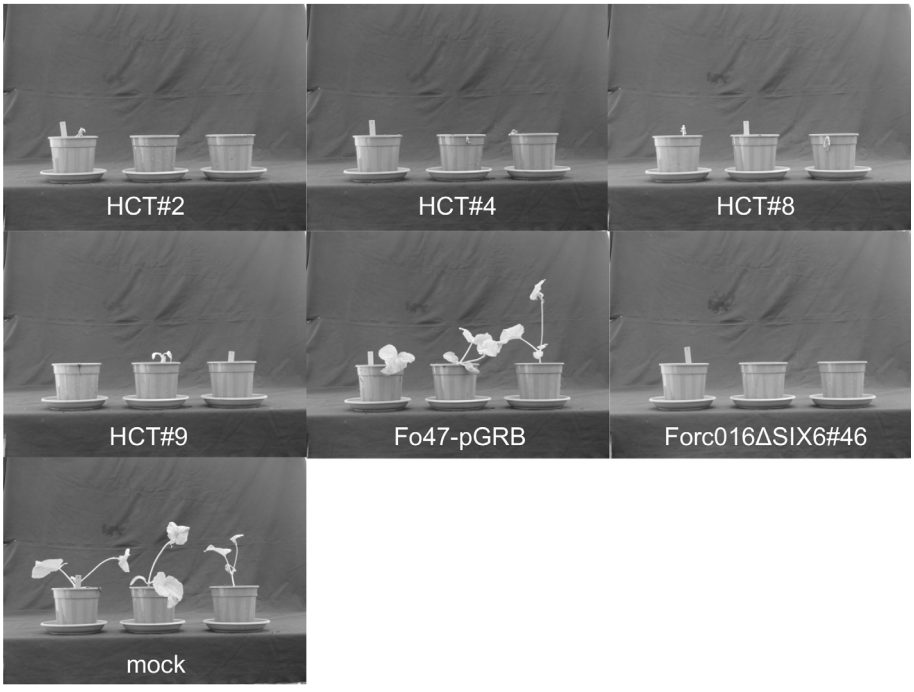
▲ Fig. S7: Normalized Illumina read mapping to the SMRT assembly of Forc016 shows large scale chromosome rearrangements and duplications upon horizontal transfer of chr^{RC} into a Fo47 background of HCT strains #1 and #3.
(A) Reads mapped more abundantly to the transferred chr^{RC} sequence than the rest of the assembly. **(B)** HCT strain #2 was included as a control that obtained a single copy of chr^{RC} in the Fo47 background (showing a relative coverage ±4x along the entire chromosome). HCT #1 has relative coverage depths that vary between 12x, 8x and 4x along the entire length of chr^{RC}, suggesting large segmental duplications of parts of the chromosome. HCT #3 displays coverage along the entire chromosome except for the terminal part, where the coverage drops to 0.

▶ Fig. S8: Horizontal chromosome transfer (HCT) of Forc chr^{RC} to Fo47 results in strains that are pathogenic on cucurbits.
(A) Cucumber, **(B)** melon and **(C)** watermelon plants two weeks after inoculation with four HCT strains. **(D)** Typical root and shoot rot symptoms (maceration and lesion formation along the hypocotyl) associated with Forc also develop when plants are inoculated with the HCT-strains. Figure continues on next page.

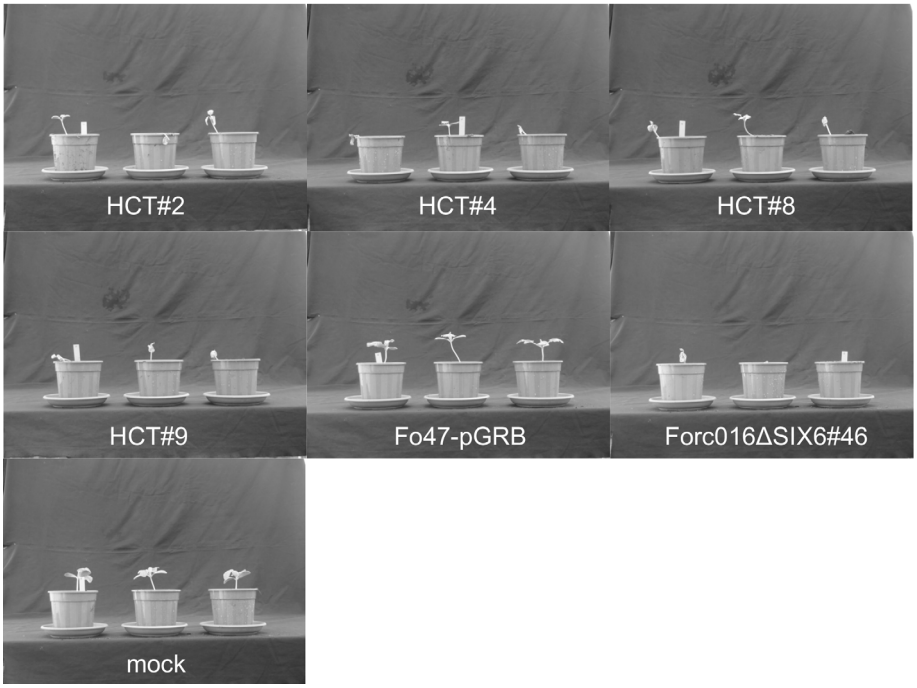
A. ***C. sativus*** **cv. Paraiso**



B. ***C. melo*** **cv. Cha-T**



C. ***C. lanatus*** **cv. Black Diamond**



D. ***C. sativus*** **cv. Paraiso (symptoms)**

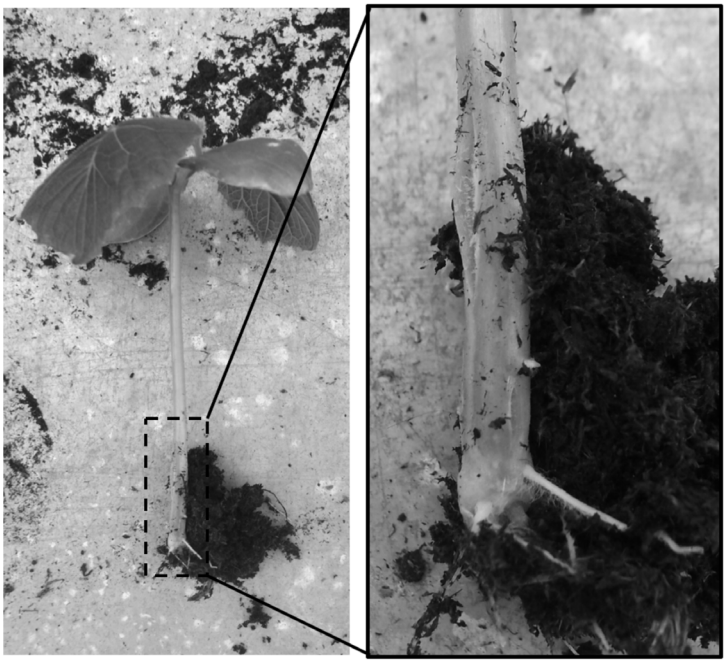


Fig. S8 (continued)

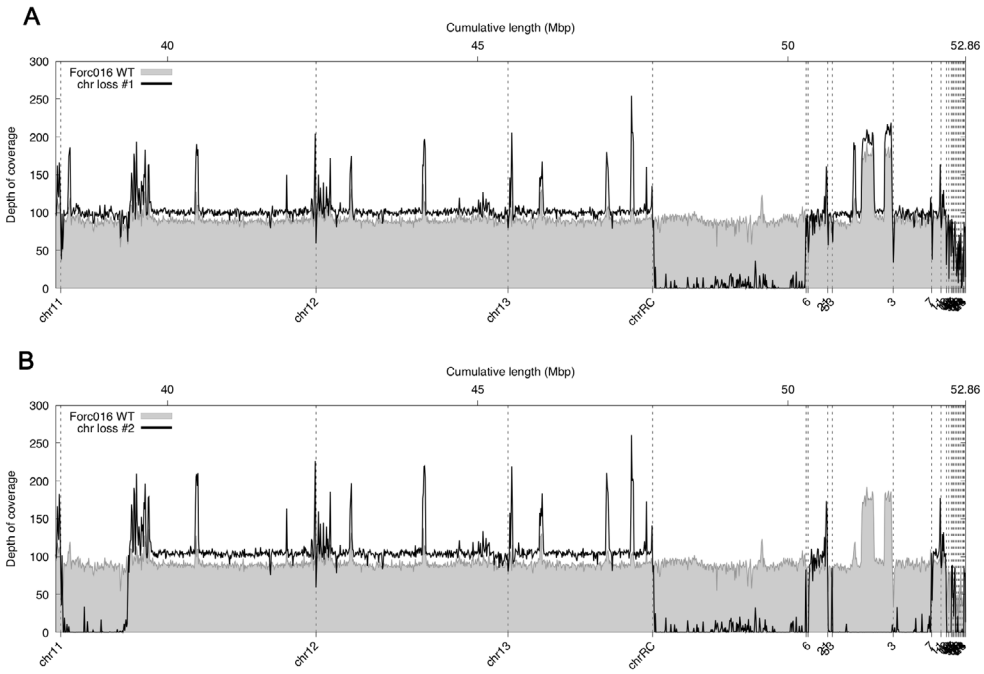


Fig. S9: Illumina read mapping to the SMRT assembly of Forc016 (shown from chromosome 11 onwards) demonstrates the loss of chr^{RC} and, in strain #2, additional sequences corresponding to the two smallest chromosomes of Forc016.

(A) Read coverage in chr loss strain #1 indicates specific loss of chr^{RC} from the genome, while **(B)** in chr loss strain #2 this chromosome was lost along with the two small accessory chromosomes. Since no coverage was found for part of chr11 and contigs 53, 3, 21 and several smaller contigs, these together likely correspond to the two smallest chromosomes of Forc016.

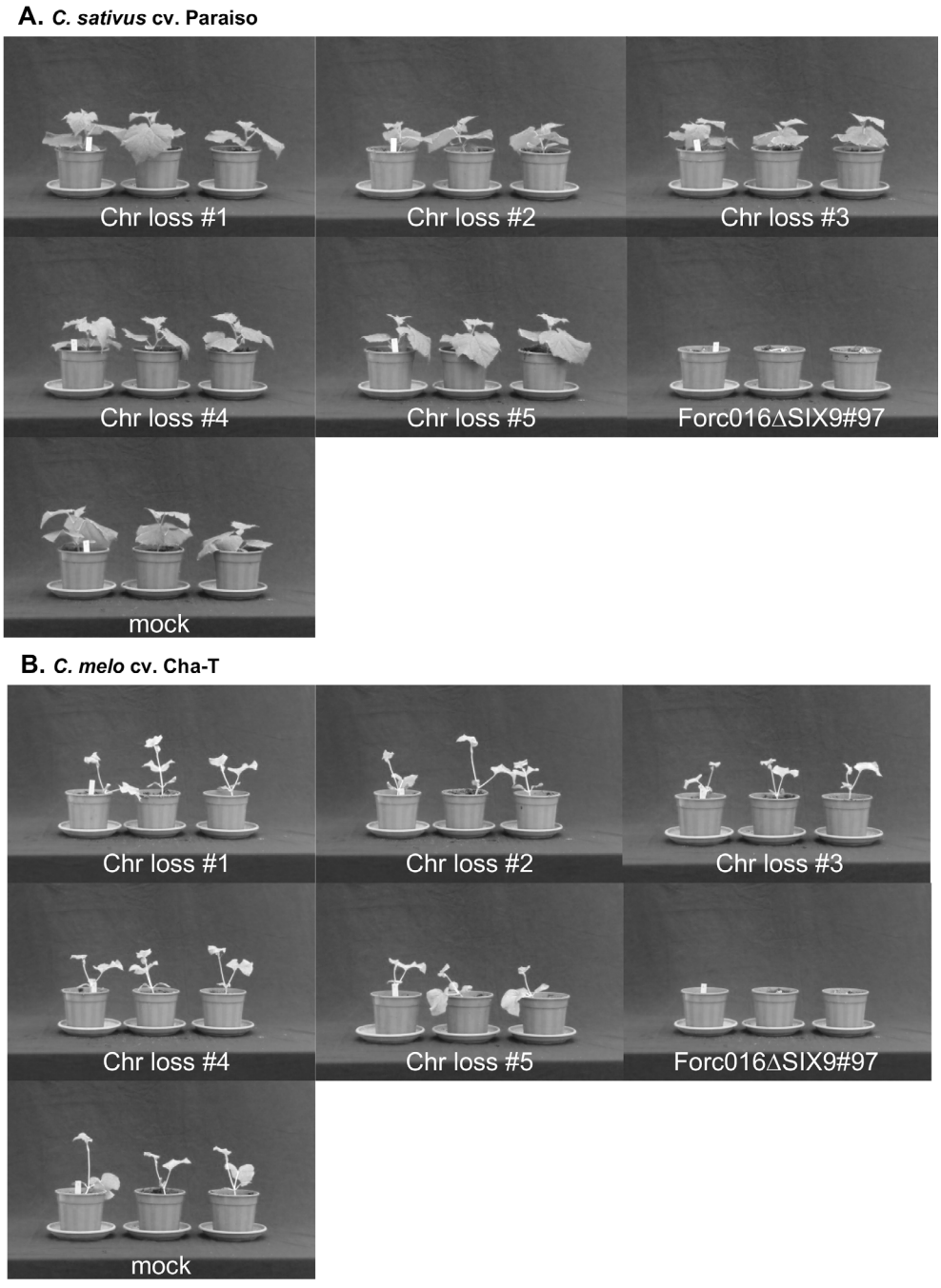


Fig. S10: Forc016 strains that lost chr^{RC} have completely lost their virulence.
(A) Cucumber, **(B)** melon and **(C)** watermelon plants two weeks after inoculation with five strains that lost chr^{RC}, their parent strain (Forc016Δ*SIX9*#97) and mock clearly illustrate that only treatment with the parent strain results in disease development.

C. ***C. lanatus*** **cv. Black Diamond**

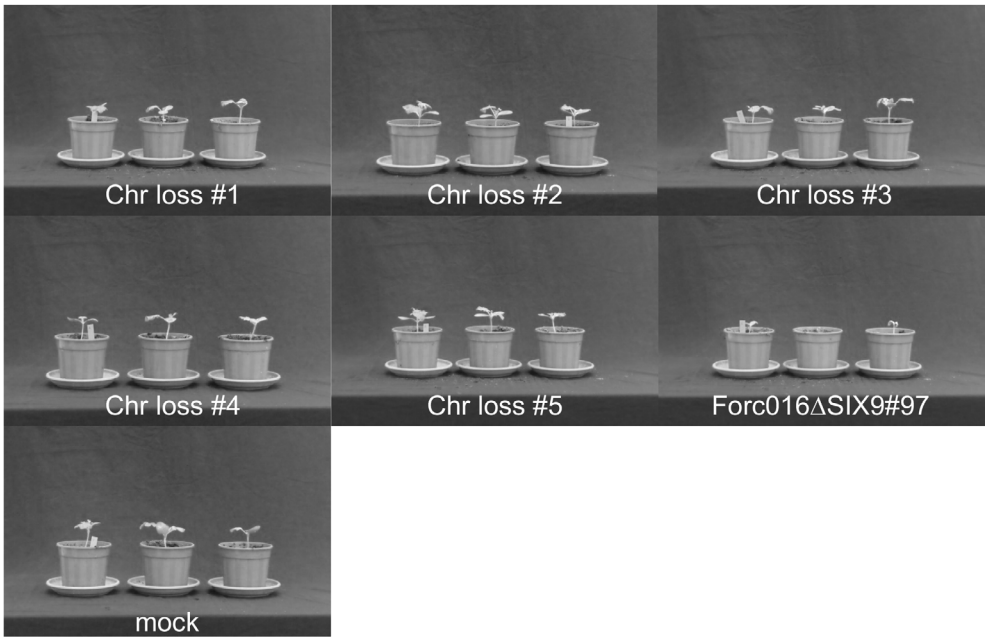


Fig. S10 (continued)

Table S1: Genes identified through GO term enrichment analysis (Fig. S5) on chr^{RC}

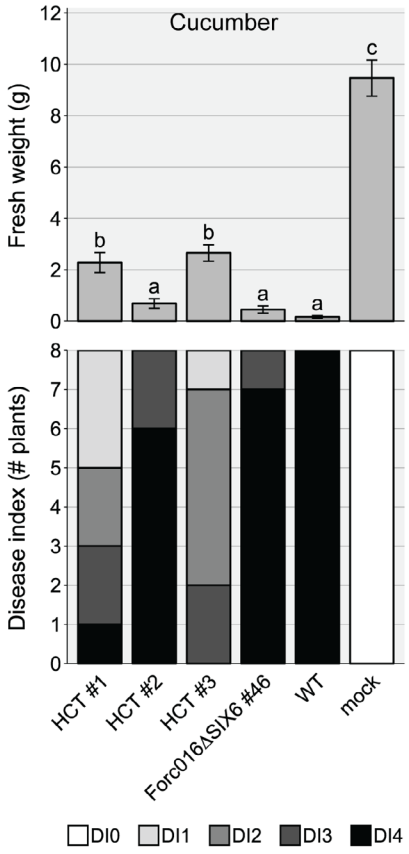
GO term	GO term name	Gene ID	Predicted function or domain	Start coordinate	End coordinate	Orientation
GO:0005975	carbohydrate metabolic process	g15643	Glycosyl hydrolases family 16 (GH16) domain profile	366128	367216	+
		g15837	Glycosyl hydrolase family 3 C-terminal domain	1194231	1196864	-
		g15838	Galactose mutarotase-like	1198981	1201953	+
		g15874	NodB homology domain profile	1327871	1328549	-
		g15955	Glycosyl hydrolase family 3 N terminal domain	1628362	1628772	+
GO:0005975 / GO:0006032	carbohydrate / chitin metabolic process	g15626	Chitin-binding type-1 domain profile	288485	290001	+
		g15628	Glycosyl hydrolases family 18	295552	296358	-
		g15710	Glycosyl hydrolases family 18	675899	677125	+
		g15711	Glycosyl hydrolases family 18	681983	682963	-
		g16070	Glycosyl hydrolases family 18	2106916	2107896	+
		g16071	Glycosyl hydrolases family 18	2112754	2113980	-
		g16127	Glycosyl hydrolases family 18	2384373	2385353	+
		g16128	Chitin-binding type-1 domain profile	2389593	2391109	-
GO:0006281 / GO:0000723	DNA repair / telomere maintenance	g15574	PIF1-like helicase	101447	108284	-
		g15593	Helitron helicase-like domain at N-terminus	164699	171001	-
		g15829	PIF1-like helicase	1172771	1179609	+
		g15869	PIF1-like helicase	1307727	1309605	-
		g15892	PIF1-like helicase	1382184	1389022	+
		g16110	PIF1-like helicase	2318774	2325612	-
GO:0006302	double-strand break repair	g15732	Protein involved in double-strand break repair	808370	809446	+
GO:0006464	cellular protein modification process	g15709	Tubulin-tyrosine ligase domain	669313	670359	-
		g16072	Tubulin-tyrosine ligase domain	2119755	2120552	+
GO:0006487	protein N-linked glycosylation	g15703	protein N-linked glycosylation	635263	635820	+
		g16078	protein N-linked glycosylation	2148719	2149249	-
GO:0006730	one-carbon metabolic process	g15954	Alpha-carbonic anhydrases profile	1627113	1627999	+
GO:0006816	calcium ion transport	g15679	Predicted membrane-bound protein	505326	506780	+
GO:0015074 / GO:0007059	DNA integration / chromosome segregation	g16096	Integrase catalytic domain profile	2261720	2264886	-
GO:0015936	coenzyme A metabolic process	g15638	Hydroxymethylglutaryl-coenzyme A reductases family profile	332081	333238	-
		g16116	Hydroxymethylglutaryl-coenzyme A reductases family profile	2346009	2347166	+
GO:0019684	photosynthesis; light reaction	g15742	Predicted membrane-bound protein	841688	843079	+
GO:0032259	methylation	g15562	FtsJ-like methyltransferase	54702	55857	-

Table S2: Primers used in this study

Primer	Target sequence	Oligo sequence (5'-3') [a]
FP5020	eGFP	aaaGGTACCaAGATCTaACTAGTaCTTAAGcctccggattttgagctttcg
FP4313	eGFP	aaaGGTACCgcgacacgatccagcattaatg
FP4983	HSVtk	aaaAAGCTTaCCTGCAGGaTGATCAaGGTCACCcgcggtggaattcgaattgg
FP4317	HSVtk	aaaAAGCTTgaccatgattacgccaagctcg
FP6031	*SIX6* upstream flank	aaaTTAATTAAccgaagagctggatcgtttgaag
FP6032	*SIX6* upstream flank	aaaACTAGTgatgtgacaggacaagttgatggttc
FP6033	*SIX6* downstream flank	aaaGGCGCGCCctataaagccaatacgattcgaag
FP6106	*SIX6* downstream flank	aaaCCTGCAGGgtagaataacctagatcacgtgc
FP6036	*SIX9* upstream flank	aaaTTAATTAAgtgactactgtggtccttgctg
FP6037	*SIX9* upstream flank	aaaACTAGTgactataggctgaagttagactgg
FP6038	*SIX9* downstream flank	aaaGGCGCGCCggcagagattgtccttacaaac
FP6107	*SIX9* downstream flank	aaaGGTCACCgagttttatcgcatatcatcgtcc
FP6064	*SMP1* upstream flank	aaaTTAATTAAgtaaagagaacacgttgcaaataag
FP6065	*SMP1* upstream flank	aaaACTAGTgaagtttcttgtgtaaaaagtgtgg
FP6066	*SMP1* downstream flank	aaaGGCGCGCCctggcgatgtcgaacagac
FP6115	*SMP1* downstream flank	aaaCCTGCAGGctgtcgagcaggaaaggatac
FP1490	*SIX6* southern probe	CTCTCCTGAACCATCAACTT
FP1491	*SIX6* southern probe	CAAGACCAGGTGTAGGCATT

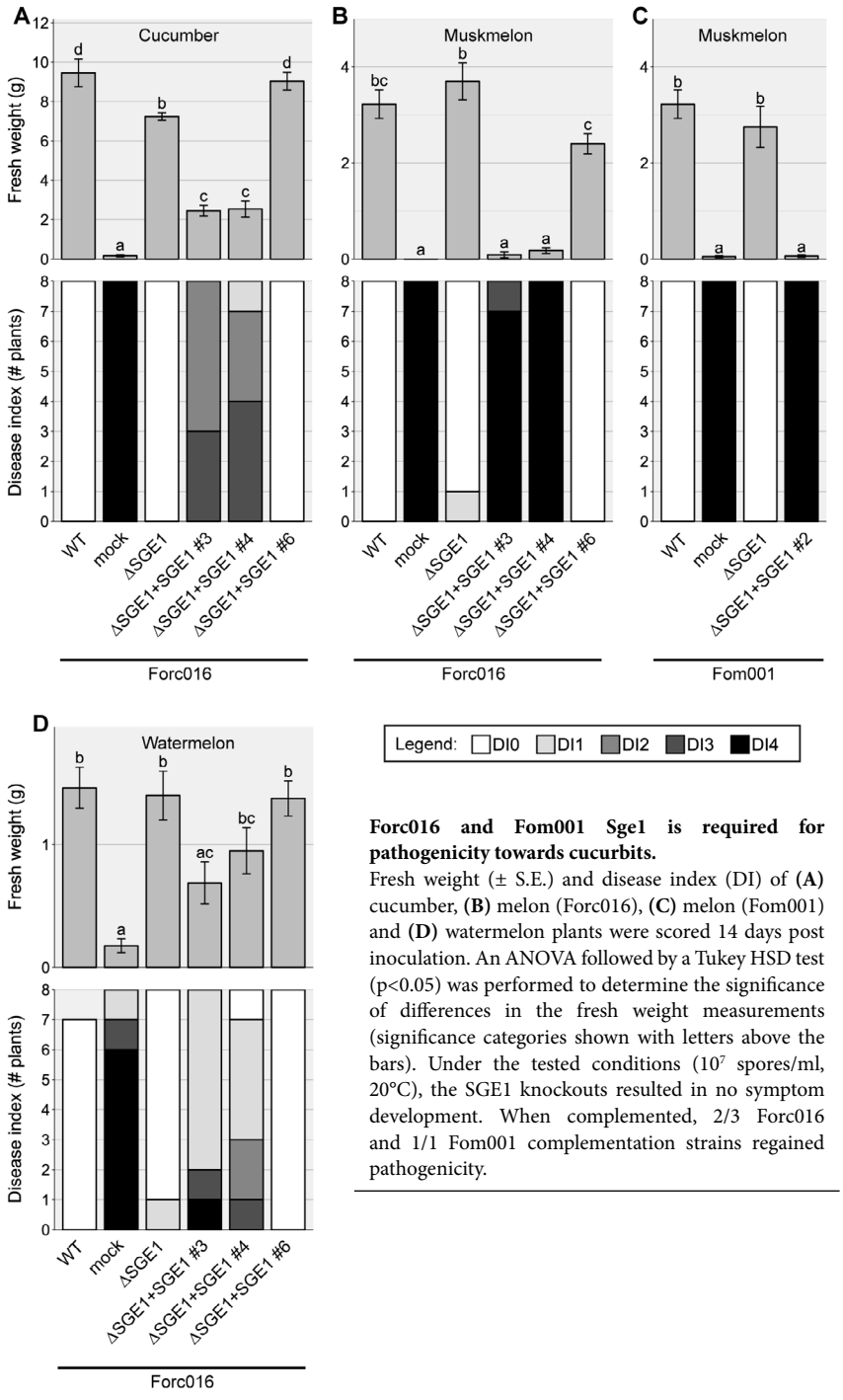
[a] restriction sites are written in uppercase

Appendix A



Horizontal chromosome transfer (HCT) strains #1 and #3, that underwent large scale chromosome rearrangements, are less pathogenic than the other tested HCT strains.

Appendix B



Forc016 and Fom001 Sge1 is required for pathogenicity towards cucurbits.
Fresh weight (± S.E.) and disease index (DI) of **(A)** cucumber, **(B)** melon (Forc016), **(C)** melon (Fom001) and **(D)** watermelon plants were scored 14 days post inoculation. An ANOVA followed by a Tukey HSD test ($p<0.05$) was performed to determine the significance of differences in the fresh weight measurements (significance categories shown with letters above the bars). Under the tested conditions (10^7 spores/ml, 20°C), the SGE1 knockouts resulted in no symptom development. When complemented, 2/3 Forc016 and 1/1 Fom001 complementation strains regained pathogenicity.

Chapter 4

Comparative genomics-based markers: discrimination of host-specificity in *Fusarium oxysporum*

This chapter has been submitted for publication

Authors: Peter van Dam, Mara de Sain, Anneliek ter Horst, Michelle van der Gragt and Martijn Rep

Comparative genomics-based markers: discrimination of host-specificity in *Fusarium oxysporum*

Abstract

The polyphyletic nature of many *formae speciales* of *Fusarium oxysporum* (Fo) prevents molecular identification of newly encountered strains based on conserved, vertically inherited genes. Alternative molecular detection methods that could replace labour- and time-intensive disease assays are therefore highly desired. Effectors are functional elements in the pathogen-host interaction and have been found to show very limited sequence diversity between strains of the same forma specialis, which makes them potential markers for host-specific pathogenicty. We therefore compared candidate effector genes extracted from 60 existing and 22 newly generated genome assemblies, specifically targeting strains affecting cucurbit plant species. Based on these candidate effector genes, in total eighteen PCR primer pairs were designed to discriminate between each of the seven *Cucurbitaceae*-affecting *formae speciales*. When tested on a collection of strains encompassing different clonal lineages of these *formae speciales*, non-pathogenic strains and strains of other formae speciales, they allowed clear recognition of the host range of each evaluated strain. Within Fo f. sp. *melonis* more genetic variability exists than anticipated, resulting in three *melonis* marker patterns that partially overlapped with the cucurbit-infecting *formae speciales cucumerinum, niveum, momordicae* and/or *lagenariae*. For Fo f. sp. *niveum*, a multiplex TaqMan assay was evaluated, which allowed quantitative and specific detection of template DNA quantities as low as 2.5 pg. These results provide ready-to-use marker sequences for the mentioned Fo pathogens. Additionally, the method can be applied to find markers distinguishing other host-specific forms of *F. oxysporum*.

Introduction

Accurate and rapid pathogen detection is necessary to take appropriate action against plant diseases. *Fusarium oxysporum* (Fo) is a soil-borne fungus that includes both non-pathogenic and plant-pathogenic strains. Pathogenic strains of *F. oxysporum* cause vascular wilt and cortical rot disease in a wide variety of agricultural crop species. They are classified into host-specific forms (*formae speciales,* ff. spp.) and are often further subdivided into races based on their capacity to infect different cultivars of a plant species (Armstrong and Armstrong, 1978; Del Mar Jiménez-Gasco and Jiménez-Díaz, 2003; Inami *et al.*, 2012).

Fusarium wilt and root rot in cucurbits are amongst the most prominent and destructive diseases affecting this plant family (Chen *et al.*, 2003; Vakalounakis *et al.*, 2004; Martyn, 2014). In total, seven cucurbit-infecting *formae speciales* have been described: Fo f. sp. *cucumerinum,* f. sp. *melonis,* f. sp. *niveum,* f. sp. *radicis-cucumerinum,* f. sp. *lagenariae,* f. sp. *momordicae* and f. sp. *luffae* (Table 1). The latter three are mostly restricted to South-East Asia (Kim *et al.*, 1993), while the *formae speciales* affecting cucumber, melon and watermelon are globally distributed and more important from an economic standpoint (Kim *et al.*, 1993; Zhang *et al.*, 2005; Martyn, 2014).

Currently, there are no effective curative treatments for *Fusarium* disease (Lievens *et al.*, 2006). Use of resistant varieties or rootstocks is the only practical measure for controlling the disease in the field (Pavlou *et al.*, 2002; Cohen *et al.*, 2014, 2015). In greenhouses, soil sterilization by fumigation with methyl bromide can be performed (Pavlou *et al.*, 2002; Michielse and Rep, 2009). Most efforts are directed towards prevention of the disease. Routine methods that provide reliable subspecific identification, sensitive detection, and accurate quantification of *F. oxysporum* are of high importance (Lievens and Thomma, 2005) and could prevent unnecessary efforts to suppress harmless fungal populations (Van Der Does *et al.*, 2008). Development of these types of markers has thusfar been complicated by the polyphyletic nature of most *formae speciales* of Fo (Lievens and Thomma, 2005).

As many Fo strains have been found to be non-pathogenic, endophytic or even applicable as biocontrol strain (Fravel *et al.*, 2003; Alabouvette *et al.*, 2009; Aimé *et al.*, 2013), discrimination between pathogenic and abundantly present non-virulent strains is very important (Wang *et al.*, 2013). Discrimination of Fo *formae speciales* and races is routinely done through labour- and time-intensive disease assays (Recorbet *et al.*, 2003; Lievens *et al.*, 2007; Covey *et al.*, 2014). Molecular detection methods are therefore highly desired.

Formae speciales are often of polyphyletic origin (Baayen *et al.*, 2000), and pathogenic strains may share a higher level of sequence similarity of conserved genes with strains that are non-pathogenic or pathogenic towards another host (Kistler, 1997; Lievens *et al.*, 2009). Diagnostics based on genes like *Translation Elongation Factor 1-alpha (EF1α)* or the ribosomal intergenic spacer (IGS) are therefore only useful to discriminate between fungal species (Cenis *et al.*, 2003; Haegi *et al.*, 2013). In several cases they have been suggested for

subspecies discrimination, but these often prove to be unreliable for this purpose (Zhang *et al.*, 2005; Lin *et al.*, 2010; Haegi *et al.*, 2013).

Several molecular markers for the cucurbit-infecting ff. spp. *cucumerinum, radicis-cucumerinum, niveum and luffae* have been developed. These are all based on Random Amplified Polymorphic DNA (RAPD) fragments, resulting in Sequence Characterized Amplified Region (SCAR) markers. SCAR markers are suboptimal for *forma specialis* discrimination because they are based on genomic regions that are not necessarily required for virulence. Furthermore, as they can be localized anywhere on the genome, there is often little to no sequence data available in public databases for comparison with other sequences. The robustness of the markers can only be verified by screening against a large collection of strains (Lievens and Thomma, 2005).

Interestingly, closer inspection of previously developed f. sp.-distinguishing SCAR markers showed that the selected sequences were often (part of) a transposable element, such as *Fot1* (Fo f. sp. *albedinis, chrysantemi, dianthi*), *Folyt1* (Fo f. sp. *radicis-cucumerinum*) and *Impala* (Fo f. sp. *ciceris, dianthi)* (Lievens *et al.*, 2008) or pathogenicity-associated genes like *FTF1* (Fo f. sp. *phaseoli*) (Alves-Santos *et al.*, 2002). A race 1 specific Fo f. sp. *lactucae* marker was developed by amplifying and cloning regions between long terminal repeats of retrotransposons in the genome (Pasquali *et al.*, 2007). For the ff. spp. *lagenariae, momordicae* and *melonis,* only DNA fingerprinting results have been described thus far (Namiki *et al.*, 1994).

It was recently shown that host-specificity is associated with the suite of effector genes present in the genome of *F. oxysporum* strains (van Dam *et al.*, 2016). Both presence-absence polymorphisms as well as the sequence type of individual effector genes turned out to be predictive for a strain's host range. These genes therefore form the most solid base for discrimination of *formae speciales* within the *F. oxysporum* species complex (FOSC) (Recorbet *et al.*, 2003; Lievens and Thomma, 2005; Lievens *et al.*, 2009). Indeed, use of virulence genes to identify fungal plant pathogens has proven successful in the past for other *Fusarium* species (Hogg *et al.*, 2007; Mbofung *et al.*, 2011). Within the FOSC, this approach has been applied to distinguish Fo f. sp. *cubense* tropical race 4 by targeting a candidate effector gene (Aguayo *et al.*, 2017). Additionally, Fo f. sp. *lycoperici* and Fo f. sp. *cubense* can be discriminated from other *formae speciales* through the use of polymerase chain reaction (PCR) primers designed to detect specific *Secreted In Xylem (SIX)* effector gene sequences (Van Der Does *et al.*, 2008; Lievens *et al.*, 2009; Fraser-Smith *et al.*, 2014). At the time of these studies however, no (or limited) comparative genomics analyses could be performed due to the lack of available genome sequences. All *SIX* genes have homologs in other host-pathogenic forms of Fo (e.g. *SIX1*, *SIX5* and *SIX6* (Chakrabarti *et al.*, 2011; Thatcher *et al.*, 2012; Taylor *et al.*, 2015; van Dam *et al.*, 2016)). For these marker-specificity could not be evaluated beforehand and cross-reaction with non-target *formae speciales* was found (Lievens *et al.*, 2009).

Since it is not yet viable to sequence every individual strain encountered, we decided to design effector candidate-based markers. In this way, we aimed to be able to distinguish cucurbit-

affecting ff. spp. from i) each other, ii) other *formae speciales* and iii) non-pathogenic strains. Therefore, we used whole-genome sequences of a number of representative cucurbit-infecting Fo strains as a starting point and identified putative effector genes suitable as markers. An advantage of using molecular markers over whole genome sequencing is that they can also be applied to infected soil or plant tissue samples; the fungus does not need to be isolated and cultured (Lin *et al.*, 2010). Techniques such as TaqMan Real-Time PCR even allow for a quantitative evaluation of pathogen abundance, e.g. on DNA isolated from soil (Huang *et al.*, 2016).

The genetic basis for host-specificity of FOSC strains towards plants belonging to the *Cucurbitaceae* family is similar (van Dam *et al.*, 2016), making these *formae speciales* relatively difficult to separate. This means that this is a good test case for host-specificity discrimination and the results presented here can be exemplary for application to other plant species where disease caused by *F. oxysporum* is a pressing problem.

Table 1: *Formae speciales* of *F. oxysporum* affecting members of the *Cucurbitaceae* family.
A five-letter abbreviation was used to distinguish between *formae speciales* in the species complex.

Forma specialis	Abbreviation	Host	Reference
cucumerinum	Focuc	Cucumber *(Cucumis sativus)*	Owen, 1956
melonis	Fomln	Muskmelon *(Cucumis melo)*	Leach *et al.*, 1937
niveum	Foniv	Watermelon *(Citrullus lanatus)*	Armstrong and Armstrong, 1981
radicis-cucumerinum	Forcu	Multiple cucurbits (including cucumber, melon, watermelon, gourd)	Vakalounakis, 1996
momordicae	Fomom	Bitter gourd *(Momordica charantia)*	Sun and Huang, 1982
lagenariae	Folag	Calabash gourd *(Lagenaria spp.)*	Matuo and Yamamoto, 1967; Namiki *et al.*, 1994
luffae	Foluf	Sponge gourd *(Luffa cylindrica)*	Kim *et al.*, 1993

Results

Several cucurbit-infecting *formae speciales* have a polyphyletic origin

In order to be able to select *forma specialis*-wide marker sequences, it is necessary to collect the genetic variety for that *forma specialis* as completely as possible. We made use of 66 previously published genome sequences and added *de novo* genome assemblies generated from Illumina paired-end read data of 22 new strains (see Supplementary data S1).

Vegetative compatibility groups (VCGs) tend to represent clonal lineages (Correll, 1991; Leslie, 1993; Gordon and Martyn, 1997; Kistler, 1997; Baayen *et al.*, 2000), but some exceptions are known. For example, Fo f. sp. *radicis-cucumerinum* was described as having two VCGs. RAPD fingerprinting analyses and phylogenetic studies based on conserved genes, however, showed that the two VCGs are very similar and appear to be clonally related (Vakalounakis

and Fragkiadakis, 1999; Lievens *et al.*, 2007; van Dam *et al.*, 2016).
In order to assess the genetic diversity of the *formae speciales* under investigation, we generated a phylogenetic tree based on over 400 core genomic gene sequences from each of their genomes (Fig. 1). This showed that the *formae speciales cucumerinum, melonis, niveum* and *lagenariae* occupied multiple clades in the tree (5, 3, 3 and 3, respectively), indicating that they belong to different clonal lines. In our set of strains, we have 6 of 7 described *cucumerinum* VCGs (Vakalounakis and Fragkiadakis, 1999; Vakalounakis *et al.*, 2004), 3 of 9 *melonis* VCGs (Mirtalebi and Banihashemi, 2014), all 3 *niveum* VCGs (Katan, 1999) and both *radicis-cucumerinum* VCGs (Katan, 1999). For the *formae speciales lagenariae* (3 VCGs described (Cumagun *et al.*, 2008)), *momordicae* (4 VCGs described (Cumagun *et al.*, 2008)) and *luffae* (unknown number of VCGs), no VCG information was available for our strains, although they group into three, one and one clade(s), respectively (Fig. 1).

Candidate effector gene phylogenies display clear grouping of host specificity

Unlike conserved core genes, virulence-related genes tend to be identical across members belonging to the same polyphyletic *forma specialis* of *F. oxysporum* (Van Der Does *et al.*, 2008; van Dam *et al.*, 2016). For this reason, they have predictive value for a strain's host range. *Forma specialis* markers are essentially the smallest possible set of effector genes that is shared by all strains of a *forma specialis* and absent or different in sequence (at least as a set) in all other strains (van Dam *et al.*, 2016).

We extracted the sequences plus 150 nucleotides up- and downstream of the open reading frame of each of the described candidate effector genes from van Dam *et al.* (2016) and generated a multiple sequence alignment (MSA, Supplementary Data S2) and phylogenetic tree for each of them (three examples in Fig. 2, continued in Supplementary data S3). A custom python script identified those genes in which *all* members of a *forma specialis* grouped together in a separate clade. From the genes displaying such grouping, the genes that facilitated the best discrimination were selected based on manual inspection of the MSA to come to a final selection of marker sequences per *forma specialis* (Table 2).
Some of the selected genes show multiple *forma specialis*-specific clades, therefore multiple markers targeting different *forma specialis* could be designed on these genes. An example is candidate effector #99, a hypothetical protein-encoding gene that is used as a marker for Foniv, Folag, Focuc and Foluf (Fig. 2B). *Melonis* strain Fomln010 possesses an identical copy to both candidate effector #99 homologs present in the Foniv strains as well as a copy identical to the Folag gene sequence. To still be able to distinguish these *formae speciales* from one another, it is therefore of importance to use multiple markers for each *forma specialis*.

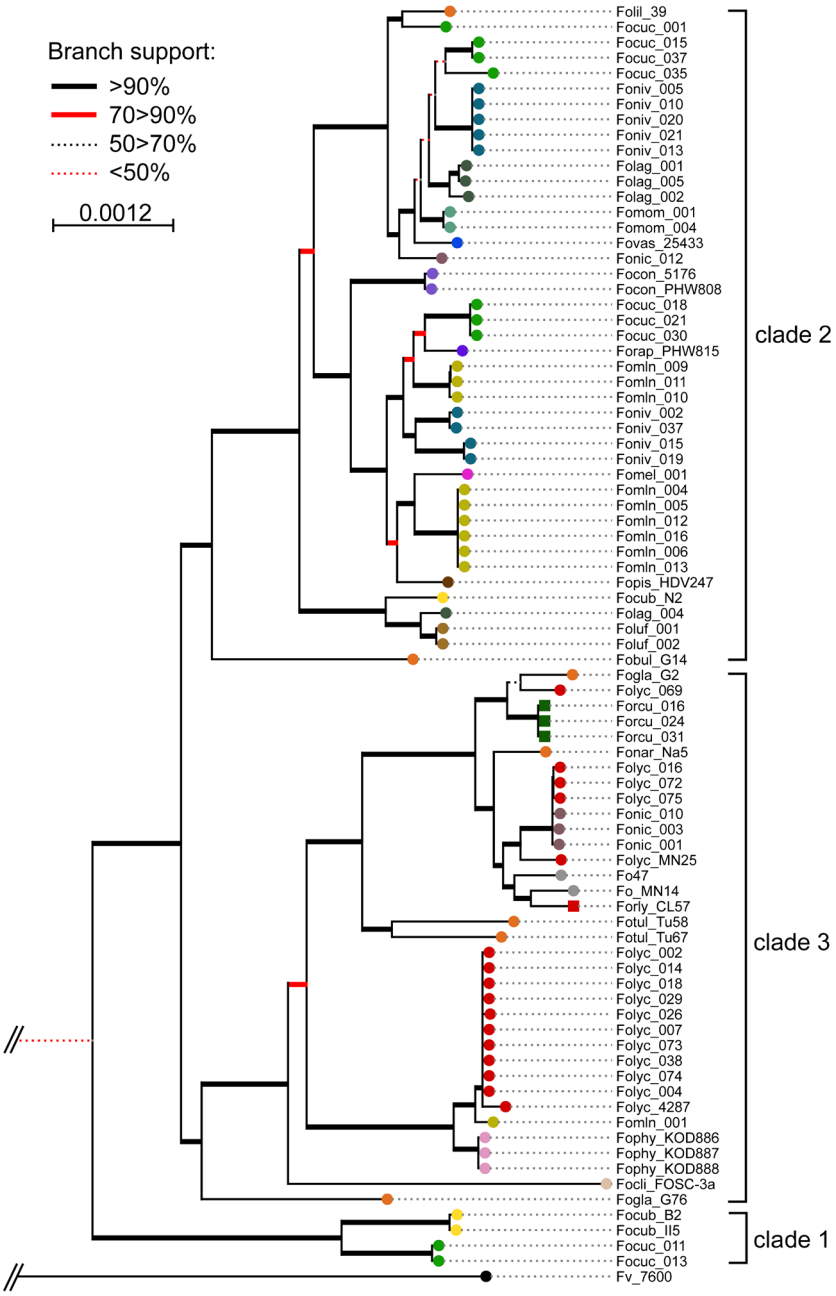


Fig. 1: *Formae speciales cucumerinum, melonis, niveum* and *lagenariae* are of polyphyletic origin. 441 conserved core genes from all genomes were extracted, aligned and concatenated into a multiple sequence alignment. Phylogeny was inferred with 100 bootstrap iterations. All strains fall within the three main clades of the FOSC. (Focuc= Fo f. sp. *cucumerinum*; Fomln= Fo f. sp. *melonis*; Foniv= Fo f. sp. *niveum*; Forcu= Fo f. sp.*radicis-cucumerinum*; Folag= Fo f. sp. *lagenariae; Foluf*= Fo f. sp. *luffae*; Fomom=Fo f. sp. *momordicae*. For abbreviations of other *formae speciales,* see Supplementary data S1).

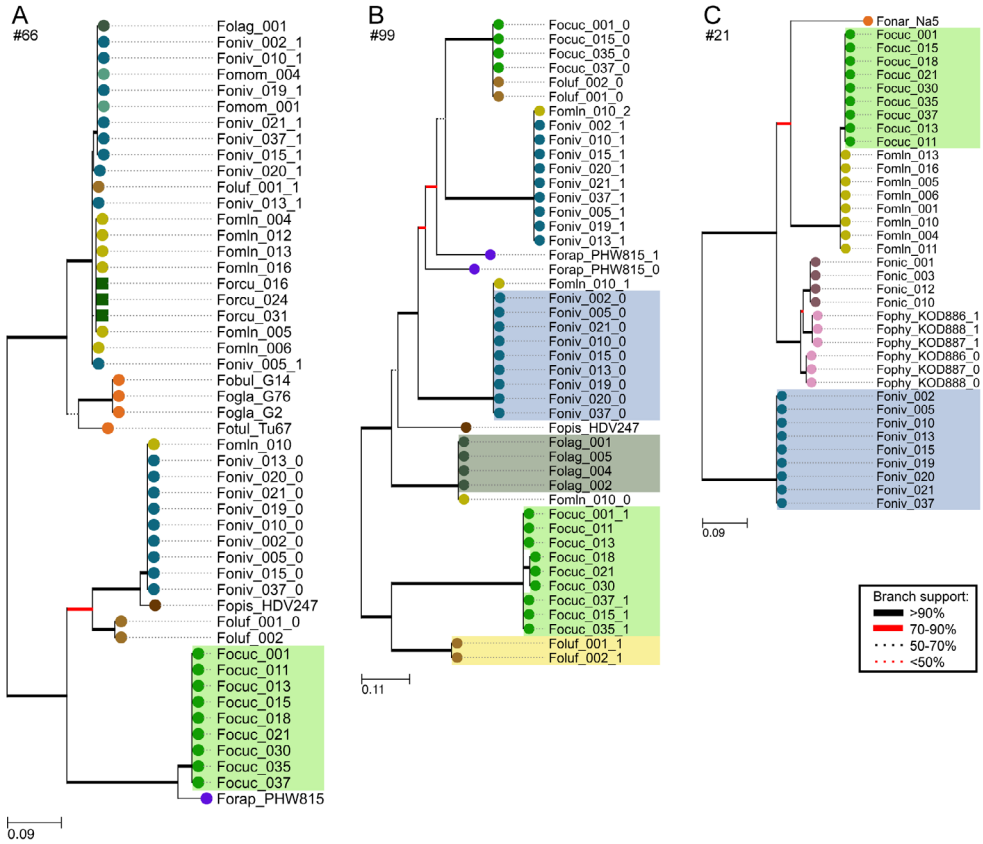


Fig. 2: Phylogenetic trees of three genes selected as marker for Fo f. sp. *cucumerinum:* (A) #66, (B) #99, (C) #21.

Separation of a clade that includes all strains belonging to a *forma specialis* indicates sequence similarity within and sequence dissimilarity between *formae speciales*. Coloured highlights in the tree reflect the target f. sp. of the marker. Hypothetical-protein encoding genes #99 and #21 are used as markers for multiple *formae speciales*.

Discrimination of cucurbit-infecting ff. spp. by PCR

PCR primers were designed specifically on polymorphic regions of the selected DNA sequences (Table 2, Supplementary Data S2), aiming to generate a PCR product sized above 120 and below 700 nt for quick and reliable application. *Fusarium Extracellular Matrix 1 (FEM1)* (Schoffelmeer *et al.*, 2001) was taken along as a positive control. To verify the applicability of the markers, PCRs were executed for each of the primer pairs on a subset of the strains that were used for marker design, i.e. of which the host range has been confirmed and the genome had been sequenced. This included strains belonging to the cucurbit-infecting *formae speciales*, several other ff. spp. (*vasinfectum, Arabidopsis*-infecting, *lycopersici, radicis-lycopersici, nicotianae, melongenae, Physalis*-infecting, *cubense, pisi, tulipae, gladioli*) and two non-pathogenic *F. oxysporum* strains: Fo47 (Aimé *et al.*, 2013) and MN14 (van Dam *et al.*, 2016). The strains were selected based on their differential phylogenetic distribution in Fig. 1

Table 2: Selected marker genes and their respective target *forma specialis*.

Gene ID	Target gene[a]	Target *forma specialis*	Gene tree
Positive control	*FEM1*	positive control	-
94	HPEG	all cucurbit-infecting	Suppl. data S3A
13	*SIX13*	*radicis-cucumerinum*	Suppl. data S3B
70	HPEG	*radicis-cucumerinum*	Suppl. data S3C
66	HPEG	*cucumerinum*	Fig. 2A
99	HPEG	*cucumerinum*	Fig. 2B
21	Fom Effector 7	*cucumerinum*	Fig. 2C
1	*SIX1*	*melonis*	Suppl. data S3D
20	Fom Effector 6	*melonis*	Suppl. data S3E
18	Fom Effector 3	*melonis* + *niveum*	Suppl. data S3G
99	HPEG	*niveum*	Fig. 2B
100	HPEG	*niveum*	Suppl. data S3H
21	Fom Effector 7 (pseudogenized)	*niveum*	Fig. 2C
98	HPEG	*momordicae*	Suppl. data S3I
130	HPEG	*momordicae*	Suppl. data S3J
1	*SIX1*	*lagenariae* + *momordicae*	Suppl. data S3D
71	HPEG	*lagenariae*	Suppl. data S3F
99	HPEG	*lagenariae*	Fig. 2B
99	HPEG	*luffae*	Fig. 2B

[a] HPEG: Hypothetical protein encoding gene

as well as presence and absence of selected marker sequences in their genome assembly.

All except one of the *forma specialis*-specific PCR markers behaved like expected (Table 3), showing PCR products only in the expected combinations of genomic DNA and marker primers. One false-positive PCR product was found, in the combination of Fo f. sp. *pisi* HDV247 and marker #130 (Fomom). In the genome assembly of HDV247, this gene was found to be present with 97% sequence similarity, although the downstream region of this gene provided sufficient sequence diversity for primer design (Supplementary data S2).

Marker #94, targeting all cucurbit-affecting *formae speciales,* gave a band of the correct size for all cucurbit-affecting isolates tested, except Fomln010. Furthermore, strain Fomln010 displayed an atypical *F. oxysporum* f. sp. *melonis* marker pattern, as it yielded PCR products that were not seen in the other Fomln isolates for markers Foniv #99, Foniv #100 and Folag #99. This pattern, designated in Table 3 as "B", was not unexpected, since presence of identical sequences for these three markers as well as absence of the gene encoding hypothetical protein #94 had been observed in the genome assembly of Fomln010 (sequence similarity of marker #99 shown in Fig. 2B). The presence of identical effector candidate sequences across *formae speciales* affecting similar plant species was not surprising since they share part of their genetic toolset allowing for pathogenic colonization of these plants (van Dam *et al.*, 2016). It does, however, make marker selection more challenging. While screening for specific differentiation of, for instance, Fomln and Foniv, it is therefore important to check multiple markers.

Table 3: PCR Markers allow discrimination of cucurbit-affecting *formae speciales* of *F. oxysporum*
The following symbols are used in the table: + for positive test result, - for negative test result, ± for weak positive test result (very faint PCR product of the expected size present).

Strain	*forma specialis*	FEM (+)	94 cucurbits	13 Forcu	70 Forcu	66 Focuc	99 Focuc	21 Focuc	1 Fomln	20 Fomln	18 Fomln + Foniv	99 Foniv	100 Foniv	21 Foniv	98 Fomom	130 Fomom	1 Fomom + Folag	71 Folag	99 Folag	99 Foluf	Marker profile
Forcu016	*radicis-cucumerinum*	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*radicis-cucumerinum*
Forcu031	*radicis-cucumerinum*	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*radicis-cucumerinum*
Focuc001	*cucumerinum*	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	*cucumerinum*
Focuc013	*cucumerinum*	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	*cucumerinum*
Focuc015	*cucumerinum*	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	*cucumerinum*
Focuc018	*cucumerinum*	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	*cucumerinum*
Focuc035	*cucumerinum*	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	*cucumerinum*
Fomln001	*melonis*	+	+	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	*melonis "A"*
Fomln006	*melonis*	+	+	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	*melonis "A"*
Fomln010	*melonis*	+	-	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-	+	-	*melonis "B"*
Foniv002	*niveum*	+	+	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	*niveum*
Foniv010	*niveum*	+	+	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	*niveum*
Foniv015	*niveum*	+	+	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	*niveum*
Foniv020	*niveum*	+	+	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	*niveum*
Fomom001	*momordicae*	+	+	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	*momordicae*
Fomom004	*momordicae*	+	+	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	*momordicae*
Folag001	*lagenariae*	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	*lagenariae*
Folag004	*lagenariae*	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	*lagenariae*
Foluf001	*luffae*	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	*luffae*
Foluf002	*luffae*	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	*luffae*
Fovas 25433	*vasinfectum*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Focon5176	*conglutinans*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Folyc 4287	*lycopersici*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Folyc MN25	*lycopersici*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Folyc069	*lycopersici*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Folyc072	*lycopersici*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Forly CL57	*radicis-lycopersici*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Fonic001	*nicotianae*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Fomel001	*melongenae*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Fophy KOD886	*physali*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Focub II5	*cubense*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Fopis HDV247	*pisi*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	±	-	-	-	-	-
Fotul Tu67	*tulipae*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Fogla G76	*gladioli*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Fo47	non-pathogenic	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FoMN14	non-pathogenic	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H_2O	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Evaluating *forma specialis* classification using markers

After testing the markers on sequenced strains to verify that they worked as anticipated, an extended set of strains originating from around the world (strain info in Supplementary data S1) was screened. Most strains were isolated from *Fusarium*-affected cucurbit plants and were described as one of the pathogenic forms listed in Table 1. The aim was to either confirm or reject their reported host-specificity with our markers. A number of strains isolated from non-cultivated soil samples was also taken along. The expectation was that these non-specialized strains do not possess many effector genes and therefore would test negative for all of the 18 markers.

As can be seen from Table 4, marker analysis confirmed the reported *forma specialis* of most strains that were tested. However, some strains behaved differently than expected. For

example, PCR products were identified for Fomln017, 021, 024 and 026 for *cucumerinum* markers #66 and #21, as well as *momordicae/lagenariae* marker #1. Intriguingly, none of the *melonis* markers tested positive in these strains (marker pattern "C" in Table 4). Additionally, a third *melonis* pattern was observed with strain Fomln023 that was nearly identical to the pattern of Fomln010 (pattern "B", Table 3). If these strains are indeed pathogenic on melon plants, this is a clear indication that there is more genetic diversity present amongst strains belonging to Fo f. sp. *melonis* than anticipated during marker design. Finally, Fomln marker #1 cross-reacted with Foniv041 genomic DNA, showing that this marker is not 100% specific for Fomln.

Cucurbit marker #94 did not test positive for four individual Fo f. sp. *melonis* strains and one Fo f. sp. *momordicae* strain. However, while it does not detect all cucurbit-infecting strains, it did not result in false positives (Table 3 and 4), meaning that it can still be used in addition to the other f. sp.-specific markers.

Several strains showed a marker pattern typically observed for another *forma specialis,* indicating that their reported host specificity might not be accurate. Strains Focuc014 and Focuc040, reported as Fo f. sp. *cucumerinum*, clearly showed a positive result for both *radicis-cucumerinum* markers and absence of all three *cucumerinum* markers, suggesting that they are in fact f. sp. *radicis-cucumerinum* strains. Another interesting candidate was strain 14150, reportedly an isolate belonging to Fo f. sp. *'cucurbitacearum',* a *forma specialis* proposed to encompass all *formae speciales* affecting cucurbits (Gerlagh and Blok, 1988). This strain also showed the marker pattern typically observed for Fo f. sp. *radicis-cucumerinum*. Four strains (one reported as *cucumerinum* and three as *niveum* strains) displayed absence of all 18 markers tested, while another strain (reported as *cucumerinum*) tested positive only for Focuc marker #21, suggesting that they are not capable of infecting any of the cucurbit plants. As expected, each of the environmental strains tested negative for all of the markers.

Disease assays confirm marker predictions

The strains of which the reported *forma specialis* did not match the marker pattern were tested in a bioassay on susceptible cucurbit varieties to evaluate their actual host range (Table 4, right column; Supplementary data S4). Strains Focuc014, Focuc040 and 14150 caused severe crown rot symptoms both in cucumber and melon, meaning that they are in fact Fo f. sp. *radicis-cucumerinum* strains, as predicted by our PCR analysis. The strains that were predicted to be non-pathogenic based on their marker pattern indeed did not cause symptom development when tested on susceptible cucumber (Focuc028, Focuc039) or watermelon (Foniv034, Foniv035, Foniv038) plants. Strains Fomln017, 021, 024, 026 (profile C), as well as 023 (profile B) were all able to cause disease in susceptible musk melon plants, even though their marker pattern was different from the most common profile in our set of isolates (profile A, Tables 3 and 4). Fomln023, which tested positive for two of the three *niveum* markers, was also tested

Table 4: PCR testing the markers on a set of 48 world-wide isolates verifies their reported *forma specialis* in most cases.

The following symbols are used in the table: + for positive test result, - for negative test result, ± for weak positive test result (very faint PCR product of the expected size present).

Strain	Reported f.sp.	FEM (+)	94 cucurbits	13 Forcu	70 Forcu	66 Focuc	99 Focuc	21 Focuc	1 Fomln	20 Fomln	18 Fomln + Foniv	99 Foniv	100 Foniv	21 Foniv	98 Fomom	130 Fomom	1 Fomom + Folag	71 Folag	99 Folag	99 Foluf	Marker profile [a]	Bio-assay result [b]
Focuc014	*cucumerinum*	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*radicis-cucumerinum*	*radicis-cucumerinum*
Focuc040	*cucumerinum*	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*radicis-cucumerinum*	*radicis-cucumerinum*
Forcu005	*radicis-cucumerinum*	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*radicis-cucumerinum*	not tested
Forcu017	*radicis-cucumerinum*	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*radicis-cucumerinum*	not tested
Forcu020	*radicis-cucumerinum*	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*radicis-cucumerinum*	not tested
Forcu028	*radicis-cucumerinum*	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*radicis-cucumerinum*	not tested
Forcu029	*radicis-cucumerinum*	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*radicis-cucumerinum*	not tested
14150	*cucurbitacearum*	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*radicis-cucumerinum*	*radicis-cucumerinum*
Focuc009	*cucumerinum*	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	*cucumerinum*	not tested
Focuc010	*cucumerinum*	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	*cucumerinum*	not tested
Focuc016	*cucumerinum*	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	*cucumerinum*	not tested
Focuc017	*cucumerinum*	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	*cucumerinum*	not tested
Focuc026	*cucumerinum*	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	*cucumerinum*	not tested
Focuc027	*cucumerinum*	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	*cucumerinum*	not tested
Focuc036	*cucumerinum*	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	*cucumerinum*	not tested
Focuc038	*cucumerinum*	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	*cucumerinum*	not tested
Fomln018	*melonis*	+	+	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	*melonis "A"*	not tested
Fomln019	*melonis*	+	+	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	*melonis "A"*	not tested
Fomln020	*melonis*	+	+	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	*melonis "A"*	not tested
Fomln027	*melonis*	+	+	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	*melonis "A"*	not tested
Fomln025	*melonis*	+	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	*melonis "A"*	*melonis*
Fomln002	*melonis*	+	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	*melonis "A"*	non-pathogenic
Fomln003	*melonis*	+	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	*melonis "A"*	not tested
Fomln023	*melonis*	+	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	+	-	*melonis "B"*	*melonis/niveum*
Fomln024	*melonis*	+	+	-	-	+	-	+	-	-	+	-	-	-	-	-	+	-	-	-	*melonis "C"*	*melonis*
Fomln017	*melonis*	+	+	-	-	+	-	+	-	-	-	-	-	-	-	-	+	-	-	-	*melonis "C"*	*melonis*
Fomln021	*melonis*	+	+	-	-	+	-	+	-	-	-	-	-	-	-	-	+	-	-	-	*melonis "C"*	*melonis*
Fomln026	*melonis*	+	+	-	-	+	-	+	-	-	-	-	-	-	-	-	+	-	-	-	*melonis "C"*	*melonis*
Foniv011	*niveum*	+	+	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	*niveum*	not tested
Foniv017	*niveum*	+	+	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	*niveum*	not tested
Foniv018	*niveum*	+	+	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	*niveum*	not tested
Foniv033	*niveum*	+	+	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	*niveum*	not tested
Foniv039	*niveum*	+	+	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	*niveum*	not tested
Foniv040	*niveum*	+	+	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	*niveum*	not tested
Foniv041	*niveum*	+	+	-	-	-	-	-	±	-	+	+	+	+	-	-	-	-	-	-	*niveum*	not tested
Fomom002	*momordicae*	+	+	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	*momordicae*	not tested
Fomom003	*momordicae*	+	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	*momordicae*	not tested
Folag003	*lagenariae*	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	*lagenariae*	not tested
Folag006	*lagenariae*	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	*lagenariae*	not tested
Folag007	*lagenariae*	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	*lagenariae*	not tested
Folag008	*lagenariae*	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	*lagenariae*	not tested
Focuc022	*cucumerinum*	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	non-pathogenic
Focuc028	*cucumerinum*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	non-pathogenic
Focuc039	*cucumerinum*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	non-pathogenic
Foniv034	*niveum*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	non-pathogenic
Foniv035	*niveum*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	non-pathogenic
Foniv038	*niveum*	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	non-pathogenic
RBG1687	environmental	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	not tested
RBG1693	environmental	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	not tested
RBG5713	environmental	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	not tested
RBG5786	environmental	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	not tested
RBG5789	environmental	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	not tested
RBG5791	environmental	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	not tested
RBG5798	environmental	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	not tested
RBG5820	environmental	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	not tested
RBG5824	environmental	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	not tested
RBG5827	environmental	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	non-pathogenic	not tested
H_2O	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		

[a]: "non-pathogenic" means not pathogenic towards any of the seven *formae speciales* listed in Table 1.

[b]: "non-pathogenic" means no symptom development in susceptible host plants of the originally reported f. sp.

on susceptible watermelon plants. This strain was found to also be capable to cause disease in these plants, whereas Fomln017, with an almost identical marker pattern, was not. Fomln002 did not cause symptoms in susceptible melon plants, showing that possessing effector gene sequences alone is not always sufficient for pathogenicity and false positives may show up.
The fact that the bioassay data confirmed the suspected *forma specialis* predicted by the reported markers indicates that they provide a robust tool for identifying whether an isolate indeed belongs to the suspected *forma specialis* or not. PCR cross-reaction between *cucumerinum, melonis* and *niveum* markers and the cross-pathogenicity of strain Fomln023 suggest a shared evolutionary origin of the *formae speciales* affecting cucumber, melon and watermelon.

Specific detection of Fo f.sp. *niveum* using a TaqMan assay

TaqMan Real-Time PCR has added benefits over traditional PCR: samples can easily be multiplexed, the fluorescent probe provides additional sequence specificity and the technique allows for quantification of a target DNA sequence, for example on DNA isolated from soil or diseased plant tissue. A TaqMan experiment was conducted using two of the marker genes in this study, Foniv #21 and #100. These markers showed good specificity and displayed no cross-reaction with non-target strains in the PCR tests (Tables 3 and 4). Taqman-specific primers and probes were designed in such a way that a 116 and 138 bp Fo f. sp. *niveum* specific amplicon was formed, respectively. As a fluorescent dye, HEX ($\lambda_{emission}$ = 556nm) was used for marker #21 and FAM ($\lambda_{emission}$ = 518nm) was used for #100. As an internal control for sample/DNA quality that will allow for normalization of the tested markers during multiplexing experiments, a set of primers and a probe with a different fluorescent dye (TAMRA, $\lambda_{emission}$ = 580nm) was designed on a region of *EF1α* conserved in all Fo strains. To test the efficiency of the primers and probe sets, a dilution series of Fo genomic DNA was made and used as template in a TaqMan assay.
A linear relationship was found between Foniv002 genomic DNA concentration and real-time quantification cycles (Fig. 3; $R^2_{marker\#21}$ = 0.999, $R^2_{marker\#100}$ = 0.998, $R^2_{EF1\alpha}$ = 0.999). The pathogen could be detected at template concentrations as low as ~2.5 pg (Fig. 3).
The TaqMan assay was performed on isolates for which the marker genes #21 and #100 were identified in the genome assembly (Fig. 2C, Supplementary data S3H). Each sequence type was included, with the addition of strains of *F. proliferatum* and *Fusarium* sp. that were identified to have candidate effector #100 in a recent study (van Dam and Rep, 2017). Fo f. sp. *lycopersici* 4287, Fo f.sp. *cubense* II5 and biocontrol strain Fo47 were included as negative controls, since these do not have either of the marker genes. No cross-reactions were found, except Fomln010 that possesses an identical gene sequence to Fo f. sp. *niveum* isolates of marker #100. These results show the applicability of the TaqMan assay for specific detection of Fo f. sp. *niveum* DNA in very small quantities.

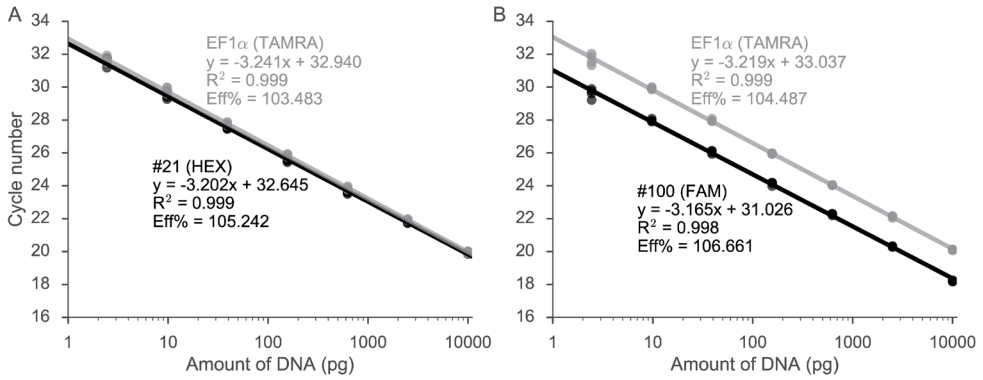


Fig. 3: Taqman assays on effector candidates can detect genomic DNA quantities as low as 2.5 pg. Standard curves for DNA quantification were generated using a 4-fold serial dilutions of Fo f. sp. *niveum* genomic DNA with (A) marker #21, (B) marker #100. Both standard curves were generated in duplex with the control gene, *EF1α*. Four technical replicates were used per dilution.

Discussion

In the current study we tried to make use of comparative genomics to design robust markers based on candidate effector genes. Effectors are functional elements in the pathogen-host interaction and have been found to show very limited sequence diversity between members of the same *forma specialis* (Rocha *et al.*, 2015; Gordon, 2017). This means that they form ideal targets for marker design (Recorbet *et al.*, 2003). Effector gene sequences are often different between *formae speciales*, although several cases of identical gene sequences have been found in a previous study in our lab (van Dam *et al.*, 2016). For example, the *SIX6* and *SIX11* homologs present in some isolates belonging to ff. spp. *niveum, melonis* and *radicis-cucumerinum* have 100% nucleotide identity. These sequences can therefore not be used for differentiation of these ff. spp. They do, however, give insight into the evolutionary history of pathogenicity of *F. oxysporum* towards cucurbits; presence of sequences that are completely identical between relatively distantly related strains implies recent horizontal transfer events of genetic material.

The benefit of using comparative genomics for marker design is that the specificity of the designed markers can directly be evaluated in other genome assemblies (as opposed to RAPD-derived marker sequences). Within the FOSC, one study has reported the use of comparative genomics for f. sp.-marker development. This resulted in markers based on unique (random) sequences distinguishing Fo f. sp. *conglutinans* from nineteen other *formae speciales* of *F. oxysporum* (Ling *et al.*, 2016).

Our goal was to differentiate between *formae speciales* affecting the *Cucurbitaceae* family. The respective hosts are highly similar to each other and incidental cross-pathogenicity between these *formae speciales* has been described (Mcmillan, 1986; Cafri *et al.*, 2005; Zhou and Everts, 2007). We designed a set of 18 primer pairs aiming to discriminate seven cucurbit-

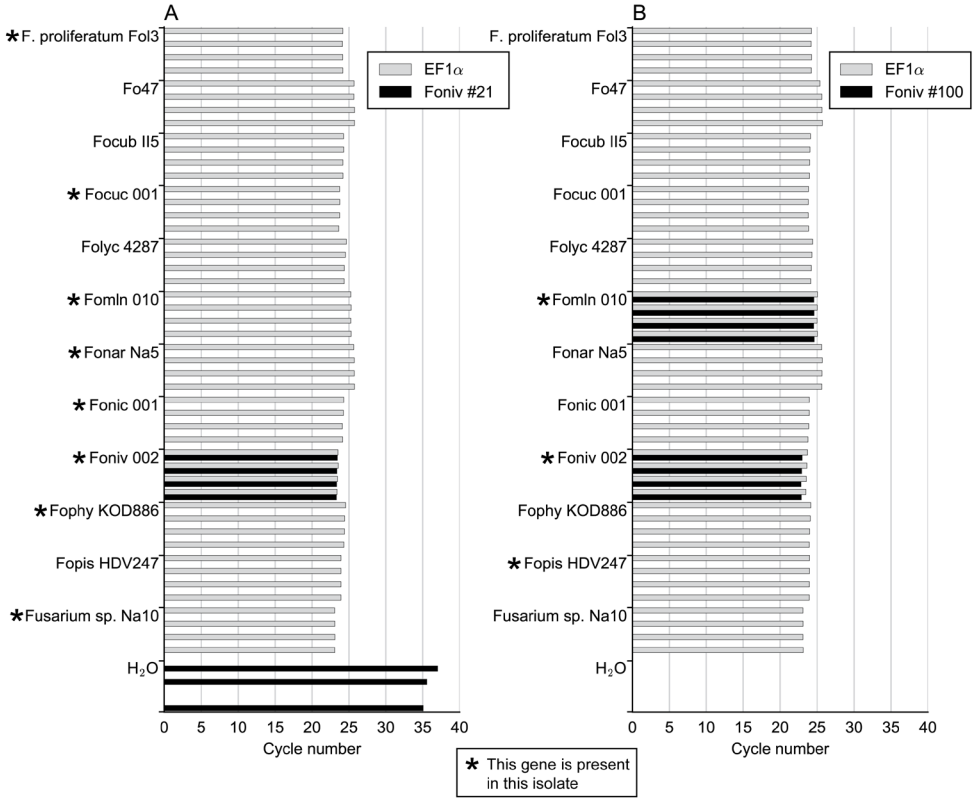


Fig. 4: TaqMan primer-probe combinations show amplification of *niveum* DNA (Foniv002) when markers (A) #21 and (B) #100 are tested in duplex with *EF1α*.
No amplification of these markers was detected in any of the non-*niveum* strains, with the exception of Fomln010 that has an identical gene sequence for hypothetical protein encoding gene #100. High Ct values (≥ 35 cycles) under the detection threshold in the water control of #21 are probably caused by primer-dimer formation under the absence of template DNA. Four technical replicates were used per sample, each represented by a bar in this graph (Focub: Fo f. sp. *cubense*, Focuc: Fo f. sp. *cucumerinum*, Folyc: Fo f. sp. *lycopersici*, Fomln: Fo f. sp. *melonis*, Fonar: Fo f. sp. *narcissi*, Fonic: Fo f. sp. *nicotianae*, Foniv: Fo f. sp. *niveum*, Fophy: Fo f. sp. *physalis*, Fopis: Fo f. sp. *pisi*).

infecting *formae speciales* from each other as well as from other host-specific forms and non-pathogenic strains of *F. oxysporum*. We found that for the *formae speciales cucumerinum, radicis-cucumerinum, niveum, lagenariae, momordicae* and *luffae* the marker sets allowed clear recognition of the host range of each evaluated strain. Marker #94, designed on a gene encoding a hypothetical protein present in all cucurbit-infecting *formae speciales,* was positive for all target strains with the exception of several *melonis* strains and one *momordicae* strain. This gene was not identified in the genome sequence of Fomln010.

Within Fo f. sp. *melonis,* more genetic variability exists than what had been taken into account as a starting point used for marker design (the ten *melonis* strains with a sequenced genome). Several *melonis* strains showed overlap in their effector gene content with cucurbit-

infecting *formae speciales cucumerinum, niveum, momordicae* and/or *lagenariae* (Table 4). So far, no SCAR or other marker sequences have been reported for *melonis*, possibly due to its heterogeneous nature. Two marker patterns were observed that were different from the marker patterns found in the majority of our *melonis* strains. Two strains (Fomln010 and Fomln023) were tested positive for *melonis* as well as *niveum* markers (pattern "B"). Interestingly, Fomln023 was capable of causing severe wilting symptoms both in melon as well as watermelon, while Fomln010 was not (van Dam *et al.*, 2016). This raises the question whether the separation of these *formae speciales* is justified, similar to the question whether strains pathogenic towards both cucumber and melon should be regarded as *cucumerinum* or *melonis*. Cafri *et al.* (2005) decided in their study that since the *cucumerinum* strains they tested were more aggressive towards cucumber than melon and no cross-pathogenicity was found the other way around, these *formae speciales* should indeed remain distinct. In the case of strain Fomln023 in the current study, disease severity was comparable between watermelon and melon plants, indicating that this strain is a 'bridging' *forma specialis*, and its marker gene pattern reflects this.

We recently demonstrated that clustering isolates based on presence-absence patterns of candidate effector genes divided *cucumerinum* into two groups, separated from each other by *melonis* and *niveum* strains (van Dam *et al.*, 2016). The cucurbit infecting isolates formed a supercluster from other *formae speciales*, indicating that they share a significant number of effector genes between them. Not much is known regarding the evolution of host-specificity of Fo towards cucurbits, but Fomln023 might contain accessory genetic material originating from both a *niveum* and a *melonis* strain. Likewise, strains Fomln017, 021, 024 and 026 tested positive for two *cucumerinum* markers and the *melonis* markers used all tested negative (pattern "C"). This indicates that these *melonis* and *cucumerinum* strains also share accessory genetic material. However, Fomln017 is, like most *melonis* isolates, highly specific to melon plants. It would be interesting to further investigate these strains, for example through long read sequencing of their genome and analysis of their pathogenicity chromosome(s) when compared to other *cucumerinum*, *melonis* and *niveum* strains. This could shed light on how pathogenicity towards cucurbits has evolved in the FOSC.

Horizontal gene and chromosome transfer has been described as an important contributor to genetic diversity and the generation of new (pathogenic) clonal lines in fungi (Ma *et al.*, 2013; Kang *et al.*, 2014). The different effector sequences and presence/absence patterns between and even within some cucurbit-infecting *formae speciales* suggest that it is possible that multiple horizontal transfer events of accessory genome material have taken place in the evolutionary trajectory resulting in pathogenicity towards cucurbits of *F. oxysporum*. This is in contrast with Fo f. sp. *lycopersici*, where the four clonal lines that were tested in van Dam *et al* (2016) all have nearly identical set of effectors and effector gene sequences (Lievens *et al.*, 2009; van Dam *et al.*, 2016).

Minor cross-reaction (a much lighter band) was found with one of the markers (Fomom

#130) with an unrelated *forma specialis*. Fo f. sp. *pisi* HDV247 (as well as Fo f. sp. *raphani* PHW815) indeed possesses this gene, although its downstream flank on which the reverse primer (FP7336) was designed was deemed to be sufficiently different from the copy in f. sp. *momordicae*; only the four 5' nucleotides matched between these two sequences. The forward primer only contained a single SNP, meaning that it probably binds in the non-target sequence of Fo f. sp. *pisi*, too. A similar observation was made between Fomln marker #1 and Foniv041 gDNA. For this isolate, however, no genome sequence is available. Through quantitative PCR techniques such as TaqMan, (cross-)reactions with a significantly lower amount of product can probably be distinguished from genuine positives.

As a proof of concept, a TaqMan test was developed for two of the markers. The TaqMan real-time PCR technique has several advantages over traditional PCR. Since it makes use of a sequence-specific fluorescently-tagged probe in addition to the primer sequences, marker specificity is potentially higher. Additionally the technique allows for quantification of the targeted DNA sequence (and thus of the pathogen in soil or infected plant tissue). Quantification of pathogenic *F. oxysporum* propagules in soil, seeds or plant tissues may aid in deciding if and when to take action. Thirdly, it is possible to test multiple markers by multiplexing, using several different fluorescent dyes at once (Weller *et al.*, 2000; Probert *et al.*, 2004; Agindotan *et al.*, 2007). The markers that were tested in duplex for f. sp. *niveum* behaved like expected: no amplification was identified in other strains (except Fomln010 with marker #100), even those that do possess the target gene. The technique allows for identification of sequences slightly different from the target sequence; compared to a positive control single copy gene like EF1α, the signal of a single copy marker with SNPs would be distinguishably higher.

These findings illustrate the hurdles that can be experienced in the design process of f. sp.-specific markers based on candidate effector genes, specifically if the *formae speciales* infect members of the same plant family and possibly arose through a shared and recent evolutionary history. Nonetheless, the combination of marker sequences described here can be used with relatively high fidelity to discriminate the seven cucurbit-affecting *formae speciales*, particularly when multiple markers are tested simultaneously in the analysis. It is possible – perhaps even likely – that more diversity exists among the seven *formae speciales* targeted here, since for several of the ff. spp. not all VCGs were sampled for genome sequencing due to unavailability of these strains. This means that the markers might require revision in the future. The availability of more whole genome sequences like the ones generated in this study will allow easier marker design and comparison in the future.

Materials and Methods

Whole genome sequencing, *de novo* assembly

F. oxysporum genomic DNA was isolated through phenol-chloroform extraction from freeze-dried mycelium that was harvested from five-day old NO_3-medium (0.17% yeast nitrogen base, 3% sucrose, 100 mM KNO_3) cultures as described in detail in (van Dam *et al.*, 2016). Library preparation of insert size 550 bp and Illumina HiSeq 2500 paired-end sequencing was performed at Keygene N.V. (Wageningen, the Netherlands).

Sequencing reads were trimmed for quality and to remove adapter sequences with FastqMcf v1.04.676 (http://code.google.com/p/ea-utils, quality threshold=20). *De novo* assemblies were generated using CLC-workbench 8.0. Default settings were used, except "minimum contig length=500".

For generating a core phylogeny, homologs of 15,956 Fol4287 core genes (including introns) were searched in all genomes using BLASTN with default parameters. We selected all sequences that overlapped >70% with the query sequence and with more than 80% identity to the query. We then selected query genes for which we found only a single hit in each genome, leaving us with 440 genes. We used ClustalO (Sievers *et al.*, 2014) to construct a multiple sequence alignment for each gene and a custom python script to concatenate these alignments. This alignment was subsequently trimmed using trimAl -strictplus. We used PhyML v20120412 (Guindon *et al.*, 2009) with 100 bootstraps to infer phylogeny and ETE v3.0.0b35 (Huerta-Cepas *et al.*, 2016) to visualize the tree.

Marker discovery and primer design

A custom python script was written to extract the sequence (plus 150 bp up- and downstream) of candidate effector genes from each of the genome assemblies using BLASTN with default parameters. MUSCLE v3.8.31 (Edgar, 2004) was used to generate alignments of each gene and phylogeny was inferred using PhyML v20120412 with 100 bootstraps. Another python script was used to traverse the tree in ETE v3 to identify instances where all isolates belonging to a *forma specialis* were clustering together in a separate clade, indicating sequence similarity that could potentially be used for primer design. Highlighting, drawing and rendering of the gene trees were done using ETE v3. Visual inspection of each of the gene trees allowed for the selection of a final set of marker genes per *forma specialis*. Scripts are available upon request. Primers were designed manually based on the sequence alignment per gene (see Supplementary data S2). In cases where only a few SNPs were identified to separate host-specificity of isolates, we aimed to target the mismatching nucleotides towards the 3' end of the primer, as described in (Liu *et al.*, 2012).

DNA isolation

Genomic DNA isolation for testing of markers was performed using 10-20 day-old mycelium

scraped off a PDA plate as starting material. The tissue was disrupted by shaking it in a TissueLyser (Qiagen) for 2 min. at 30 Hz in the presence of 400µl TE, 300µl phenol:chloroform (1:1) and glass beads. The aqueous phase was transferred to a fresh tube and an equal volume of chloroform was added. The DNA in the aqueous phase was transferred to a fresh tube and diluted 10x with sterile MilliQ water prior to use in Polymerase Chain Reaction (PCR). DNA quantity was estimated for the TaqMan standard curve using a Qubit Fluorometer (ThermoFisher Scientific).

Polymerase Chain Reaction

PCR was executed using Sphaero-Q Supertaq (Gorinchem, the Netherlands) in 20µl reaction volumes which included the following components (final concentration): 1X Sphaero-Q Supertaq buffer, 0.25 units of Sphaero-Q Supertaq, 5 pmol of each primer, dNTPs (0.2 mM each) and 1 µl template DNA. The following PCR program was used: 2' 94°C; [30" 94°C, 30" T_{ann}, 40" 72°C]$_{35x}$; 5' 72°C; pause 16°C. The PCR primer sequences and corresponding annealing temperatures are listed in Supplementary data S5. *Fusarium Extracellular Matrix 1* (*FEM1)* primers were used as a positive control and sterile MilliQ was used as a negative control for each of the primer combinations instead of template DNA.

TaqMan Real-Time PCR

TaqMan RT PCRs were performed on a QuantStudio® 3 system (ThermoFisher Scientific). Primers and probes were designed using Primer3web v4.0.0 (http://primer3.ut.ee/) and their sequences can be found in Supplementary data S6. A total volume of 10 µl of the reaction mixture included the following components (final concentration): 1X Sphaero-Q Supertaq buffer, 0.25 units of Sphaero-Q Supertaq (Gorinchem, the Netherlands), 3 pmol of each primer, 1 pmol of each probe, dNTPs (0.2 mM each), 0.1X ROX reference dye (ThermoFisher Scientific), 1 µl template DNA. Four simultaneous amplifications were performed for each sample to confirm reproducibility of the results. A negative control sample consisted of sterile MilliQ substituted for the DNA template. The PCR program was set as follows: 2' 94°C; [30" 94°C, 48" 60°C, 12" 60°C (data collection)]$_{40x}$.

Disease assays

Pathogenicity tests were performed using the root dip method (Wellman, 1939). In short, conidia were isolated from five-day-old cultures NO3-medium (0.17% yeast nitrogen base, 3% sucrose, 100 mM KNO3) by filtering through miracloth (Merck; pore size of 22–25 µm). Spores were centrifuged, resuspended in sterile MilliQ water, counted and brought to a final concentration of 10^7 spores/mL. When the first true leaves were emerging (after ±10 days), 5-8 seedlings per treatment were uprooted, inoculated, individually potted and kept at 20°C (Fo f. sp. *radicis-cucumerinum*) or 25°C (all other *formae speciales*) in the greenhouse. The following plant cultivars were used: *Cucumis sativus* cv. Paraiso, *Cucumis melo* cv. Cha-T,

Citrullus lanatus cv. Black Diamond. Two weeks after inoculation, disease was scored using a disease index from 0-4 as described in detail in van Dam *et al.*, 2016 (van Dam *et al.*, 2016).

Data access

Whole-Genome Shotgun projects for the newly sequenced strains of *F. oxysporum* have been deposited at Genbank under the BioProject PRJNA389501. Raw sequence data have been deposited into the Sequence Read Archive under the accession number SRP109253. All publically available genome sequences that were used were obtained from Genbank. Their NCBI accession numbers can be found in Supplementary Table S1.

Acknowledgments

This work was supported by the Horizon programme (project 93512007) of the Netherlands Genomics Initiative (NGI) through a grant to M. Rep.

References

Agindotan, B.O., Shiel, P.J., and Berger, P.H. (2007) Simultaneous detection of potato viruses, PLRV, PVA, PVX and PVY from dormant potato tubers by TaqMan® real-time RT-PCR. *J. Virol. Methods* **142**: 1–9.

Aguayo, J., Mostert, D., Fourrier-Jeandel, C., Cerf-Wendling, I., Hostachy, B., Viljoen, A., and Ioos, R. (2017) Development of a hydrolysis probe-based real-time assay for the detection of tropical strains of *Fusarium oxysporum* f. sp. *cubense* race 4. *PLoS One* **12**: e0171767.

Aimé, S., Alabouvette, C., Steinberg, C., and Olivain, C. (2013) The endophytic strain *Fusarium oxysporum* Fo47: a good candidate for priming the defense responses in tomato roots. *Mol. Plant-Microbe Interact.* **26**: 918–26.

Alabouvette, C., Olivain, C., Migheli, Q., and Steinberg, C. (2009) Microbiological control of soil-borne phytopathogenic fungi with special emphasis on wilt-inducing *Fusarium oxysporum*. *New Phytol.* **184**: 529–544.

Alves-Santos, F.M., Ramos, B., García-Sánchez, M.A., Eslava, A.P., and Díaz-Mínguez, J.M. (2002) A DNA-based procedure for In planta detection of *Fusarium oxysporum* f. sp. *phaseoli*. *Phytopathology* **92**: 237–244.

Armstrong, G.M. and Armstrong, J.K. (1981) *Formae speciales* and races of *Fusarium oxysporum* causing wilt diseases. In *Fusarium: diseases, biology, and taxonomy*. Pennsylvania State University Press, University Park, PA, pp. 391–399.

Armstrong, G.M. and Armstrong, J.K. (1978) *Formae speciales* and Races of *Fusarium oxysporum* Causing Wilts of the *Cucurbitaceae*. *Phytopathology* **68**: 19.

Baayen, R.P., O'Donnell, K., Bonants, P.J., Cigelnik, E., Kroon, L.P., Roebroeck, E.J., and Waalwijk, C. (2000) Gene Genealogies and AFLP Analyses in the *Fusarium oxysporum* Complex Identify Monophyletic and Nonmonophyletic *Formae Speciales* Causing Wilt and Rot Disease. *Phytopathology* **90**: 891–900.

Breeuwsma, S.J. and De Boer, M. (2004) *Fusarium* in bloembolgewassen: detectiemethoden en vruchtwisselingsproblematiek PPO Bloembollen en Bomen (in Dutch), PPO project 320689.Cafri, D., Katan, J., and Katan, T. (2005) Cross-pathogenicity between *formae speciales* of *Fusarium oxysporum*, the pathogens of cucumber and melon. *J. Phytopathol.* **153**: 615–622.

Cai, G., Gale, L.R., Schneider, R.W., Kistler, H.C., Davis, R.M., Elias, K.S., and Miyao, E.M. (2003) Origin of Race 3 of *Fusarium oxysporum* f. sp. *lycopersici* at a Single Site in California. *Phytopathology* **93**: 1014–1022.

Cenis, J.L., Tello, J., and Cifuentes, D. (2003) Genetic relationships among seven specialized forms of Fusarium oxysporum determined by DNA sequencing of the ITS region and AFLPs. *Spanish J. Agric. Res.* **1**: 55–63.

Chakrabarti, A., Rep, M., Wang, B., Ashton, A., Dodds, P., and Ellis, J. (2011) Variation in potential effector genes distinguishing Australian and non-Australian isolates of the cotton wilt pathogen *Fusarium oxysporum* f.sp. *vasinfectum*. *Plant Pathol.* **60**: 232–243.

Chen, K.S., Chang, P.-F., Liou, T.D., and Huang, J.W. (2003) Identification of physiological races of *Fusarium oxysporum* f. sp. *niveum* and breeding for *Fusarium* wilt-resistant watermelon line. *Plant Pathol. Bull.* **12**:

247–254.
Clark, C.A., Hyun, J.-W., and Hoy, M.W. (1998) Relationships Among Wilt-Inducing Isolates of Fusarium oxysporum from Sweetpotato and Tobacco. *Plant Dis.* **82**: 530–536.
Cohen, R., Orgil, G., Burger, Y., Saar, U., Elkabetz, M., Tadmor, Y., et al. (2015) Differences in the responses of melon accessions to fusarium root and stem rot and their colonization by *Fusarium oxysporum* f. sp. *radicis-cucumerinum*. *Plant Pathol.* **64**: 655–663.
Cohen, R., Tyutyunik, J., Fallik, E., Oka, Y., Tadmor, Y., and Edelstein, M. (2014) Phytopathological evaluation of exotic watermelon germplasm as a basis for rootstock breeding. *Sci. Hortic.* **165**: 203–210.
Correll, J.C. (1991) The relationship between formae speciales, races, and vegetative compatibility groups in *Fusarium oxysporum*. *Phytopathology* **81**: 1061–1064.
Covey, P. a, Kuwitzky, B., Hanson, M., and Webb, K.M. (2014) Multilocus analysis using putative fungal effectors to describe a population of *Fusarium oxysporum* from sugar beet. *Phytopathology* **104**: 886–896.
Cumagun, J., Oribiana, Z., Tolentino, M., Relevante, C., and Balatero, C. (2008) Vegetative compatibility among *Fusarium oxysporum* isolates from bitter gourd and bottle gourd in the Philippines. *J. Plant Prot. Res.* **48**: 283–293.
van Dam, P., Fokkens, L., Schmidt, S.M., Linmans, J.H.J., Kistler, H.C., Ma, L.-J., and Rep, M. (2016) Effector profiles distinguish *formae speciales* of *Fusarium oxysporum*. *Environ. Microbiol.* **18**: 4087–4102.
van Dam, P. and Rep, M. (2017) The Distribution of *Miniature Impala* Elements and *SIX* Genes in the *Fusarium* Genus is Suggestive of Horizontal Gene Transfer. *J. Mol. Evol.*
Van Der Does, H.C., Lievens, B., Claes, L., Houterman, P.M., Cornelissen, B.J.C., and Rep, M. (2008) The presence of a virulence locus discriminates *Fusarium oxysporum* isolates causing tomato wilt from other isolates. *Environ. Microbiol.* **10**: 1475–1485.
Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792–1797.
Fraser-Smith, S., Czislowski, E., Meldrum, R. a., Zander, M., O'Neill, W., Balali, G.R., and Aitken, E. a. B. (2014) Sequence variation in the putative effector gene *SIX8* facilitates molecular differentiation of *Fusarium oxysporum* f. sp. *cubense*. *Plant Pathol* 10.1111/ppa.12184.
Fravel, D., Olivain, C., and Alabouvette, C. (2003) *Fusarium oxysporum* and its biocontrol. *New Phytol.* **157**: 493–502.
Gerlagh, M. and Blok, W.J. (1988) *Fusarium oxysporum* f.sp. *cucurbitacearum* n.f. embracing all *formae speciales* of *F. oxysporum* attacking Cucurbitaceous crops. *Netherlands J. Plant Pathol.* **94**: 17–31.
Gordon, T.R. (2017) *Fusarium oxysporum* and the *Fusarium* wilt Syndrome. *Annu. Rev. Phytopathol.* **55**: 1–17.
Gordon, T.R. and Martyn, R.D. (1997) The evolutionary biology of *Fusarium oxysporum*. *Annu. Rev. Phytopathol.* **35**: 111–128.
Guo, L., Han, L., Yang, L., Zeng, H., Fan, D., Zhu, Y., et al. (2014) Genome and Transcriptome Analysis of the Fungal Pathogen *Fusarium oxysporum* f. sp. *cubense* Causing Banana Vascular Wilt Disease. *PLoS One* **9**: e95543.
Guindon, S., Dufayard, J.F., Hordijk, W., Lefort, V., and Gascuel, O. (2009) PhyML: fast and accurate phylogeny reconstruction by maximum likelihood. In, *Infection Genetics and Evolution*., pp. 384–385.
Hadar, E. and Katan, J. (1989) The Use of Nitrate-Nonutilizing Mutants and a Selective Medium for Studies of Pathogenic Strains of *Fusarium oxysporum*. *Plant Dis.* **73**: 800.
Haegi, A., Catalano, V., Luongo, L., Vitale, S., Scotton, M., Ficcadenti, N., and Belisario, A. (2013) A newly developed real-time PCR assay for detection and quantification of *Fusarium oxysporum* and its use in compatible and incompatible interactions with grafted melon genotypes. *Phytopathology* **103**: 802–10.
Hogg, A.C., Johnston, R.H., and Dyer, A.T. (2007) Applying Real-Time Quantitative PCR to Fusarium Crown Rot of Wheat. *Plant Dis.* **91**: 1021–1028.
Huang, C.-H., Tsai, R.-T., and Vallad, G.E. (2016) Development of a TaqMan Real-Time Polymerase Chain Reaction Assay for Detection and Quantification of *Fusarium oxysporum* f. sp. *lycopersici* in Soil. *J. Phytopathol.* **164**: 455–463.
Huerta-Cepas, J., Serra, F., and Bork, P. (2016) ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **33**: 1635–1638.
Inami, K., Yoshioka-Akiyama, C., Morita, Y., Yamasaki, M., Teraoka, T., and Arie, T. (2012) A genetic mechanism for emergence of races in *Fusarium oxysporum* f. sp. *lycopersici*: Inactivation of avirulence gene *AVR1* by transposon insertion. *PLoS One* **7**: e44101.
Kang, S., Demers, J., del Mar Jimenez-Gasco, M., and Rep, M. (2014) *Fusarium oxysporum*. In, *Genomics of Plant-*

Associated Fungi and Oomycetes: Dicot Pathogens. Springer, pp. 99–119.
Katan, T. (1999) Current status of vegetative compatibility groups in *Fusarium oxysporum*. *Phytoparasitica* **27**: 51–64.
Kim, D.H., Martyn, R.D., and Magill, C.W. (1993) Mitochondrial DNA (mtDNA)-Relatedness among *formae speciales* of *Fusarium oxysporum* in the Cucurbitaceae. *Phytopathology* **83**: 91–97.
Kistler, H.C. (1997) Genetic Diversity in the Plant-Pathogenic Fungus *Fusarium oxysporum*. *Phytopathology* **87**: 474–479.
Leach, J.G., Currence, T.M., and others (1937) *Fusarium* wilt of muskmelons in Minnesota. *Univ. Minnesota, Agric. Exp. Stn. Bull.* **129**:.
Leslie, J.F. (1993) Fungal vegetative compatibility. *Annu. Rev. Phytopathol.* **31**: 127–150.
Lievens, B., Brouwer, M., Vanachter, A.C.R.C., Cammue, B.P.A., and Thomma, B.P.H.J. (2006) Real-time PCR for detection and quantification of fungal and oomycete tomato pathogens in plant and soil samples. *Plant Sci.* **171**: 155–165.
Lievens, B., Claes, L., Vakalounakis, D.J., Vanachter, A.C.R.C., and Thomma, B.P.H.J. (2007) A robust identification and detection assay to discriminate the cucumber pathogens *Fusarium oxysporum* f. sp. *cucumerinum* and f. sp. *radicis-cucumerinum*. *Environ. Microbiol.* **9**: 2145–2161.
Lievens, B., Houterman, P.M., and Rep, M. (2009) Effector gene screening allows unambiguous identification of *Fusarium oxysporum* f. sp. *lycopersici* races and discrimination from other *formae speciales*. *FEMS Microbiol. Lett.* **300**: 201–215.
Lievens, B., Rep, M., and Thomma, B.P.H.J. (2008) Recent developments in the molecular discrimination of *formae speciales* of *Fusarium oxysporum*. *Pest Manag. Sci.* **64**: 781–788.
Lievens, B. and Thomma, B.P.H.J. (2005) Recent developments in pathogen detection arrays: implications for fungal plant pathogens and use in practice. *Phytopathology* **95**: 1374–1380.
Lin, Y.-H., Chen, K.-S., Chang, J.-Y., Wan, Y.-L., Hsu, C.-C., Huang, J.-W., and Chang, P.-F.L. (2010) Development of the molecular methods for rapid detection and differentiation of *Fusarium oxysporum* and *F. oxysporum* f. sp. *niveum* in Taiwan. *N. Biotechnol.* **27**: 409–418.
Ling, J., Zhang, J., Zeng, F., Cao, Y., Xie, B., and Yang, Y. (2016) Comparative genomics provide a rapid detection of *Fusarium oxysporum* f. sp. *conglutinans*. *J. Integr. Agric.* **15**: 822–831.
Liu, J., Huang, S., Sun, M., Liu, S., Liu, Y., Wang, W., et al. (2012) An improved allele-specific PCR primer design method for SNP marker analysis and its application. *Plant Methods* **8**: 34.
Ma, L.-J., Geiser, D.M., Proctor, R.H., Rooney, A.P., O'Donnell, K., Trail, F., et al. (2013) *Fusarium* pathogenomics. *Annu. Rev. Microbiol.* **67**: 399–416.
Del Mar Jiménez-Gasco, M. and Jiménez-Díaz, R.M. (2003) Development of a Specific Polymerase Chain Reaction-Based Assay for the Identification of *Fusarium oxysporum* f. sp. *ciceris* and Its Pathogenic Races 0, 1A, 5, and 6. *Phytopathology* **93**: 200–209.
Martyn, R.D. (2014) *Fusarium* Wilt of Watermelon: 120 Years of Research. *Hortic. Rev. Vol. 42* **42**: 349–442.
Matuo, T. and Yamamoto, I. (1967) On *Fusarium oxysporum* f.sp. *lagenariae* n.f. causing wilt of *Lagenaria vulgaris* var. hispida. *Trans. Mycol. Soc. Japan* **8**: 61–63.
Mbofung, G.C.Y., Fessehaie, A., Bhattacharyya, M.K., and Leandro, L.F.S. (2011) A new TaqMan real-time polymerase chain reaction assay for quantification of *Fusarium virguliforme* in soil. *Plant Dis.* **95**: 1420–1426.
Mcmillan, T. (1986) Cross pathogenicity studies with isolates of *Fusarium oxysporum* from either cucumber or watermelon pathogenic to both crop species. *Ann. Appl. Biol.* **109**: 101–105.
Mes, J.J., Doorn, Van, J., Roebroeck, E.J.A., Van Egmond, E., Van Aartrijk, J., and Boonekamp, P.M. (1994) Restriction fragment length polymorphisms, races and vegetative compatibility groups within a worldwide collection of *Fusarium oxysporum* f. sp. *gladioli*. *Plant Pathol.* **43**: 362–370.
Mes, J.J., Weststeijn, E.A., Herlaar, F., Lambalk, J.J.M., Wijbrandi, J., Haring, M.A., and Cornelissen, B.J.C. (1999) Biological and molecular characterization of *Fusarium oxysporum* f. sp. *lycopersici* divides race 1 isolates into separate virulence groups. *Phytopathology* **89**: 156–160.
Michielse, C.B. and Rep, M. (2009) Pathogen profile update: *Fusarium oxysporum*. *Mol. Plant Pathol.* **10**: 311–324.
Mirtalebi, M. and Banihashemi, Z. (2014) Genetic Relationship among *Fusarium oxysporum* f. sp. *melonis* Vegetative Compatibility Groups and Their Relatedness to Other *F. oxysporum formae speciales*. *J. Agric. Sci. Technol.* **16**: 931–943.
Namiki, F., Shiomi, T., Kayamura, T., and Tsuge, T. (1994) Characterization of the *formae speciales* of *Fusarium oxysporum* causing wilts of cucurbits by DNA fingerprinting with nuclear repetitive DNA sequences. *Appl.*

Environ. Microbiol. **60**: 2684–2691.
O'Donnell, K., Cigelnik, E., and Nirenberg, H.I. (1998) Molecular Systematics and Phylogeography of the Gibberella fujikuroi Species Complex. *Mycologia* **90**: 465–493.
Owen, J.H. (1956) Cucumber wilt, caused by *Fusarium oxysporum* f. *cucumerinum*. *Phytopathology* **46**: 153–157.
Pasquali, M., Dematheis, F., Gullino, M.L., and Garibaldi, A. (2007) Identification of Race 1 of *Fusarium oxysporum* f. sp. *lactucae* on Lettuce by Inter-Retrotransposon Sequence-Characterized Amplified Region Technique. *Phytopathology* **97**: 987–996.
Pavlou, G.C., Vakalounakis, D.J., and Ligoxigakis, E.K. (2002) Control of Root and Stem Rot of Cucumber, Caused by *Fusarium oxysporum* f. sp. *radicis-cucumerinum*, by Grafting onto Resistant Rootstocks. *Plant Dis.* **86**: 379–382.
Probert, W.S., Schrader, K.N., Khuong, N.Y., Bystrom, S.L., and Graves, M.H. (2004) Real-Time Multiplex PCR Assay for Detection of *Brucella* spp ., *B. abortus*, and *B. melitensis*. *J. Clin Microbiol.* **42**: 1290–1293.
Recorbet, G., Steinberg, C., Olivain, C., Edel, V., Trouvelot, S., Dumas-Gaudot, E., et al. (2003) Wanted: pathogenesis-related marker molecules for *Fusarium oxysporum*. *New Phytol.* **159**: 73–92.
Rocha, L.O., Laurence, M.H., Ludowici, V.A., Puno, V.I., Lim, C.C., Tesoriero, L.A., et al. (2015) Putative effector genes detected in *Fusarium oxysporum* from natural ecosystems of Australia. *Plant Pathol.* **65**: 914–929
Schmidt, S.M., Lukasiewicz, J., Farrer, R., van Dam, P., Bertoldo, C., and Rep, M. (2016) Comparative genomics of *Fusarium oxysporum* f. sp. *melonis* reveals the secreted protein recognized by the *Fom-2* resistance gene in melon. *New Phytol.* **209**: 307–318.
Schoffelmeer, E.A., Vossen, J.H., van Doorn, A.A., Cornelissen, B.J., and Haring, M.A. (2001) FEM1, a *Fusarium oxysporum* glycoprotein that is covalently linked to the cell wall matrix and is conserved in filamentous fungi. *Mol. Genet. Genomics* **265**: 143–52.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., et al. (2014) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**: 539–539.
Skovgaard, K., Nirenberg, H.I., O'Donnell, K., and Rosendahl, S. (2001) Evolution of Fusarium oxysporum f. sp. vasinfectum Races Inferred from Multigene Genealogies. *Phytopathology* **91**: 1231–1237.
Sun, S. and Huang, J. (1982) A new *Fusarium* wilt of bitter gourd in Taiwan. *Plant Dis.* **67**: 226–227.
Taylor, A., Vagany, V., Jackson, A.C., Harrison, R.J., Rainoni, A., and Clarkson, J.P. (2015) Identification of pathogenicity-related genes in *Fusarium oxysporum* f. sp. *cepae*. *Mol. Plant Pathol.*
Thatcher, L.F., Gardiner, D.M., Kazan, K., and Manners, J.M. (2012) A highly conserved effector in *Fusarium oxysporum* is required for full virulence on *Arabidopsis*. *Mol. Plant. Microbe. Interact.* **25**: 180–90.
Vakalounakis, D.J. (1996) Root and Stem Rot of Cucumber Caused by *Fusarium oxysporum* f.sp. *radicis-cucumerinum* f.sp. nov. *Plant Dis.* **80**: 313–316.
Vakalounakis, D.J. and Fragkiadakis, G.A. (1999) Genetic Diversity of *Fusarium oxysporum* Isolates from Cucumber: Differentiation by Pathogenicity, Vegetative Compatibility, and RAPD Fingerprinting. *Phytopathology* **89**: 161–168.
Vakalounakis, D.J., Wang, Z., Fragkiadakis, G.A., Skaracis, G.N., and Li, D.-B. (2004) Isolates Obtained from Cucumber in China by Pathogenicity, VCG, and RAPD. *Plant Dis.* **88**: 645–649.
Wang, C., Lin, Y., Lin, Y., and Chung, W. (2013) Modified Primers for the Identification of Nonpathogenic *Fusarium oxysporum* Isolates That Have Biological Control Potential against Fusarium Wilt of Cucumber in Taiwan. *PLoS One* **8**: e65093.
Weller, S.A., Elphinstone, J.G., Smith, N.C., Boonham, N., and Stead, D.E. (2000) Detection of *Ralstonia solanacearum* Strains with a Quantitative, Multiplex, Real-Time, Fluorogenic PCR (TaqMan) Assay. *Appl. Environ. Microbiol.* **7**: 2853–2858.
Wellman, F.L. (1939) A technique for studying host resistance and pathogenicity in tomato Fusarium wilt. *Phytopathology* **29**: 945–956.
Zhang, Z., Zhang, J., Wang, Y., and Zheng, X. (2005) Molecular detection of *Fusarium oxysporum* f. sp. *niveum* and *Mycosphaerella melonis* in infected plant tissues and soil. *FEMS Microbiol. Lett.* **249**: 39–47.
Zhou, X.G. and Everts, K.L. (2003) Races and Inoculum Density of *Fusarium oxysporum* f . sp . *niveum* in Commercial Watermelon Fields in Maryland and Delaware. *Plant Dis.* **87**: 692–698.
Zhou, X.G. and Everts, K.L. (2007) Characterization of a Regional Population of *Fusarium oxysporum* f. sp. *niveum* by Race, Cross Pathogenicity, and Vegetative Compatibility. *Phytopathology* **97**: 461–469.

Supplementary Data

Supplementary data S1: Overview of *Fusarium* strains used in this study

Strain	Original designation	*Forma specialis*	VCG	Race	Origin of strain	Reference	GenBank assembly accession
Folyc 4287	NRRL34936	*lycopersici*	0030	2	Spain	Broad institute	GCA_000149955.2
Focon 5176	Fo5176	(on *Brassica)*	-	-	Australia	Broad institute	GCA_000222805.1
Folyc MN25	NRRL54003	*lycopersici*	0033	3	USA	Broad institute	GCA_000259975.2
Fopis HDV247	NRRL37622	*pisi*	-	-	-	Broad institute	GCA_000260075.2
Forly CL57	NRRL26381	*radicis-lycopersici*	0094	-	USA, Florida	Broad institute	GCA_000260155.3
Fovas 25433	NRRL25433	*vasinfectum*	-	-	China	Broad institute	GCA_000260175.2
Focub II-5	NRRL54006	*cubense*	01213	TR4	Indonesia	Broad institute	GCA_000260195.2
Focon PHW808	NRRL54008	*conglutinans*	0101	2	-	Broad institute	GCA_000260215.2
Forap PHW815	NRRL54005	*raphani*	0102	-	-	Broad institute	GCA_000260235.2
Fomln001	NRRL26406	*melonis*	0136	1	Mexico	Broad institute	GCA_000260495.2
Fo47	NRRL54002	(non-pathogen, biocontrol)	-	-	France	Broad institute	GCA_000271705.2
Focli FOSC 3-a	NRRL32931	(clinical isolate, from human blood)	-	-	USA, Massachusetts	Broad institute	GCA_000271745.2
Focub N2	N2	*cubense*	-	1	China	Guo et al., 2014	GCA_000350345.1
Focub B2	B2	*cubense*	-	4	China	Guo et al., 2014	GCA_000350365.1
Focuc013	9904-1	*cucumerinum*	0186	-	China	Lievens et al., 2007	MABJ01000000
Focuc015	9906-3	*cucumerinum*	0184	-	China	Lievens et al., 2007	MABK01000000
Focuc021	ATCC 16416	*cucumerinum*	0180	-	USA, Florida	Lievens et al., 2007	MABL01000000
Focuc018	Afu-50(B)	*cucumerinum*	0180	-	Greece, Crete	Lievens et al., 2007	MABM01000000
Focuc030	FOCU-22P	*cucumerinum*	0180	-	Israel	Lievens et al., 2007	MABN01000000
Focuc035	NETH 11179	*cucumerinum*	0181	-	Netherlands	Lievens et al., 2007	MABO01000000
Focuc037	Tf-213	*cucumerinum*	0185	-	Japan	Lievens et al., 2007	MABP01000000
Forcu016	33	*radicis-cucumerinum*	0260	-	Canada	Lievens et al., 2007	MABQ02000000
Forcu024	Afu-11(A)	*radicis-cucumerinum*	0260	-	Greece, Crete	Lievens et al., 2007	MABR01000000
Forcu031	AK-2	*radicis-cucumerinum*	0261	-	Greece, Crete	Lievens et al., 2007	MABS01000000
Focuc011	9903-1	*cucumerinum*	0186	-	China	Lievens et al., 2007	MABT01000000
Fomln005	Fom 0123	*melonis*	0134	1	Spain	Schmidt et al., 2016	MAKY01000000
Focuc001	Foc-1	*cucumerinum*	0183	-	Japan	BL	MAKZ01000000
Fomln004	Fom 0122	*melonis*	0134	0	Spain	Schmidt et al., 2016	MALA01000000
Fomln006	Fom 0124	*melonis*	0134	2	Spain	Schmidt et al., 2016	MALB01000000
Fomln009	-	*melonis*	0135	2	Israel	Schmidt et al., 2016	MALC01000000
Fomln010	-	*melonis*	-	1	Israel	Schmidt et al., 2016	MALD01000000
Fomln012	ML2	*melonis*	0134	0	-	Schmidt et al., 2016	MALE01000000
Fomln016	Fom26	*melonis*	0134	1	-	Schmidt et al., 2016	MALF01000000
Fomln013	-	*melonis*	0134	2	Spain	Schmidt et al., 2016	MALG01000000
Folyc004	IPO1530 / B1	*lycopersici*	0030	1	Netherlands	Mes et al., 1999	MALH01000000
Folyc007	D2	*lycopersici*	0030	2	France	Mes et al., 1999	MALI01000000

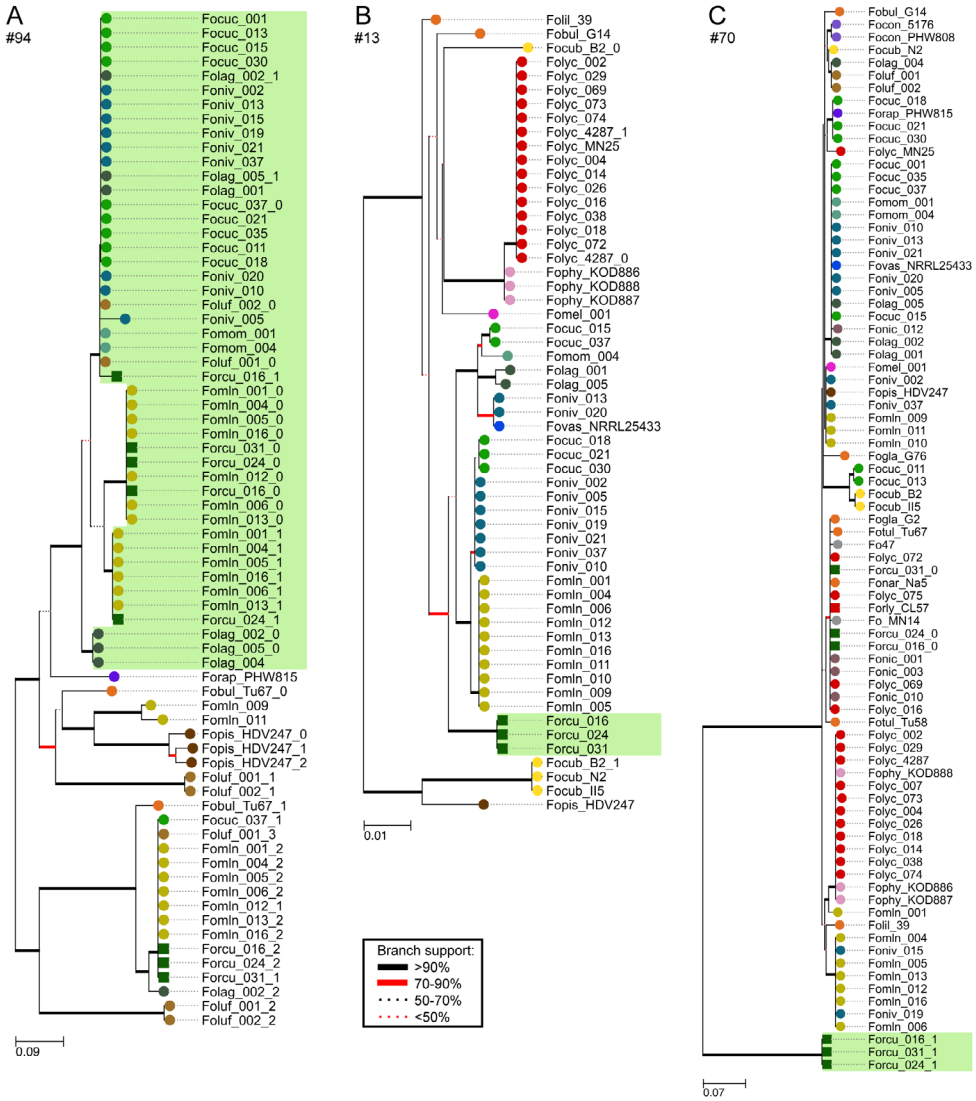
Folyc014	LSU-3	*lycopersici*	0030	1	USA, Louisiana	Mes et al., 1999	MALJ01000000
Folyc026	BRIP 14844 (M1943)	*lycopersici*	0030	3	Australia	Mes et al., 1999	MALK01000000
Folyc018	LSU-7	*lycopersici*	0030	2	USA, Louisiana	Mes et al., 1999	MALL01000000
Folyc016	BFOL-51	*lycopersici*	0031	1	USA, Louisiana	Mes et al., 1999	MALM01000000
Folyc029	5397	*lycopersici*	0030	3	USA, Florida	Mes et al., 1999	MALN01000000
Folyc038	CA92/95	*lycopersici*	0030	3	USA, California	Lievens et al., 2009	MALO01000000
Folyc069	DF0-23	*lycopersici*	0035	2	USA, California	Cai et al., 2003	MALP01000000
Folyc072	DF0-38	*lycopersici*	0031	2	USA, California	Cai et al., 2003	MALQ01000000
Folyc073	DF0-40	*lycopersici*	0030	2	USA, California	Cai et al., 2003	MALR01000000
Folyc074	DF0-41	*lycopersici*	0030	3	USA, California	Cai et al., 2003	MALS01000000
Folyc075	DF0-62	*lycopersici*	0031	2	USA, California	Cai et al., 2003	MALT01000000
FoMN14	MN-14	(non-pathogen, from tomato plant)	-	-	USA, California	Gale et al., 2003	MALU01000000
Foniv002	CBS 418.90	*niveum*	-	-	Israel	Lievens et al., 2007	MALX01000000
Foniv005	TX-471-1	*niveum*	0080	0	USA, Texas	Zhou and Everts, 2007	MALY01000000
Foniv010	F-016-1	*niveum*	0082	1	USA, Maryland	Zhou and Everts, 2007	MALZ01000000
Foniv013	F-014-2	*niveum*	0082	2	USA, Maryland	Zhou and Everts, 2007	MAMA01000000
Foniv015	F-063-1	*niveum*	0082	2	USA, Maryland	Zhou and Everts, 2007	MAMB01000000
Foniv019	TX-X1D	*niveum*	0082	2	USA, Texas	Zhou and Everts, 2007	MAMC01000000
Foniv020	F-099-1	*niveum*	0083	2	USA, Delaware	Zhou and Everts, 2007	MAMD01000000
Foniv021	MD-ZE622	*niveum*	-	3	USA, Maryland	Zhou and Everts, 2007	MAME01000000
Foniv037	NRRL38539	*niveum*	-	-	Israel	Hadar and Katan, 1989	MAMF01000000
Folyc002	WCS862 / E241	*lycopersici*	0030	2	Netherlands	Mes et al., 1999	MAMG01000000
Fomln011	-	*melonis*	-	0	Israel	Schmidt et al., 2016	MAMH01000000
Fogla G14	G14	*gladioli*	0341	-	Netherlands	Mes et al., 1994	NJCM01000000
Fogla G2	G2	*gladioli*	0340	-	France	Mes et al., 1994	NJCL01000000
Fogla G76	G76	*gladioli*	0343	-	Italy	Mes et al., 1994	NJCK01000000
Folag001	01-03008	*lagenariae*	-	-	Japan	Namiki et al., 1994	NJCJ01000000
Folag002	03-05118	*lagenariae*	-	-	Japan	Namiki et al., 1994	NJCI01000000
Folag004	Lag:3-1 (JCM9293)	*lagenariae*	-	-	Japan	Namiki et al., 1994	NJCH01000000
Folag005	Lag:1-1	*lagenariae*	-	-	Japan	Namiki et al., 1994	NJCG01000000
Folil Fol39	Fol39	*lilii*	-	-	Netherlands	Breeuwsma & de Boer, 2004; JVD	NJCF01000000
Foluf001	Fol-114	*luffae*	-	-	Taiwan	Lievens et al., 2009	NJCE01000000
Foluf002	Fol-167	*luffae*	-	-	Taiwan	Lievens et al., 2009	NJCD01000000
Fomel001	J-71	*melongenae*	-	-	-	IPO	NJCC01000000
Fomom001	NRRL26413	*momordicae*	-	-	Taiwan	ARS	NJCB01000000
Fomom004	90NF2-1 (JCM9292)	*momordicae*	-	-	Japan	Namiki et al., 1994	NJCA01000000
Fonar Na5	Na5	*narcissi*	2	-	Netherlands	Breeuwsma & de Boer, 2004; JVD	NJCV01000000
Fonic001	FON-1	*nicotianae*	-	-	USA, Connecticut	Clark et al., 1998	NJBZ01000000
Fonic003	10913	*nicotianae*	0373	-	USA, Maryland	Clark et al., 1998	NJBY01000000
Fonic010	Ft-Rob	*nicotianae*	0378	-	USA, North Carolina	Clark et al., 1998	NJBX01000000
Fonic012	Ft-1512	*nicotianae*	-	-	USA, North Carolina	Clark et al., 1998	NJCU01000000
Fophy KOD886	KOD886	*physali*	-	-	USA, California	KOD	NJBW01000000

Fophy KOD887	KOD887	*physali*	-	-	USA, California	KOD	NJBV01000000
Fophy KOD888	KOD888	*physali*	-	-	USA, California	KOD	NJBU01000000
Fo Tu58	Tu58	(non-pathogenic, from symptomous tulip)	-	-	Netherlands	Breeuwsma & de Boer, 2004; JVD	NJBT01000000
Fotul Tu67	Tu67	*tulipae*	-	-	Netherlands	Breeuwsma & de Boer, 2004; JVD	NJBS01000000
14150	14150	*cucurbitacearum* (redesignated as *radicis-cucumerinum)*	-	-	Netherlands	NAKT	Not available
Focuc009	0020	*cucumerinum*	0187	-	China	Lievens et al., 2007	Not available
Focuc010	9901-2	*cucumerinum*	0186	-	China	Lievens et al., 2007	Not available
Focuc014	9906-2	*cucumerinum* (redesignated as *radicis-cucumerinum)*	0184	-	China	Lievens et al., 2007	Not available
Focuc016	9909-2	*cucumerinum*	0185	-	China	Lievens et al., 2007	Not available
Focuc017	9909-3	*cucumerinum*	0186	-	China	Lievens et al., 2007	Not available
Focuc022	ATCC 36330	*cucumerinum* (redesignated as *non-pathogen)*	0180	-	Israel	Lievens et al., 2007	Not available
Focuc026	ATCC 201950	*cucumerinum*	0180	-	USA, Florida	Lievens et al., 2007	Not available
Focuc027	Cu:4-1 Koma 4	*cucumerinum*	0181	-	Japan	Lievens et al., 2007	Not available
Focuc028	Cu:5-0 Koma 5	*cucumerinum* (redesignated as *non-pathogen)*	0183	-	Japan	Lievens et al., 2007	Not available
Focuc036	NRRL26437	*cucumerinum*	-	-	USA, South Carolina	ARS	Not available
Focuc038	NRRL26744	*cucumerinum* (redesignated as *non-pathogen)*	-	-	Japan	ARS	Not available
Focuc039	NRRL38591	*cucumerinum*	-	-	New Zealand	ARS	Not available
Focuc040	07-08969	*cucumerinum*(redesignated as *radicis-cucumerinum)*	-	-	Netherlands	NAKT	Not available
Folag003	07-27503	*lagenariae*	-	-	Japan	Namiki et al., 1994	Not available
Folag006	Lag:7-1	*lagenariae*	-	-	Japan, Kumamoto	Namiki et al., 1994	Not available
Folag007	No. 87	*lagenariae*	-	-	Japan, Tochigi	Namiki et al., 1994	Not available
Folag008	No. 134	*lagenariae*	-	-	Japan, Tochigi	Namiki et al., 1994	Not available
Fomln002	CBS 420.90	*melonis*	-	-	Israel	Lievens et al., 2007	Not available
Fomln003	CBS 423.90	*melonis*	-	-	Israel	Lievens et al., 2007	Not available
Fomln017	NRRL22518	*melonis*	-	-	USA	O'Donnell et al., 1998	Not available
Fomln018	NRRL22519	*melonis*	-	-	France	O'Donnell et al., 1998	Not available
Fomln019	NRRL22520	*melonis*	-	-	USA	O'Donnell et al., 1998	Not available
Fomln020	NRRL22521	*melonis*	-	-	Belgium	O'Donnell et al., 1998	Not available
Fomln021	NRRL26172	*melonis*	-	-	China	ARS	Not available
Fomln023	NRRL26174	*melonis*	-	-	China	ARS	Not available
Fomln024	NRRL26745	*melonis*	-	-	Japan	ARS	Not available
Fomln025	NRRL26746	*melonis*	-	-	Japan	ARS	Not available
Fomln026	NRRL38516	*melonis*	-	-	New Zealand	ARS	Not available
Fomln027	NRRL38524	*melonis*	-	-	New Zealand	ARS	Not available
Fomom002	NRRL26748	*momordicae*	-	-	Japan	Skovgaard et al., 2001	Not available
Fomom003	90NF1-2	*momordicae*	-	-	Japan, Kagoshima	Namiki et al., 1994	Not available
Foniv011	F-086-1	*niveum*	0082	1	USA, Maryland	Zhou and Everts, 2003	Not available
Foniv017	F-097-2	*niveum*	0082	-	USA, Delaware	Zhou and Everts, 2007	Not available
Foniv018	F-100-2	*niveum*	0082	2	USA, Delaware	Zhou and Everts, 2007	Not available

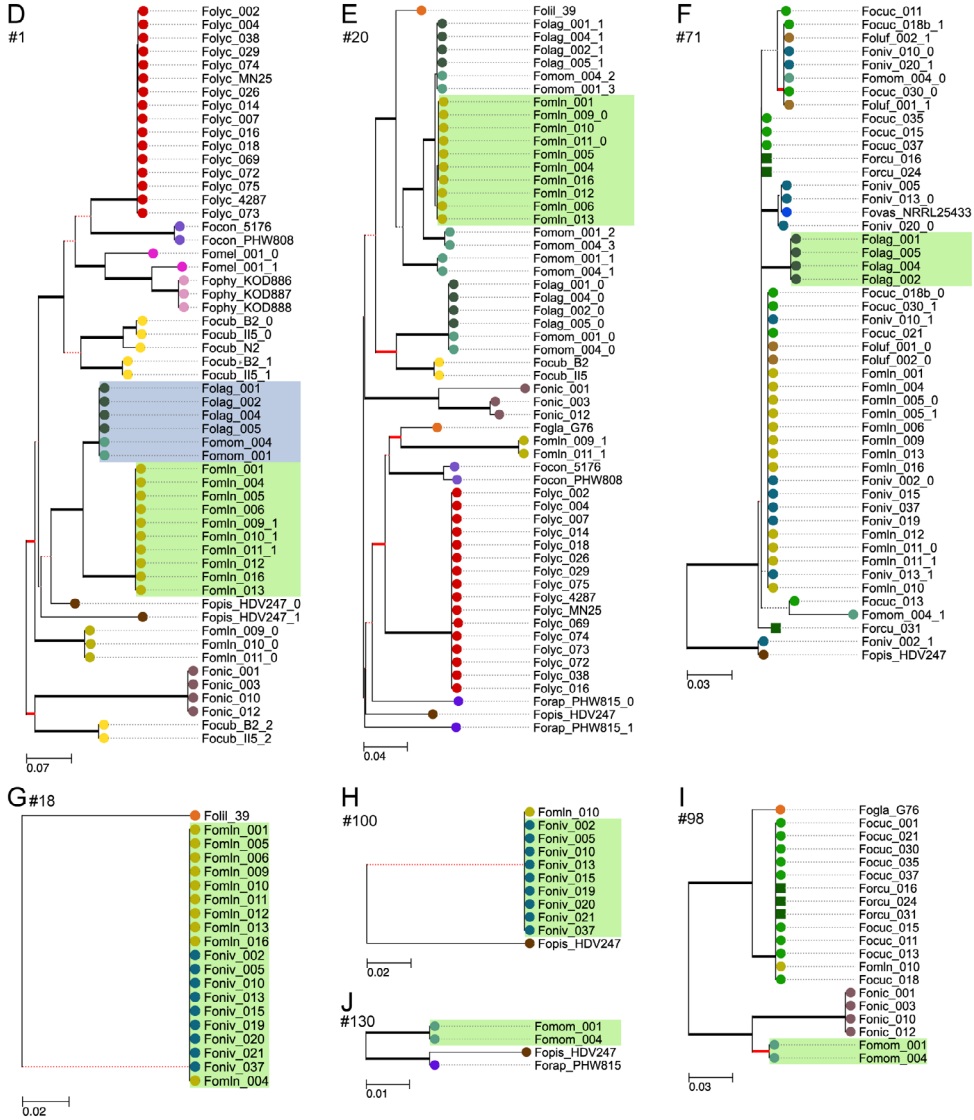
Foniv033	NRRL26747	*niveum*	-	-	Japan	ARS	Not available
Foniv034	NRRL36275	*niveum (redesignated as non-pathogen)*	-	-	-	ARS	Not available
Foniv035	NRRL38278	*niveum (redesignated as non-pathogen)*	-	-	USA	ARS	Not available
Foniv036	NRRL38503	*niveum*	-	-	New Zealand	ARS	Not available
Foniv038	NRRL38552	*niveum (redesignated as non-pathogen)*	-	-	Israel	ARS	Not available
Foniv039	LB	*niveum*	-	0	-	NAKT	Not available
Foniv040	IPO 30288	*niveum*	-	1	-	IPO	Not available
Foniv041	CBS 419.90	*niveum*	-	-	Israel	Lievens et al., 2007	Not available
Forcu005	14	*radicis-cucumerinum*	-	-	Canada	Lievens et al., 2007	Not available
Forcu017	34	*radicis-cucumerinum*	-	-	Canada	Lievens et al., 2007	Not available
Forcu020	38	*radicis-cucumerinum*	-	-	France	Lievens et al., 2007	Not available
Forcu028	Afu-58	*radicis-cucumerinum*	0260	-	Greece, Crete	Lievens et al., 2007	Not available
Forcu029	Afu-68(A)	*radicis-cucumerinum*	0260	-	Greece, Crete	Lievens et al., 2007	Not available
RBG1687	RBG1687	(non-pathogen, from *Wollemia nobilis* seedling leaves)	-	-	Australia	ML	Not available
RBG1693	RBG1693	(from flannel flower roots)	-	-	Australia	ML	Not available
RBG5713	RBG5713	(non-pathogen, from soil)	-	-	Australia	Rocha et al., 2015	Not available
RBG5786	RBG5786	(non-pathogen, from soil)	-	-	Australia	Rocha et al., 2015	Not available
RBG5789	RBG5789	(non-pathogen, from soil)	-	-	Australia	Rocha et al., 2015	Not available
RBG5791	RBG5791	(non-pathogen, from soil)	-	-	Australia	Rocha et al., 2015	Not available
RBG5798	RBG5798	(non-pathogen, from soil)	-	-	Australia	Rocha et al., 2015	Not available
RBG5820	RBG5820	(non-pathogen, from soil)	-	-	Australia	Rocha et al., 2015	Not available
RBG5824	RBG5824	(non-pathogen, from soil)	-	-	Australia	Rocha et al., 2015	Not available
RBG5827	RBG5827	(non-pathogen, from soil)	-	-	Australia	Rocha et al., 2015	Not available
F. proliferatum Fol3	Fol3	(from *Lilium* sp.)	-	-	Netherlands	van Dam and Rep, 2017	NJCT01000000
Fusarium sp. Na10	Na10	(from *Narcissus* sp.)	-	-	Netherlands	van Dam and Rep, 2017	NJCS01000000

NAKT: NAKtuinbouw, Netherlands Inspection Service for Horticulture, Roelofarendsveen, the Netherlands
BL: Bart Lievens, Scientia Terrae Research Institute, Belgium
JVD: Joop van Doorn, PPO Research Centre, Lisse, The Netherlands
ML: Matthew Laurence, Plant Disease Diagnostic Unit of the Royal Botanic Gardens and Domain Trust, Sydney, Australia
ARS: Agricultural Research Service, USDA USA
IPO: Plant Research International (formerly Instituut voor Planteziektenkundig Onderzoek), Wageningen, The Netherlands
KOD: Kerry O'Donnell, USDA ARS, Peoria IL, USA

Due to size restrictions, Supplementary Data S2 (alignments of selected marker sequences with primers highlighted in them) is not included in this thesis, but can be accessed online via http://gdurl.com/M-TI.
(or https://drive.google.com/file/d/0B29G0ryqK2xgWUNGellwUEhIMWs/view?usp=sharing)



Supplementary data S3: Phylogenetic tree of genes selected as marker genes for cucurbit infecting *formae speciales*.
Separation of a clade that includes all strains belonging to a *forma specialis* indicates sequence similarity within and sequence dissimilarity between *formae speciales*. Coloured highlights in the tree reflect the target f. sp. of the marker.



Supplementary data S3 (continued)

Supplementary data S4: Average disease index found for the bio-assays performed on strains that were suspected to belong to a different *forma specialis* than reported.
Values given are the average disease indices (0-4) that were given to 5-8 individual plants, 2 weeks post inoculation. N.t. = not tested.

Isolate	Reported f. sp.	*Cucumis sativus* cv Paraiso	*Cucumis melo* cv Cha-T	*Citrullus lanatus* cv Black Diamond	Conclusion
14150	*cucurbitacearum*	3.2	4	n.t.	*radicis-cucumerinum*
Focuc014	*cucumerinum*	2.63	3.88	n.t.	*radicis-cucumerinum*
Focuc022	*cucumerinum*	0	n.t.	n.t.	*non-pathogenic*
Focuc028	*cucumerinum*	0	n.t.	n.t.	*non-pathogenic*
Focuc039	*cucumerinum*	0	0	n.t.	*non-pathogenic*
Focuc040	*cucumerinum*	3.2	4	4	*radicis-cucumerinum*
Fomln002	*melonis*	n.t.	0.25	n.t.	*non-pathogenic*
Fomln017	*melonis*	0	4	0	*melonis*
Fomln021	*melonis*	n.t.	4	n.t.	*melonis*
Fomln023	*melonis*	n.t.	3.63	3	*melonis/niveum*
Fomln024	*melonis*	n.t.	4	n.t.	*melonis*
Fomln025	*melonis*	n.t.	3.75	n.t.	*melonis*
Fomln026	*melonis*	n.t.	4	n.t.	*melonis*
Foniv034	*niveum*	n.t.	n.t.	0	*non-pathogenic*
Foniv035	*niveum*	n.t.	n.t.	0	*non-pathogenic*
Foniv038	*niveum*	n.t.	n.t.	0	*non-pathogenic*

Supplementary data S5: Primers and annealing temperatures used in this study

Gene ID	Target gene[a]	Target f. sp.	Primer name	Primer sequence (5'-3')	T_{Anneal}	Product size (nt)
+	*FEM1*	Pos. ctrl	fp157	ATGAAGTACACTCTCGCTACC	54°C	274
			fp158	GGTGAAAGTGAAAGAGTCACC		
94	HPEG	all cucurbit infecting	fp7304	GCCTCATTGAAGTTTCAACA	54°C	346
			fp7321	TGGTAAAGGACACGACCATT		
13	*SIX13*	Forcu	fp7305	TTGCCCAAAATGGCATGTTT	56°C	328
			fp7322	CATTGACACTGTAAGTGGG		
70	HPEG	Forcu	fp7306	TACAACCTCTCTCTTTCCTT	54°C	454
			fp7323	GCTGAATTCTAGCAGAGAAT		
66	HPEG	Focuc	fp7307	CCGTTATGGCCAGAGATC	54°C	425
			fp7324	CCAACAAACAGAGCAAAACTAA		
99	HPEG	Focuc	fp7308	CTACCAATCTCTCCTGAGTG	54°C	445
			fp7325	GTCGATTGCAGTGCTAGTCT		
21	Fom Effector 7	Focuc	fp7309	CAGTCTAACCCTGTCTCATT	54°C	381
			fp7326	CGCCAATAGATAGTGATGGA		
1	*SIX1*	Fomln	fp7310	CCTCTCAGTCCTTGGGTCT	54°C	397
			fp7327	ACTCGCTTCAGCTTACCGA		
20	Fom effector 6	Fomln	fp7406	TGAAAGTCTTGGCGGGTGT	56°C	305
			fp7328	TCCTCTCCATCCTCATCAGT		
18	Fom effector 3	Fomln+ Foniv	fp7312	TTAGTGCAGCTTTTCTCCTC	54°C	299
			fp7329	AGTGGTTAGTCAAGTGGTAA		
99	HPEG	Foniv	fp7313	TGCCGGGCTAGTTAATATAGT	54°C	406
			fp7330	ACCATTTTTCTGTTGGGGTTG		
100	HPEG	Foniv	fp7314	ATTTTGCTAGCTTCAGCAGTT	54°C	482
			fp7331	ATCCTGAACGGTGACTAGAG		
21	Fom Effector 7	Foniv	fp7315	CGCTCGCTATAATTCAAACG	54°C	139
			fp7332	GGAGGAGCACTACAACTAAT		
71	HPEG	Folag	fp7407	TAGTCCAATCTGCCTCAGCAA	54°C	270
			fp7410	GGAAGTGAGCATTCTTCCGTA		
99	HPEG	Folag	fp7408	TCGTATCTCTCAGTAGTATGG	54°C	367
			fp7411	AATGGATACCTTATAAGGGCT		
1	*SIX1*	Folag + Fomom	fp7409	TTGGGATTGCGGCTTATGCT	56°C	463
			fp7412	AAAGTGGTACACTCCGTGC		
98	HPEG	Fomom	fp7318	AGGTGCAGCGTTTTTAGGT	60°C	469
			fp7335	GAGGGCTGGTTGAGAACTA		
130	HPEG	Fomom	fp7319	TCTACGCTTCGAGGATGGTA	56°C	368
			fp7336	TCGTTTAGACGACTACAACC		
99	HPEG	Foluf	fp7320	TACTCTCCTAGAGTCAGTCT	54°C	606
			fp7337	CACGCCATCATCCTTTATTC		

[a] HPEG: Hypothetical protein encoding gene

Supplementary data S6: Primers and probes used for the TaqMan experiments

Gene ID	Target gene[a]	Target f. sp.	Primer name	Oligo sequence (5'-3')	Product size (nt)
21	Fom Effector 7	Foniv	fp7589	CCGGTACCCCAGCTTTATGT	116
			fp7590	CAGCAACGTTCTGAAAGCGT	
			probe_3	**HEX**-TGCAGGTTGGCAGGCCCCTG-**BHQ1**	
100	HPEG	Foniv	fp7591	CACCAACAACTATGCGGCAC	138
			fp7592	GCAATTGACCCAGCTGCAAT	
			probe_4	**FAM**-AGTCGCCGGCCACCACATTGA-**BHQ1**	
EF1α	Elongation Factor 1 alpha	All strains	fp7710	CGCTGAGCTCGGTAAGGG	97
			fp7711	CCAGAGAGCAATATCGATGGTGA	
			probe_7	**TAMRA**-ACGCCTGGGTTCTTGACAAGCTCA-**BHQ2**	

[a] HPEG: Hypothetical protein encoding gene

Chapter 5

The distribution of *miniature impala* elements and *SIX* genes in the *Fusarium* genus is suggestive of horizontal gene transfer

This chapter has been published as:

Peter van Dam and Martijn Rep (2017) The Distribution of *Miniature Impala* Elements and *SIX* Genes in the *Fusarium* Genus is Suggestive of Horizontal Gene Transfer. *Journal of Molecular Evolution*. doi:10.1007/s00239-017-9801-0

The distribution of *miniature impala* elements and *SIX* genes in the *Fusarium* genus is suggestive of horizontal gene transfer

Abstract

The *mimp* family of miniature inverted-repeat transposable elements was previously found only in genomes of *Fusarium oxysporum* and is contextually associated with virulence genes in this species. Through extensive comparative analysis of 83 *F. oxysporum* and 52 other *Fusarium* genomes, we uncovered the distribution of different *mimp* families throughout the genus. We show that i) *mimps* are not exclusive to *F. oxysporum*; ii) pathogenic isolates generally possess more *mimps* than non-pathogenic strains and iii) two isolates of *F. hostae* and one *F. proliferatum* isolate display evidence for horizontal transfer of genetic material to or from *F. oxysporum*. Multiple instances of *mimp* elements identical to *F. oxysporum mimps* were encountered in the genomes of these isolates. Moreover, homologs of effector genes *(SIX1, 2, 6, 7, 11* and *AVRFOM2)* were discovered here, several with very high (97-100%) pairwise nucleotide sequence identity scores. These three strains were isolated from infected flower bulbs (*Hyacinthus* and *Lilium* spp.). Their ancestors may thus have lived in close proximity to pathogenic strains of *F. oxysporum* f. sp. *hyacinthi* and f. sp. *lilii*. The Fo f. sp. *lycopersici SIX2* effector gene was found to be widely distributed (15/18 isolates) throughout the *F. fujikuroi* species complex, exhibiting a predominantly vertical inheritance pattern. These findings shed light on the potential evolutionary mechanism underlying plant-pathogenicity in *Fusarium* and show that interspecies horizontal gene transfer may have occurred.

Introduction

Transposable elements (TEs) are DNA sequences that can duplicate or move from one site to another within a genome. Two different TE classes are distinguished based on their transposition intermediate, RNA or DNA. Class I transposons (also called retrotransposons) transpose by transcription into an RNA intermediate and reverse transcription into cDNA before insertion into a new site. Class II TEs (or DNA transposons) on the other hand, transpose through a "cut-and-paste" mechanism. This latter class of TEs is flanked by Terminal Inverted Repeats (TIRs) that facilitate the recognition for DNA excision.

Miniature inverted-repeat transposable elements (MITEs) are short (<500 base pairs (bp)) non-autonomous class II TEs (Bergemann *et al.*, 2008). Their structure resembles defective DNA transposons and they are thought to originate through the deletion of the transposase open reading frame (ORF) between the TIRs (Feschotte and Pritham, 2007). Several studies have shown that MITEs can be mobilized by full-length class II TEs (Feschotte *et al.*, 2002; Dufresne *et al.*, 2007; Bergemann *et al.*, 2008).

A MITE called *mimp* (for *miniature impala)* has so far only been described in the filamentous fungus *Fusarium oxysporum* (Fo). Using different approaches, six families of *mimps* have been described in the reference genome *F. oxysporum* f. sp. *lycopersici* 4287, based on the consensus sequence of the inverted repeats (Bergemann *et al.*, 2008). They are characterized by a uniformity in size of about 180-220 bp (Dufresne *et al.*, 2007) and appear to have originated from full-length *impala* elements.

The *impala* family of transposons belongs to the *Tc1/mariner* superfamily of class II transposons (Hua-Van, Langin, *et al.*, 2001). This particular TE occurs at a low copy number in the genome of *F. oxysporum*. Although 1-5 copies were detected in most isolates and the TE has been described as an ancient component of the *F. oxysporum* genome (Hua-Van, Langin, *et al.*, 2001), they have also been found to be absent in some isolates (Hua-Van, Pamphile, *et al.*, 2001). *Impala* elements contain a single ORF encoding a transposase of 340 amino acids flanked by TIRs of 37 bp (Hua-Van, Langin, *et al.*, 2001). They have been shown to be active in at least some strains of *F. oxysporum* (Hua-Van, Pamphile, *et al.*, 2001). Interestingly, the transposase remains functional when transferred into the genome of closely related *(F. moniliforme*, *F. culmorum* and *F. graminearum* (Hua-Van, Pamphile, *et al.*, 2001; Dufresne *et al.*, 2007; Spanu *et al.*, 2012)) as well as more distantly related fungal species (*Magnaporthe grisea, Aspergillus nidulans, A. fumigatus, Colletotrichum gloeosporioides* and *Penicillium griseoroseum* (Villalba *et al.*, 2001; Hua-Van *et al.*, 2002; de Queiroz and Daboussi, 2003; Firon *et al.*, 2003; Li Destri Nicosia *et al.*, 2004)). Reinsertion in the genome occurs at TA residues, that are duplicated upon insertion (Dufresne *et al.*, 2007).

Genomes of *F. oxysporum* strains are divided into two compartments. A set of conserved 'core' chromosomes is dedicated to housekeeping and vegetative growth, while one or several accessory chromosomes harbour high numbers of TEs and sometimes large segmental

duplications. These accessory chromosomes are sometimes directly linked to virulence of the isolate due to the presence of virulence (effector) genes on these chromosomes (Ma *et al.*, 2010). Moreover, they can be horizontally transferred from pathogenic to non-pathogenic strains, thereby conferring the host-specific pathogenicity upon the recipient strain (Ma *et al.*, 2010). In the genome of Fo f. sp. *lycopersici* 4287, 95% of the class II TEs and the majority of *mimps* are present on the accessory chromosomes (Schmidt *et al.*, 2013). Intriguingly, they were found to be significantly overrepresented in the promoter regions (<1500 bp) of known effector genes (named *SIX,* for *Secreted In Xylem*) and other genes that are expressed during plant infection (Schmidt *et al.*, 2013). This association with virulence genes was used to predict novel candidate effectors in the genomes of Fo f. sp. *lycopersici*, Fo f. sp. *melonis*, Fo f. sp. *cucumerinum*, Fo f. sp. *radicis-cucumerinum* and Fo f. sp. *niveum* (Schmidt *et al.*, 2013, 2016; van Dam *et al.*, 2016).

The goal of the current study was to evaluate the distribution of different classes of *mimps* throughout the *Fusarium* genus based on published and novel whole genome sequences. We searched for *mimp*-like elements from the genomes of isolates belonging to six different *Fusarium* species complexes. We find that *mimp* elements are not exclusive to the *F. oxysporum* species complex. Moreover, we find that several *SIX* genes are present in non-*oxysporum Fusarium* strains that also have many *mimps*. Based on these results we explore the possibility of horizontal transfer of genetic material between *Fusarium* species.

Results

Identification of *mimps* in whole genome assemblies

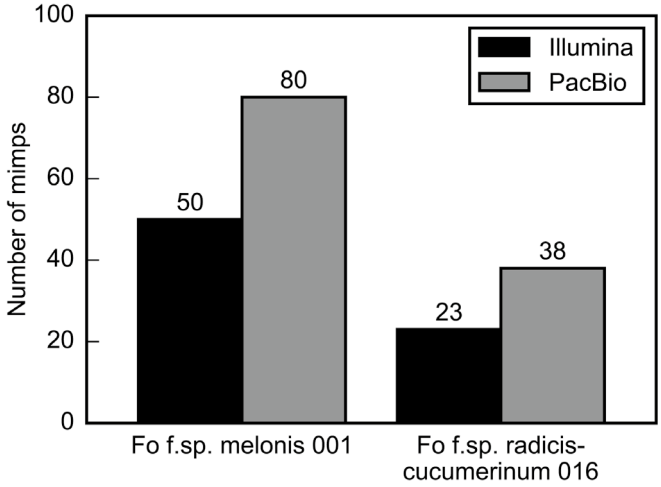
Based on the described inverted repeats in Bergemann *et al.*(2008), we extracted the sequences of *mimps* from all currently available *Fusarium* genome assemblies. These include 83 *F. oxysporum* genomes and 52 genomes from other *Fusarium* species (Supplemental Table S1). We used a consensus sequence generated from the first 16 nucleotides of TIRs of all six previously described *mimp* families to search for the presence of *mimp*-like elements in each of the genomes ('AGT[GA][GA]G[GAT][TGC]GCAA[TAG]AA'). Stretches of sequence where an instance of this motif was found within 400 bp of another instance in reverse orientation were extracted. In total, 2,688 *mimps* were identified. The vast majority (2,572) of these were extracted from *F. oxysporum* genomes.

On average, 31 intact *mimps* were found per *F. oxysporum* genome, with numbers ranging from zero in *F. oxysporum* f. sp. *cubense* N2 and two copies in *F. oxysporum* f. sp. *cubense* B2 to 74 in *F. oxysporum* f. sp. *raphani* PHW815. Other isolates where few *mimps* were encountered include *F. oxysporum* f. sp. *nicotianae* and non-plant pathogenic isolates such as MN14 (saprophytic strain isolated from tomato), FOSC3-a (a clinical isolate) and Fo47 (a biocontrol strain) (Supplemental Table S1). These latter three isolates were previously shown to possess

relatively few candidate effector genes and also lack copies of *SIX* virulence genes (van Dam *et al.*, 2016). The reason for the absence of a high number of *mimps* and putative virulence genes in these isolates may be attributed to a smaller amount of accessory material.

The degree of genome assembly fragmentation influences number of *mimps* identified

Fusarium genome assemblies generated from short-read sequence data are typically assembled into hundreds or even thousands of contigs of 500 bp or larger. Especially repeat-rich regions such as the accessory chromosomes of *F. oxysporum*, where most of the *mimps* are typically located (Schmidt *et al.*, 2013; Kang *et al.*, 2014), are highly fragmented. As most of the genomes in our dataset were sequenced with Illumina short read technology, we wondered whether the number of *mimps* was underestimated. We therefore compared the number of *mimps* identified in the Illumina assemblies of two individual *F. oxysporum* isolates to their respective long-read assemblies that were generated from PacBio sequencing data (van Dam *et al.*, accepted). A higher number of intact *mimps* was indeed identified in the PacBio assemblies of *F. oxysporum* f. sp. *melonis* 001 (60% more; 80/50) and *F. oxysporum* f. sp. *radicis-cucumerinum* 016 (65% more; 38/23) (Fig. 1).



◀ **Fig. 1: 60% (Fo f. sp. *melonis* 001) and 65% (Fo f. sp. *radicis-cucumerinum* 016) more intact *mimps* were found in PacBio assemblies compared to Illumina assemblies.**

Distribution of mimps in the *Fusarium* genus

Although *mimps* have thus far only been described in the *F. oxysporum species complex,* we find that they are not exclusive to this species complex. The genomes of most other *Fusarium* species did show complete absence of *mimps*. However, occurrence of one or a few elements was identified in amongst others *F. verticillioides, F. proliferatum, F. nygamai* and *F. avenaceum* (Fig. 2). Two previously sequenced and four *de novo* sequenced non-*F. oxysporum* genomes stood out in the analysis because they displayed a high number of *mimps*, similar to the numbers found in *F. oxysporum*. These *Fusarium* strains were all isolated from diseased bulb

flowers that were affected by bulb rot or leaf and stem spot. *F. proliferatum* Fol3 (isolated from *Lilium*) was earlier identified as *F. oxysporum* (Baayen *et al.*, 1998) but is now reclassified as *F. proliferatum* based on the concatenated sequence of the *EF1alpha, RPB1* and (partial) *RPB2* genes (Fig. 1). *F. hostae* isolates Hy9 and Hy14 were isolated from diseased *Hyacinthus* bulbs (Breeuwsma and De Boer, 2004), *F. agapanthi* NRRL31653 and NRRL54464 were isolated from diseased *Agapanthus* plants (African lily) (Edwards *et al.*, 2016) and *Fusarium* sp. Na10 was isolated from disaesed *Narcissus* bulbs (Breeuwsma and De Boer, 2004).
To identify whether a potentially active intact *impala* was present in the genome, a TBLASTN search was performed using the full-length FOM24 *impala* transposase ORF as a query (Fig. 2). Many *F. oxysporum* isolates showed mutations of the transposase, inducing a premature stop codon in the ORF. However, several isolates such as Fo f. sp. *conglutinans* PHW808, Fo f. sp. *melonis* 009 and Fo47 still possess an intact transposase ORF. Intriguingly, the non-FOSC bulb-infecting isolates *F. hostae* Hy9 and Hy14 as well as *F. proliferatum* Fol3 also contain a largely intact transposase ORF, only interrupted by the end of the contig or by a stretch of ambiguous nucleotides (Ns) caused by contig scaffolding (Fig. 2). This means that an intact *impala* may be present in these isolates. *Fusarium* sp. Na10 and both *F. agapanthi* isolates did not return a significant hit.
Based on reciprocal BLAST hits, all the extracted *mimps* were classified into either of the six previously described families, or novel (unclassified) families and plotted next to a phylogenetic tree of a subselection of the genomes (Fig. 2). The most common *mimp* families are families 1, 2, 3 and 4, that on average make up 86% of the total *F. oxysporum mimp* content. The largest *mimp* family identified in the two *F. agapanthi* isolates (type '09') was not very common in the FOSC. *Mimp* categories in the genomes of bulb-infecting isolates Fol3, Hy9 and Hy14 showed a very similar distribution of families to that found in most *F. oxysporum* isolates. To investigate whether *mimp* elements have been subject to horizontal transfer (HT), we examined their nucleotide sequence in greater detail.

Mimp elements identical to *F. oxysporum mimps* occur outside the FOSC

In order to find out whether some of the *mimps* identified in *F. hostae* Hy9/Hy14, *F. proliferatum* Fol3, *Fusarium* sp. Na10 or either of the *F. agapanthi* genomes were (nearly) identical to a copy in one of the 83 *F. oxysporum* genomes, the sequence of each of their *mimps* was compared to all *F. oxysporum mimps* in pairwise comparisons. We found that *F. hostae* Hy9 has five and Hy14 six *mimps* that are 100% identical to copies found in *F. oxysporum* f. sp. *vasinfectum*, *raphani, conglutinans, pisi* and *tulipae* (Fig. 3A and B). *F. proliferatum* Fol3, isolated from infected lily bulbs, has three *mimps* that are identical to a *F. oxysporum* copy. Interestingly, all three of these elements matched with *mimps* in the genomes of *F. oxysporum* f. sp. *lilii* Fol39 and *F. oxysporum* f. sp. *gladioli* G2, isolates that are also pathogenic to bulb flowers.
The *mimps* belonging to families 2, 3, 4, 5 and category '09' that were identified in the 47

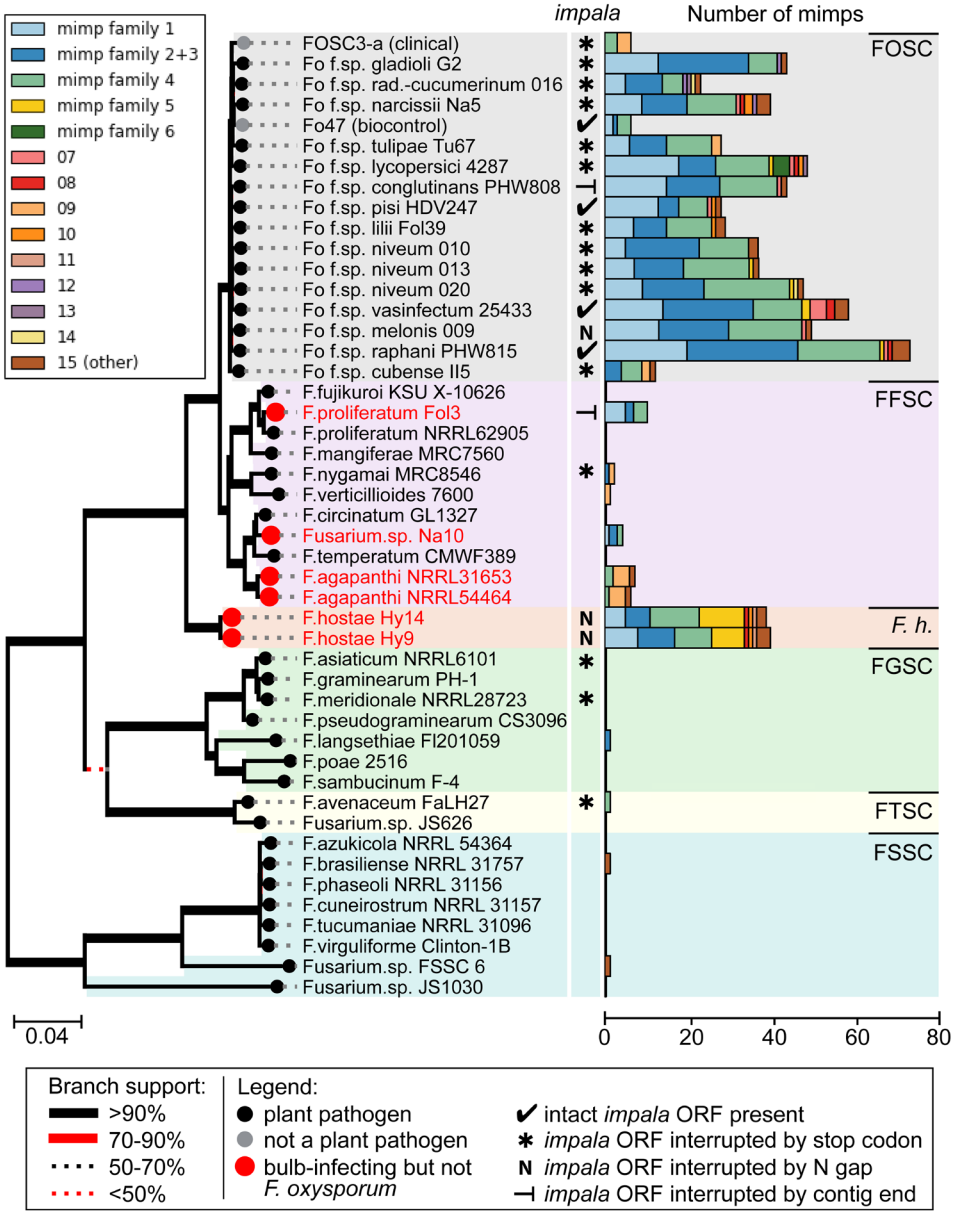


Fig. 2: Most *Fusarium* species outside the FOSC lack *mimps* in their genome, but a relatively high number of *mimps* was found in bulb-infecting isolates of *F. hostae* and the FFSC.
At least one representative genome was selected per *Fusarium* species. Phylogeny was inferred from a concatenated sequence alignment of three conserved genes: *EF1α, RPB1* and *RPB2* (partial), using 100 bootstrap replicates. TBLASTN (evalue < 1e-100) was performed to identify intact *impala* open reading frames in the assemblies. Intact *mimps* were divided into categories and these were plotted next to the phylogenetic tree. In total, the analysis covered six *Fusarium* species complexes (FOSC = *F. oxysporum species complex;* FFSC = *F. fujikuroi species complex;* F. h. = *F. hostae;* FGSC = *F. graminearum species complex;* FTSC = *F. tricinctum species complex*; FSSC = *F. solani species complex*).

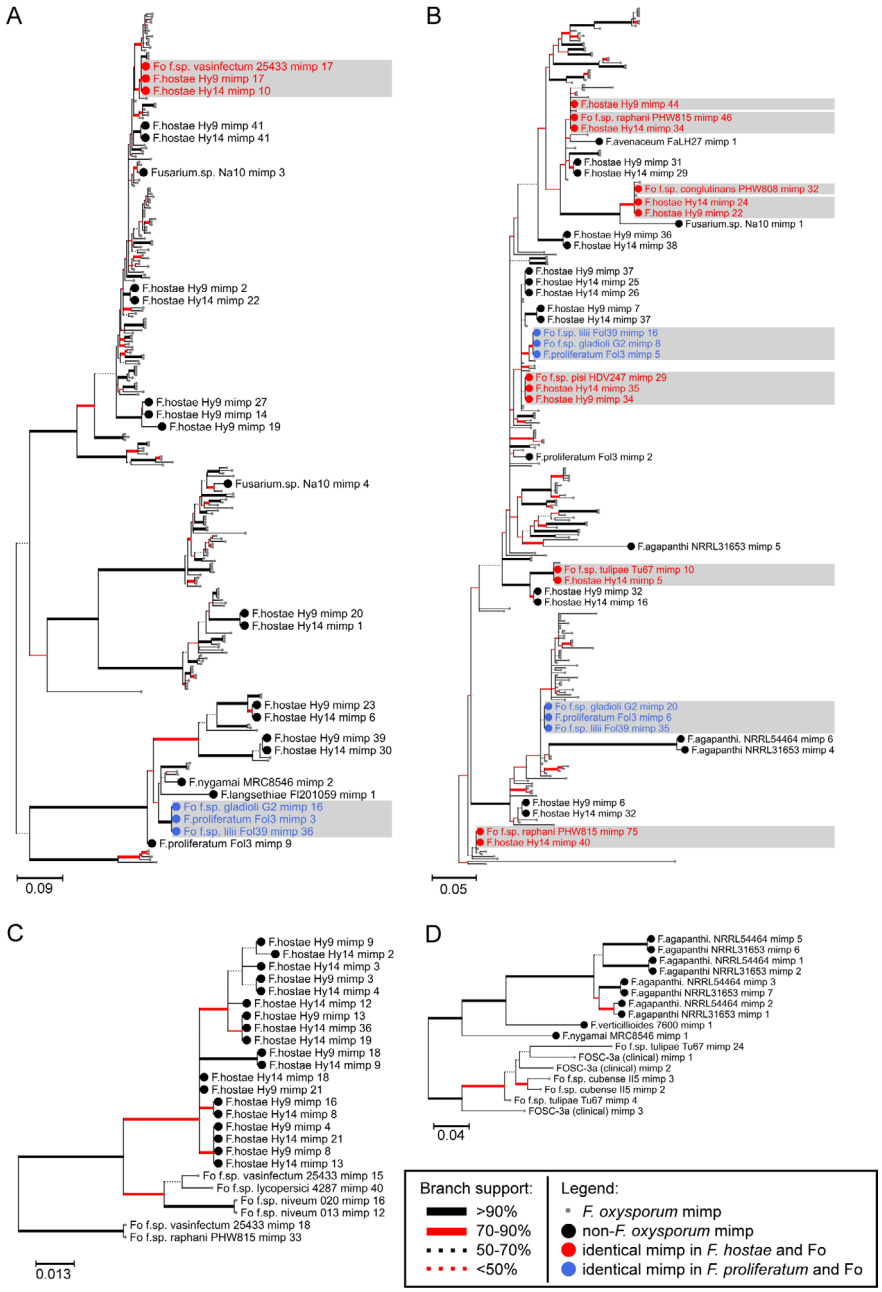


Fig. 3: Phylogenies of *mimps* belonging to four different families show the presence of *mimps* in *F. hostae* and *F. proliferatum* Fol3 that are identical to elements found in pathogenic *F. oxysporum* isolates.
Sequence alignment was performed on *mimp* sequences extracted from 49 *Fusarium* genomes belonging to **(A)** *mimp* family 2/3, **(B)** family 4, **(C)** family 5 and **(D)** category '09' with MAFFT and phylogeny was inferred using PhyML with 100 bootstraps. *Mimps* extracted from *F. hostae* and *F. proliferatum* genomes having a 100% identity match in a *F. oxysporum* genome are highlighted in red and blue, respectively. Node annotations of other *F. oxysporum mimps* have been omitted for clarity.

genomes shown in Fig. 2 were aligned per family and visualized in a phylogenetic tree (Fig. 3). This analysis shows that several *F. hostae mimps* and *mimps* extracted from *Fusarium* sp. Na10, *F. proliferatum* Fol3 and *F. nygamai* MRC8546 are present in clades close to *F. oxysporum* elements (Fig. 3A and B), meaning that they are highly similar. Additionally, most *F. agapanthi mimps* (Fig. 3D) and a large number of *F. hostae mimps* (Fig. 3C) are relatively distantly related to *F. oxysporum mimps*, indicating that they might have evolved separately in these species.

SIX genes identified in *Fusarium* species outside the FOSC

Several sequences with high levels of similarity to a *F. oxysporum SIX* effector gene were identified in Fol3, Hy9 and Hy14. Additionally *AVRFOM2*, encoding a small secreted protein with an avirulence function in the Fo f. sp. *melonis* – muskmelon pathosystem, was identified in these strains (Table 1). Both *SIX1* and *AVRFOM2* are 100% identical between *F. proliferatum* Fol3 and Fo f. sp. *nicotianae* isolates and several hits in *F. hostae* show a nucleotide similarity of 97% or higher. FOMG_19741 is a gene that was identified as an effector candidate in Fo f. sp. *melonis* based on contextual association with a *mimp* (Schmidt *et al.*, 2016). The reason for including it in Table 1 is the fact that the copy found in Fol3 (isolated from *Lilium* sp.) is highly similar (96.7% identity) to that of Fol39, a Fo f. sp. *lilii* isolate.

Table 1: Presence of *F. oxysporum SIX* homologs in *F. proliferatum*, *F. hostae, F. agapanthi* and *Fusarium* sp. isolates indicates possible horizontal transfer of these genes.

	Core genes *EF1alpha, RPB1, RPB2*	*SIX1*	*SIX2*	*SIX6*	*SIX7*	*SIX11*	*AVRFOM2*	FOMG_ 19741
***F. proliferatum* Fol3**	96.8% on average to Fo	100%[d] to Fonic 001	73.4% to Focub II5				100% to Fonic 003	96.7% to Folil Fol039
***F. hostae* Hy9**	95.5% on average to Fo		86.3% to Focub II5	94.9% to Foniv 019	95.2% [d] to Folil Fol39 [b]	97.6% to Fogla G14 [a]	97.4% to Fonic 003	
***F. hostae* Hy14**	95.5% on average to Fo		86.3% to Focub II5	94.9% to Foniv 019	95.2% [d] to Folil Fol39 [c]	97.6% to Fogla G14 [a]	97.4% to Fonic 003	
***Fusarium* sp. Na10**	96.8% on average to Fo		81.4% to Focub II5					
***F. agapanthi* NRRL31653**	97.1% on average to Fo		76.6% to Focub II5			96.7% to Fogla G14		
***F. agapanthi* NRRL54464**	97.1% on average to Fo		76.5% to Focub II5			96.7% to Fogla G14		

Pairwise identity percentages are based on ClustalO sequence alignments with the best hit of a *F. oxysporum* homolog (Fonic: *F. oxysporum* f. sp. *nicotianae*, Focub: *F. oxysporum* f. sp. *cubense*, Foniv: *F. oxysporum* f. sp. *niveum*, Folil: *F. oxysporum* f. sp. *lilii*, Fogla: *F. oxysporum* f. sp. *gladioli*)

[a] a *mimp* was identified on the same contig as this homolog

[b] two *mimps* were identified on the same contig as this homolog

[c] three *mimps* were identified on the same contig as this homolog

[d] this gene is interrupted by multiple stop codons and is probably a pseudogene

Strikingly, *SIX7* was only discovered in Fo f. sp. *lycopersici* and Fo f. sp. *lilii* Fol39 as an intact ORF. A pseudogenized version of the gene was found in *F. hostae* Hy9/Hy14, as well as Na5 (f. sp. *narcissii*) and G14/G2 (f. sp. *gladioli*); all isolated from diseased flower bulbs. The homologs of *SIX1* (in *F. proliferatum* Fol3) have also undergone mutations resulting in multiple stop codons in its ORFs (Fig. 4A). For the other effector homologs *(SIX2, SIX6, SIX11, AVRFOM2* and FOMG_19741), manual inspection showed that these were intact and potentially active ORFs. The position of *F. proliferatum* Fol3 *SIX1* (pseudogenized, Fig. 4A) and *F. hostae SIX6* (Fig. 4B) in the phylogenetic tree shows them nested within the tree of all *F. oxysporum* homologs.
Comparison of the scaffolds on which *SIX1* and *AVRFOM2* are located in *F. proliferatum* Fol3 and Fo f. sp. *nicotianae* 003 by nucmer (MUMmer3) alignment shows that in both cases a 2 kb region containing the ORFs is 100% identical between these strains (Fig. 5). Multiple N-gaps are present in both assemblies, indicating the presence of many repetitive, difficult to assemble sequences (potentially transposons). Nonetheless, the successfully assembled stretched of sequence between these gaps are highly conserved between the two investigated strains, with only one pairwise alignment of 340 bp having a sequence identity score of less than 100% (95.9%). In the rest of the alignments no nucleotide differences were found between Fol3 and Fonic003. This level of sequence conservation is not expected between *F. oxysporum* and *F. proliferatum* assuming vertical inheritance (average sequence identity in conserved genes between Foniv015 and Hy14 is 96.1%).

Vertical inheritance of *SIX2*

Within the FOSC, *SIX2* has only been identified in ff. spp. *lycopersici* and *cubense*. This gene, unlike the other 13 *SIX* genes, is very widespread throughout the *Fusarium fujikuroi species complex* (FFSC). Next to the two *F. agapanthi* strains*, F. proliferatum* Fol3*, F. hostae* Hy9/Hy14 and *Fusarium.* sp. Na10 that are mentioned in Table 1, a *SIX2* homolog was also identified in *F. circinatum* FSP34, *F. circinatum* GL1327, *F. fujikuroi* B14, *F. fujikuroi* IMI 58289, *F. fujikuroi* KSU 3368, *F. fujikuroi* KSU X-10626, *F. mangiferae* MRC7560, *F. temperatum* CMWF389, *F. verticillioides* 7600, *F. fujikuroi* CF-295141 and another *F. proliferatum* strain in our dataset: NRRL62905 (Fig. 6). All open reading frames were intact. The *SIX2* homologs in *F. proliferatum* strains Fol3 and NRRL62905 are identical. Moreover, comparison of the phylogenetic distribution of *SIX2* to the core phylogeny (based on the concatenated sequence alignment of *EF1alpha, RPB1* and part of *RPB2*) shows that these trees are largely congruent with each other. Fo f. sp. *cubense* II5 and B2 form a notable exception to this, since they have a *SIX2* homolog that is only 69.4% identical to *SIX2* in Fo f. sp. *lycopersici* and that clusters closer to *SIX2* homologs in other *Fusarium* species (Fig. 6).

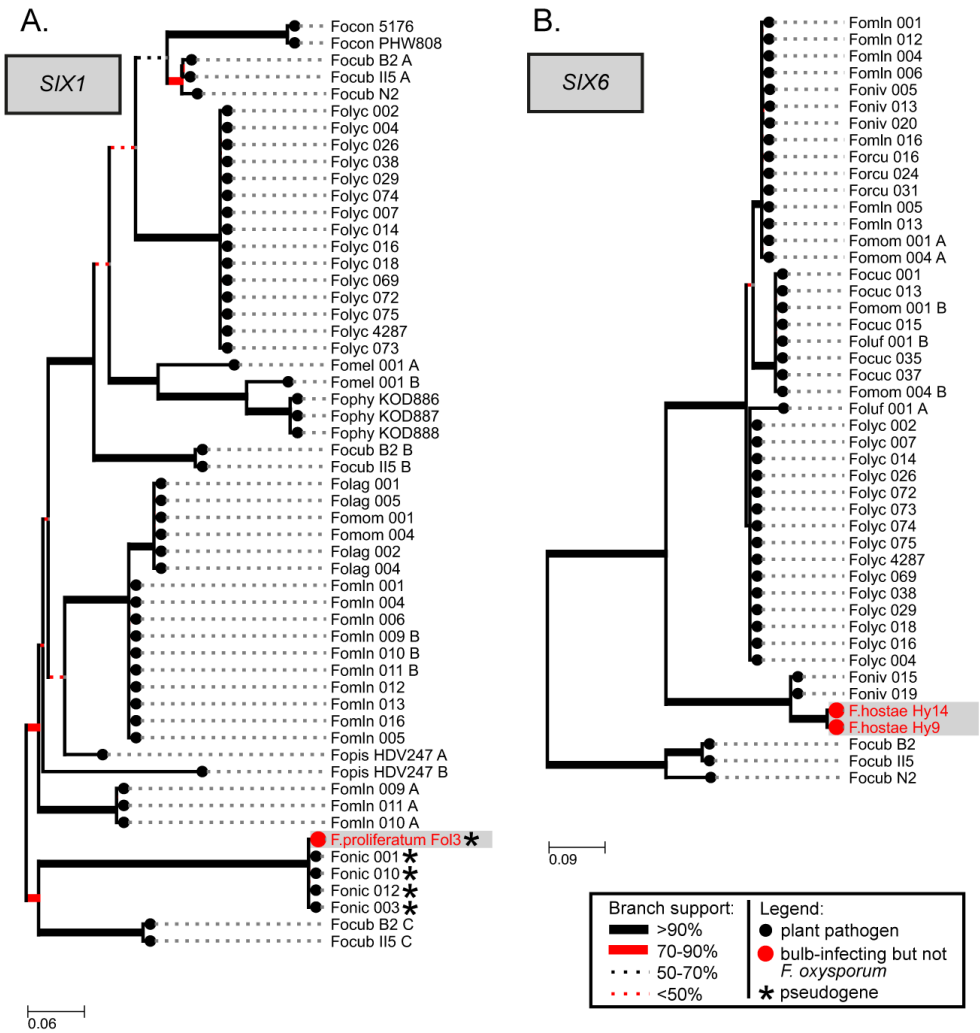
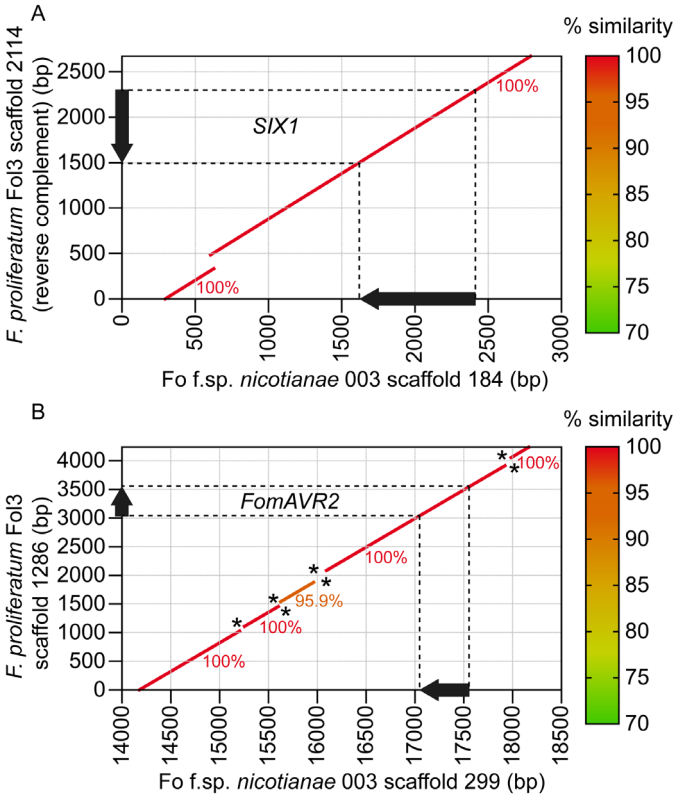


Fig. 4: *F. proliferatum* Fol3 has a (pseudogenized) *SIX1* homolog that is identical to Fo f. sp. *nicotianae* SIX1 and *F. hostae SIX6* clusters within the *F. oxysporum* clade, close the *SIX6* gene of *Fo f. sp. niveum 015 and 019.* Clustering of the non-*oxysporum SIX* genes among copies of *F. oxysporum SIX* genes supports the hypothesis of horizontal transfer. Nucleotide sequences were aligned with ClustalO and phylogeny was inferred using PhyML with 100 bootstraps (Fonic: Fo f. sp. *nicotianae*, Focub: Fo f. sp. *cubense*, Foniv: Fo f. sp. *niveum*, Fomln: Fo f. sp. *melonis*, Fopis: Fo f. sp. *pisi*, Folag: Fo f. sp. *lagenariae*; Fomom: Fo f. sp. *momordicae*, Fophy: Fo f. sp. *physali*, Fomel: Fo f. sp. *melongenae*, Focon: Fo f. sp. *conglutinans*, Forcu: Fo f. sp. *radicis-cucumerinum*, Focuc: Fo f. sp. *cucumerinum*).



◀ **Fig. 5:**
(A) *SIX1* (pseudogenized) and (B) *AVRFOM2* are located within regions of 100% nucleotide sequence identity between Fol3 and Fonic003.
Alignments were made with nucmer (with -breaklen 100 to break alignments separated by N-gaps resulting from contig scaffolding). An asterisk (*) positioned above the diagonal indicates an N-gap in Fol3 and an asterisk below the diagonal indicates an N-gap in Fonic003. The position of the *SIX1* and *AVRFOM2* ORFs is indicated on the axes with a black arrow. For visualization reasons, the *x*-axis is only partially displayed; Fonic scaffold 184 is 11,173 bp and scaffold 299 is 20,768 bp long.

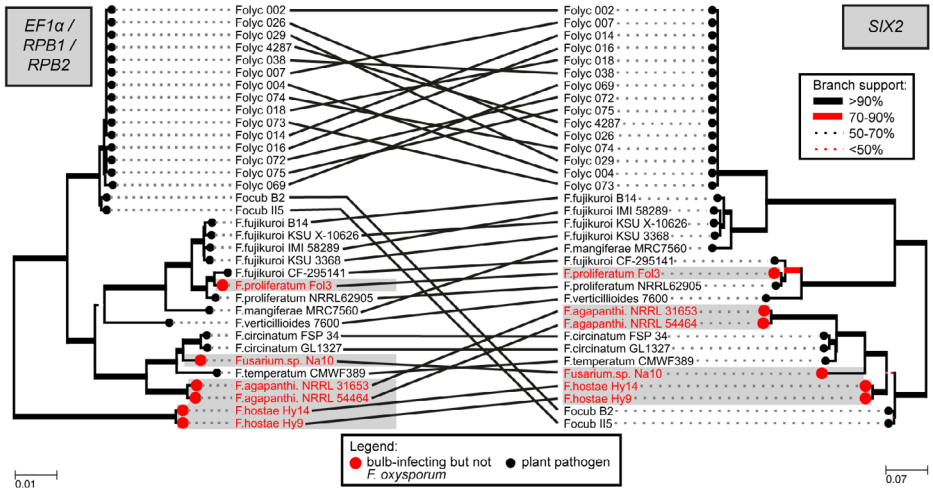


Fig. 6: A tanglegram showing that the phylogeny of *SIX2* is largely congruent with the core phylogeny of *Fusarium* species that have a *SIX2* homolog.
SIX2 occurs in Fo f. sp. *lycopersici*, Fo f. sp. *cubense, F. hostae* and numerous *Fusarium* species belonging to the FFSC. Nucleotide sequences were aligned with ClustalO and phylogeny was inferred using PhyML with 100 bootstraps (Focub: Fo f. sp. *cubense*; Folyc: Fo f. sp. *lycopersici*).

Discussion

In this study, we performed genome-wide comparative analysis of *mimp* elements and effector genes in the *Fusarium* genus. *Mimps* have so far only been described in *F. oxysporum* and are contextually associated with virulence-associated genes in this species (Schmidt *et al.*, 2013). Within *F. oxysporum*, *mimps* occur most frequently in the accessory chromosomes of plant-pathogenic strains (Bergemann *et al.*, 2008; Schmidt *et al.*, 2013). Bergemann *et al.* find that the enrichment in Fol4287 occurs only on "pathogenicity" chromosome 14 but not on chromosomes 3 and 6, that are also considered accessory but consist mostly of large segmental duplications. In another study, Dufresne *et al.*(2007) demonstrated that *mimp1* could be mobilized through the action of the transposase of *impalaE*, a *Tc1*-like autonomous element.

By searching for the conserved inverted repeat sequences that have been described for six families of *mimps* in 83 *F. oxysporum* and 52 other *Fusarium* genomes, we were able to extract 2,572 *F. oxysporum mimps* and 116 *mimps* from other *Fusaria*. The relatively high abundance within the FOSC suggests that *mimps* originate from within this species complex. Six other *Fusarium* strains drew particular attention because relatively high numbers of *mimps* could be identified in their genomes. All of these strains were isolated from diseased flower bulbs.

Comparison of the nucleotide sequence of each *mimp* encountered in these strains showed that Fol3 *(F. proliferatum)*, Hy9 and Hy14 (both *F. hostae*) all have three to five elements in their genome that are 100% identical to *F. oxysporum mimps* (Fig. 3). This, combined with the fact that they may have an intact *impala* ORF (Fig. 2; assemblies are interrupted by N-gap or end of contig) and intact *SIX* gene homologs (other than *SIX2*) with high sequence similarity (Table 1) suggests that these isolates participated in horizontal chromosome transfer between *Fusarium* species.

In contrast to strains Fol3, Hy9 and Hy14 described above, the two *F. agapanthi* isolates (isolated from African lily in Australia and Italy) displayed a distinctly different *mimp* distribution. Most of the elements identified in these strains displayed long branch lengths and separate branching (Fig. 3), indicating a high level of sequence divergence from *F. oxysporum mimps*. *Fusarium* sp. Na10, isolated from *Narcissus* sp., only has four *mimps* in its genome assembly. Although this is a relatively high number for a strain belonging to the FFSC, all of the *mimp* sequences were different from Fo *mimps*.

Presence of *SIX* genes outside of the FOSC has so far been described in *Leptosphaeria maculans*, that has a distant homolog of *SIX1* (LmCys1, 26% amino acid identity (van de Wouw *et al.*, 2010; Martin and Kamoun, 2011)), *Colletotrichum orbiculare* and *C. higginsianum*, that have *SIX1* and *SIX6* homologs (Kleemann *et al.*, 2012; Gan *et al.*, 2013), *F. verticillioides*, that has a *SIX2* homolog (van der Does and Rep, 2007) and NRRL31046, an isolate of *F. foetens* that has a (partial) *SIX1* homolog (Laurence *et al.*, 2015). Interestingly, NRRL31046 was also isolated from a bulb flower, in this case from *Begonia* sp. displaying discoloration of veins in leaves and

stems (Schroers *et al.*, 2004). No genome sequence of this strain is currently available, and it would be interesting to further investigate the distribution of *SIX* genes amongst strains of *F. foetens* since this species also belongs to the FFSC.

We now find that *SIX2,* apart from being present in *F. verticillioides,* is widely distributed amongst species in the FFSC. The phylogeny of this gene is largely congruent with the species tree indicative of vertical inheritance (Fig. 6). This is not the case, however, for most other effector genes that were identified in the analysis including *SIX1, SIX6, SIX7* and *AVRFOM2* that seem to be largely restricted within the FOSC. The identification of *AVRFOM2* and a pseudogenized copy of *SIX1*, both 100% identical to sequences in *F. oxysporum* is suggestive of horizontal transfer to or from *F. proliferatum* Fol3. This hypothesis is strengthened by the sequences surrounding these genes (multiple kb in length) that are also 100% identical between the strains (Fig. 6). Additionally, the placement of *F. hostae* Hy9/Hy14 *SIX6* amongst the *F. oxysporum SIX6* homologs also points at a HT event of this gene, supported by the high bootstrap values of the *F. hostae* clade and high sequence similarity to Fo f. sp. *niveum* 015/019 *SIX6* (Fig. 4).

Horizontal chromosome transfer (HCT) within the *species complex* and across species boundaries is believed to have contributed to genetic diversity and the generation of new (pathogenic) variants (Ma *et al.*, 2013; Kang *et al.*, 2014). This has been experimentally shown so far only between pathogenic and non-pathogenic strains of *F. oxysporum* (Ma *et al.* 2010, Vlaardingerbroek *et al.* 2016, van Dam *et al.* accepted). In these studies, the pathogenicity chromosome of Fo f. sp. *lycopersici* (chr 14) or Fo f. sp. *radicis-cucumerinum* (chr^{RC}) was transferred into the genetic background of biocontrol strain Fo47. Fo47 belongs to a different vegetative compatibility group than either of these pathogens. The recipient strain had subsequently become pathogenic towards tomato or several cucurbit species, respectively. The proposed mechanism for HCT is nuclear fusion followed by selective loss of chromosomes from one of the fusion partners, but exactly how this happens remains elusive (Vlaardingerbroek *et al.*, 2016). Hyphal fusion of genetically dissimilar strains normally leads to a vegetative incompatibility response followed by programmed cell death (Glass and Dementhon, 2006).

HT of virulence genes has been suggested for Fo f. sp. *canariensis* (Laurence *et al.*, 2015) and also in strains of Fo that were isolated from natural ecosystems of Australia (Rocha *et al.*, 2015). Interspecies HT of the virulence gene pisatin demethylase *(PDA)* from *Nectria haematococca* (the teleomorph of *F. solani*) to Fo f. sp. *phaseoli* and *pisi* was suggested by discordance between the gene genealogy of *PDA* and the organismal phylogeny (Milani *et al.*, 2012). The presence of dispensable chromosomes in *N. haematococca* and *F. oxysporum* and the fact that these can move between strains in the FOSC led the authors of this study to suggest this as a potential pathway for horizontal transmission of chromosomes containing virulence genes between *Fusarium* species.

The alternative to HT, a shared origin of all *SIX* genes in the ancestor of the FOSC and FFSC

followed by selective loss in most of the FFSC species remains a possibility. Since *SIX2* is found in multiple FFSC species as well as Fo, a shared origin at least for this gene is probable. For *SIX6, SIX7, SIX11* and the Fo f. sp. *melonis* effector candidates, no homologs in other *Fusarium* species have been described until now (Schmidt *et al.*, 2013, 2016). The compartmentalization of pathogen genomes (including that of Fo) and the concentration of effector genes on one or a few accessory chromosomes facilitates the simultaneous loss of multiple effector genes. The rate of loss of such an ancestral accessory chromosome could be high due to avirulence effects of the genes encoded on it. Still, it is hard to explain the high identity of some *mimps* and effector genes (and their up- and downstream regions) as the placement of non-Fo effector genes within the Fo clade under the assumption of exclusive vertical inheritance.

In conclusion, we describe here for the first time data suggestive of horizontal gene transfer between different *Fusarium* species. HT of (part of) an accessory chromosome may have occurred under natural conditions, such as in flower bulb fields. Data supportive of this hypothesis include the presence of many *mimps* (sometimes identical to a Fo *mimp*), as well as the occurrence of *SIX* homologs (other than *SIX2)* that were not found in other *Fusarium* species outside of the FOSC. Whole genome sequencing with long read sequencing technologies of the *F. proliferatum* and *F. hostae* strains described here, Fo f. sp. *lilii*/*hyacinthi* as well as *F. foetens* NRRL31046 would be highly interesting in order to compare their genome architectures.

Materials and Methods

Identification of *mimps* and *impala* ORFs in each genome

Mimps were identified using a custom python script (available upon request) that searches for the terminal inverted repeats using the following 16-nucleotide regular expression: 'NNCAGT[GA][GA]G[GAT][TGC]GCAA[TAG]AA'. Stretches of sequence where an instance of this motif was found within 400 bp of another instance in reverse orientation were extracted. The list of newly identified *mimps* was then compared to each other through reciprocal BLASTN (evalue < 1e-5, percent identity > 80%, alignment length > 160 nt) and clusters were formed with single linkage. This resulted in 40 clusters, of which the clusters with 10 or less *mimp* instances were grouped together in category '15' (other). The sequences of *mimp* families 1-6 were extracted from Schmidt *et al.*(2013) and Bergemann *et al.*(2008) and compared to the clusters obtained from reciprocal BLAST to see which of the categories represented which family.

Detection of an intact impala ORF was performed by manual evaluation of TBLASTN output (*e*-value < 1*e*-100) using the full-length FOM24 *impala* transposase ORF (Genbank accession AF282722.1) as a query.

Multiple sequence alignments and phylogeny

For alignment of conserved genes, the sequences of *EF1α, RPB1* and *RPB2* were extracted from the genomes based on BLASTN searches, using the gene sequences of Fol4287 as query. ClustalO v1.2.1 (Sievers *et al.*, 2014) was used to make an alignment for each gene, after which the alignments were concatenated into a single alignment. This concatenated alignment was trimmed using TrimAl (with -strictplus) (Capella-Gutierrez *et al.*, 2009) and fed to PhyML v20120412 (with --bootstrap 100) (Guindon *et al.*, 2009) in order to retrieve a phylogeny of the *Fusarium* genomes. A similar approach was taken for identification and alignment of *SIX* and Fo f. sp. *melonis* effector candidate homologs, although these alignments were not trimmed.

MAFFT v6.903b (Katoh *et al.*, 2002) was used for alignment of *mimp* sequences and PhyML was applied for phylogenetic inference as described above. All trees were visualized in ETE3 v3.0.0b35 (Huerta-Cepas *et al.*, 2016).

Whole genome sequencing, *de novo* assembly

Fusarium genomic DNA was isolated through phenol-chloroform extraction from freeze-dried mycelium that was harvested from five-day old NO_3-medium (0.17% yeast nitrogen base, 3% sucrose, 100 mM KNO_3) cultures as described in detail in (van Dam *et al.*, 2016). Library preparation of insert size 550 bp and Illumina HiSeq 2500 paired-end sequencing was performed at Keygene N.V. (Wageningen, the Netherlands).

Sequencing reads were trimmed for quality and to remove adapter sequences with FastqMcf v1.04.676 (https://expressionanalysis.github.io/ea-utils/, quality threshold=20). *De novo* assemblies were generated using CLC-workbench 8.0. Default settings were used, except 'minimum contig length=500'.

Alignment of scaffolds

Nucmer of the MUMmer package v3.1 (with '-breaklen 100' to break alignments separated by N-gaps resulting from contig scaffolding) was used for visualization of scaffold alignments.

Data access

Whole-Genome Shotgun projects for the newly sequenced strains of *F. hostae* Hy9, Hy14, *F. proliferatum* Fol3 and *Fusarium* sp. Na10 have been deposited at Genbank under the BioProject PRJNA389502. Raw sequence data have been deposited into the Sequence Read Archive under the accession number SRP109077. All publically available genome sequences that were used were obtained from Genbank. Their NCBI accession numbers can be found in Supplementary Table S1.

Acknowledgments

We would like to thank Joop van Doorn and Suzanne Breeuwsma for kindly sharing the isolates from diseased flower bulbs with us. This work was supported by the Horizon programme (project 93512007) of the Netherlands Genomics Initiative (NGI) through a grant to M. Rep.

References

Baayen, R.P., Förch, M.G., Waalwijk, C., Bonants, P.J.M., Löffler, H.J.M., and Roebroeck, E.J.A. (1998) Pathogenic, genetic and molecular characterisation of *Fusarium oxysporum* f. sp. *lilii*. *Eur. J. Plant Pathol.* **104**: 887–894.

Bergemann, M., Lespinet, O., M'Barek, S. Ben, Daboussi, M.J., and Dufresne, M. (2008) Genome-wide analysis of the *Fusarium oxysporum mimp* family of MITEs and mobilization of both native and *de novo* created mimps. *J. Mol. Evol.* **67**: 631–642.

Breeuwsma, S.J. and De Boer, M. (2004) *Fusarium* in bloembolgewassen: detectiemethoden en vruchtwisselingsproblematiek PPO Bloembollen en Bomen (in Dutch), PPO project 320689.

Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

van Dam, P., Fokkens, L., Schmidt, S.M., Linmans, J.H.J., Kistler, H.C., Ma, L.-J., and Rep, M. (2016) Effector profiles distinguish formae speciales of Fusarium oxysporum. *Environ. Microbiol.* **18**: 4087–4102.

van der Does, H.C. and Rep, M. (2007) Virulence genes and the evolution of host specificity in plant-pathogenic fungi. *Mol. Plant. Microbe. Interact.* **20**: 1175–1182.

Dufresne, M., Hua-Van, A., Abd El Wahab, H., M'Barek, S. Ben, Vasnier, C., Teysset, L., *et al.* (2007) Transposition of a fungal miniature inverted-repeat transposable element through the action of a *Tc1*-like transposase. *Genetics* **175**: 441–452.

Edwards, J., Auer, D., de Alwis, S.-K., Summerell, B., Aoki, T., Proctor, R.H., *et al.* (2016) *Fusarium agapanthi* sp. nov., a novel bikaverin and fusarubin-producing leaf and stem spot pathogen of *Agapanthus praecox* (African lily) from Australia and Italy. *Mycologia* **108**: 981–992.

Feschotte, C. and Pritham, E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**: 331–368.

Feschotte, C., Zhang, X., and Wessler, S.R. (2002) Miniature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons. *Mob. DNA II* 1147–1158.

Firon, A., Villalba, F., Beffa, R., and D'Enfert, C. (2003) Identification of essential genes in the human fungal pathogen *Aspergillus fumigatus* by transposon mutagenesis. *Eukaryot. Cell* **2**: 247–55.

Gan, P., Ikeda, K., Irieda, H., Narusaka, M., O'Connell, R.J., Narusaka, Y., *et al.* (2013) Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytol.* **197**: 1236–1249.

Glass, N.L. and Dementhon, K. (2006) Non-self recognition and programmed cell death in filamentous fungi. *Curr. Opin. Microbiol.* **9**: 553–558.

Guindon, S., Dufayard, J.F., Hordijk, W., Lefort, V., and Gascuel, O. (2009) PhyML: fast and accurate phylogeny reconstruction by maximum likelihood. In, *Infection Genetics and Evolution*., pp. 384–385.

Hua-Van, A., Cacciola, S., and Daboussi, M.J. (2002) Transposition of *impala* as a potential tool for insertional mutagenesis in *Colletotrichum gloeosporioides*. In, *6th Eur Conf on Fungal Genetics*., p. 112.

Hua-Van, A., Langin, T., and Daboussi, M.-J. (2001) Evolutionary History of the *impala* Transposon in *Fusarium oxysporum*. *Mol. Biol. Evol.* **18**: 1959–1969.

Hua-Van, A., Pamphile, J.A., Langin, T., and Daboussi, M.J. (2001) Transposition of autonomous and engineered *impala* transposons in *Fusarium oxysporum* and a related species. *Mol. Gen. Genet.* **264**: 724–731.

Huerta-Cepas, J., Serra, F., and Bork, P. (2016) ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **33**: 1635–1638.

Kang, S., Demers, J., del Mar Jimenez-Gasco, M., and Rep, M. (2014) *Fusarium oxysporum*. In, *Genomics of Plant-Associated Fungi and Oomycetes: Dicot Pathogens*. Springer, pp. 99–119.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**: 3059–3066.

Kleemann, J., Rincon-Rivera, L.J., Takahara, H., Neumann, U., van Themaat, E.V.L., van der Does, H.C., *et al.* (2012) Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathog.* **8**: e1002643.

Laurence, M.H., Summerell, B.A., and Liew, E.C.Y. (2015) *Fusarium oxysporum* f. sp. *canariensis*: evidence for horizontal gene transfer of putative pathogenicity genes. *Plant Pathol.* **64**: 1068–1075.

Li Destri Nicosia, M.G., Brocard-Masson, C., Demais, S., Hua Van, A., Daboussi, M.-J., and Scazzocchio, C. (2004) Heterologous transposition in *Aspergillus nidulans*. *Mol. Microbiol.* **39**: 1330–1344.

Ma, L.-J., van der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.-J., Di Pietro, A., *et al.* (2010) Comparative analysis reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367–373.

Ma, L.-J., Geiser, D.M., Proctor, R.H., Rooney, A.P., O'Donnell, K., Trail, F., *et al.* (2013) *Fusarium* pathogenomics. *Annu. Rev. Microbiol.* **67**: 399–416.

Martin, F. and Kamoun, S. (2011) Effectors in plant-microbe interactions John Wiley & Sons.

Milani, N.A., Lawrence, D.P., Elizabeth Arnold, A., and Van Etten, H.D. (2012) Origin of pisatin demethylase (PDA) in the genus *Fusarium*. *Fungal Genet. Biol.* **49**: 933–942.

de Queiroz, M.V. and Daboussi, M.J. (2003) *Impala*, a transposon from *Fusarium oxysporum*, is active in the genome of *Penicillium griseoroseum*. *FEMS Microbiol. Lett.* **218**: 317–21.

Rocha, L.O., Laurence, M.H., Ludowici, V.A., Puno, V.I., Lim, C.C., Tesoriero, L.A., *et al.* (2015) Putative effector genes detected in *Fusarium oxysporum* from natural ecosystems of Australia. *Plant Pathol.* **65**: 914–929

Schmidt, S.M., Houterman, P.M., Schreiver, I., Ma, L., Amyotte, S., Chellappan, B., *et al.* (2013) MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genomics* **14**: 119.

Schmidt, S.M., Lukasiewicz, J., Farrer, R., van Dam, P., Bertoldo, C., and Rep, M. (2016) Comparative genomics of *Fusarium oxysporum* f. sp. *melonis* reveals the secreted protein recognized by the *Fom-2* resistance gene in melon. *New Phytol.* **209**: 307–318.

Schroers, H.-J., Baayen, R.P., Meffert, J., de Gruyter, J., Hooftman, M., and O'Donnell, K. (2004) *Fusarium foetens*, a new species pathogenic to *begonia elatior* hybrids (*Begonia* x *hiemalis*) and the sister taxon of the *Fusarium oxysporum* species complex. *Mycologia* **96**: 393–406.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., *et al.* (2014) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**: 539–539.

Spanu, F., Pasquali, M., Scherm, B., Balmas, V., Marcello, A., Ortu, G., *et al.* (2012) Transposition of the miniature inverted-repeat transposable element *mimp1* in the wheat pathogen *Fusarium culmorum*. *Mol. Plant Pathol.* **13**: 1149–1155.

Villalba, F., Lebrun, M.-H., Hua-Van, A., Daboussi, M.-J., and Grosjean-Cournoyer, M.-C. (2001) Transposon *impala*, a Novel Tool for Gene Tagging in the Rice Blast Fungus *Magnaporthe grisea*. *Mol. Plant-Microbe Interact.* **14**: 308–315.

Vlaardingerbroek, I., Beerens, B., Rose, L., Fokkens, L., Cornelissen, B.J.C., and Rep, M. (2016) Exchange of core chromosomes and horizontal transfer of lineage-specific chromosomes in *Fusarium oxysporum*. *Environ. Microbiol.***18:** 3702–3713.

van de Wouw, A.P., Cozijnsen, A.J., Hane, J.K., Brunner, P.C., McDonald, B.A., Oliver, R.P., and Howlett, B.J. (2010) Evolution of linked avirulence effectors in *Leptosphaeria maculans* is affected by genomic environment and exposure to resistance genes in host plants. *PLoS Pathog.* **6**: e1001180.

Supplementary data

Table S1: Number of *mimps* identified in each of the genome assemblies used in this study and the GenBank accession numbers of each assembly (accessed on 10-4-2017).

species	isolate	number of mimps	GenBank accession nr	remarks
F. agapanthi	NRRL31653	7	GCA_001654555.1	
F. agapanthi	NRRL54464	6	GCA_001654545.1	
F. asiaticum	NRRL28720	0	GCA_001717835.1	
F. asiaticum	NRRL6101	0	GCA_001717845.1	
F. avenaceum	Fa05001	1	GCA_000769215.1	
F. avenaceum	FaLH03	1	GCA_000769305.1	
F. avenaceum	FaLH27	1	GCA_000769295.1	
F. azukicola	NRRL54364	0	GCA_001680625.1	
F. brasiliense	NRRL31757	1	GCA_001680685.1	
F. circinatum	FSP34	0	GCA_000497325.2	
F. circinatum	GL1327	0	GCA_000876485.1	
F. cuneirostrum	NRRL31157	0	GCA_001680505.1	
F. fujikuroi	B14	0	GCA_000315255.1	
F. fujikuroi	CF-295141	0	GCA_001705295.1	
F. fujikuroi	FGSC8932	0	GCA_001023045.1	
F. fujikuroi	IMI58289	0	GCA_900079805.1	
F. fujikuroi	KSU3368	0	GCA_001023065.1	
F. fujikuroi	KSUX-10626	0	GCA_001023035.1	
F. graminearum	233423	0	GCA_000966635.1	
F. graminearum	241165	0	GCA_000966645.1	
F. graminearum	CS3005	0	GCA_000599445.1	
F. graminearum	DAOM180378	0	GCA_001717915.1	
F. graminearum	NRRL28336	0	GCA_001717905.1	
F. graminearum	PH-1	0	GCA_000240135.3	
F. hostae	Hy14	39	NJCQ01000000	
F. hostae	Hy9	40	NJCR01000000	
F. langsethiae	Fl201059	1	GCA_001292635.1	
F. mangiferae	MRC7560	0	GCA_900044065.1	
F. meridionale	NRRL28721	0	GCA_001717825.1	
F. meridionale	NRRL28723	0	GCA_001717855.1	
F. nygamai	MRC8546	2	GCA_001262555.1	
F. oxysporum	Fo47	6	GCA_000271705.2	non-pathogenic, biocontrol
F. oxysporum	FOSC-3a	6	GCA_000350365.1	clinical isolate
F. oxysporum	MN14	6	MALU01000000	non-pathogenic
F. oxysporum	Tu58	5	NJBW01000000	non-pathogenic
F. oxysporum f. sp. *conglutinans*	5176	52	GCA_000222805.1	
F. oxysporum f. sp. *conglutinans*	PHW808	44	GCA_000260215.2	
F. oxysporum f. sp. *cubense*	B2	2	GCA_000350365.1	
F. oxysporum f. sp. *cubense*	II5	12	GCA_000260195.2	
F. oxysporum f. sp. *cubense*	N2	0	GCA_000350345.1	
F. oxysporum f. sp. *cucumerinum*	001	35	MAKZ01000000	
F. oxysporum f. sp. *cucumerinum*	011	39	MABT01000000	
F. oxysporum f. sp. *cucumerinum*	013	41	MABJ01000000	
F. oxysporum f. sp. *cucumerinum*	015	34	MABK01000000	
F. oxysporum f. sp. *cucumerinum*	018	38	MABM01000000	
F. oxysporum f. sp. *cucumerinum*	021	36	MABL01000000	
F. oxysporum f. sp. *cucumerinum*	030	38	MABN01000000	
F. oxysporum f. sp. *cucumerinum*	035	47	MABO01000000	
F. oxysporum f. sp. *cucumerinum*	037	43	MABP01000000	
F. oxysporum f. sp. *gladioli*	G14	37	NJCM01000000	
F. oxysporum f. sp. *gladioli*	G2	44	NJCL01000000	
F. oxysporum f. sp. *gladioli*	G76	25	NJCK01000000	
F. oxysporum f. sp. *lagenariae*	001	31	NJCJ01000000	
F. oxysporum f. sp. *lagenariae*	002	31	NJCI01000000	
F. oxysporum f. sp. *lagenariae*	004	21	NJCH01000000	
F. oxysporum f. sp. *lagenariae*	005	31	NJCG01000000	

Table S1 (continued)

species	isolate	number of mimps	GenBank accession nr	remarks
F. oxysporum f. sp. *lilii*	Fol39	29	NJCF01000000	
F. oxysporum f. sp. *luffae*	001	28	NJCE01000000	
F. oxysporum f. sp. *luffae*	002	27	NJCD01000000	
F. oxysporum f. sp. *lycopersici*	002	35	MAMG01000000	
F. oxysporum f. sp. *lycopersici*	004	33	MALH01000000	
F. oxysporum f. sp. *lycopersici*	007	37	MALI01000000	
F. oxysporum f. sp. *lycopersici*	014	33	MALJ01000000	
F. oxysporum f. sp. *lycopersici*	016	30	MALM01000000	
F. oxysporum f. sp. *lycopersici*	018	27	MALL01000000	
F. oxysporum f. sp. *lycopersici*	026	33	MALK01000000	
F. oxysporum f. sp. *lycopersici*	029	34	MALN01000000	
F. oxysporum f. sp. *lycopersici*	038	36	MALO01000000	
F. oxysporum f. sp. *lycopersici*	069	32	MALP01000000	
F. oxysporum f. sp. *lycopersici*	072	30	MALQ01000000	
F. oxysporum f. sp. *lycopersici*	073	30	MALR01000000	
F. oxysporum f. sp. *lycopersici*	074	36	MALS01000000	
F. oxysporum f. sp. *lycopersici*	075	33	MALT01000000	
F. oxysporum f. sp. *lycopersici*	4287	49	GCA_000149955.2	
F. oxysporum f. sp. *melongenae*	001	31	NJCC01000000	
F. oxysporum f. sp. *melonis*	001	50	GCA_000260495.2	
F. oxysporum f. sp. *melonis*	001_pacbio	80	NJCY01000000	sequenced with PacBio
F. oxysporum f. sp. *melonis*	004	41	MALX01000000	
F. oxysporum f. sp. *melonis*	005	42	MALY01000000	
F. oxysporum f. sp. *melonis*	006	39	MALZ01000000	
F. oxysporum f. sp. *melonis*	009	50	MAMA01000000	
F. oxysporum f. sp. *melonis*	010	40	MAMB01000000	
F. oxysporum f. sp. *melonis*	011	35	MAMC01000000	
F. oxysporum f. sp. *melonis*	012	33	MAMD01000000	
F. oxysporum f. sp. *melonis*	013	41	MAME01000000	
F. oxysporum f. sp. *melonis*	016	42	MAMF01000000	
F. oxysporum f. sp. *momordicae*	001	41	NJCB01000000	
F. oxysporum f. sp. *momordicae*	004	38	NJCA01000000	
F. oxysporum f. sp. *narcisi*	Na5	40	NJCV01000000	
F. oxysporum f. sp. *nicotianae*	001	7	NJBZ01000000	
F. oxysporum f. sp. *nicotianae*	003	6	NJBY01000000	
F. oxysporum f. sp. *nicotianae*	010	6	NJBX01000000	
F. oxysporum f. sp. *nicotianae*	012	13	NJCU01000000	
F. oxysporum f. sp. *niveum*	002	34	MALA01000000	
F. oxysporum f. sp. *niveum*	005	33	MAKY01000000	
F. oxysporum f. sp. *niveum*	010	37	MALB01000000	
F. oxysporum f. sp. *niveum*	013	37	MALC01000000	
F. oxysporum f. sp. *niveum*	015	26	MALD01000000	
F. oxysporum f. sp. *niveum*	019	24	MAMH01000000	
F. oxysporum f. sp. *niveum*	020	48	MALE01000000	
F. oxysporum f. sp. *niveum*	021	28	MALG01000000	
F. oxysporum f. sp. *niveum*	037	32	MALF01000000	
F. oxysporum f. sp. *physali*	KOD886	35	NJBV01000000	
F. oxysporum f. sp. *physali*	KOD887	32	NJBU01000000	
F. oxysporum f. sp. *physali*	KOD888	38	NJBT01000000	
F. oxysporum f. sp. *pisi*	HDV247	28	GCA_000260075.2	
F. oxysporum f. sp. *rad.-cumerinum*	016	23	MABQ01000000	
F. oxysporum f. sp. *rad.-cumerinum*	016_pacbio	38	MABQ02000000	sequenced with PacBio
F. oxysporum f. sp. *rad.-cumerinum*	024	29	MABR01000000	
F. oxysporum f. sp. *rad.-cumerinum*	031	25	MABS01000000	
F. oxysporum f. sp. *radicis-lycopersici*	CL57	11	GCA_000260155.3	
F. oxysporum f. sp. *raphani*	PHW815	74	GCA_000260235.2	
F. oxysporum f. sp. *tulipae*	Tu67	28	NJBS01000000	
F. oxysporum f. sp. *vasinfectum*	NRRL25433	59	GCA_000260175.2	
F. phaseoli	NRRL31156	0	GCA_001680515.1	
F. poae	2516	0	GCA_001675295.1	sequenced with PacBio

Table S1 (continued)

species	isolate	number of mimps	GenBank accession nr	remarks
F. proliferatum	ET1	1	GCA_900067095.1	
F. proliferatum	Fol3	10	NJCT01000000	
F. proliferatum	NRRL62905	0	GCA_900029915.1	
F. pseudograminearum	CS3096	0	GCA_000303195.2	
F. pseudograminearum	CS3270	0	GCA_000974265.2	
F. pseudograminearum	RBG5266	0	GCA_001703955.1	
F. sambucinum	F-4	0	GCA_001567575.1	
F. temperatum	CMWF389	0	GCA_001513835.1	
F. tucumaniae	NRRL31096	0	GCA_001680535.1	
F. tucumaniae	NRRL31781	0	GCA_001680595.1	
F. tucumaniae	NRRL34546	0	GCA_001680725.1	
F. verticillioides	7600	1	GCA_000149555.1	
F. virguliforme	Clinton-1B	0	GCA_001680665.1	
F. virguliforme	LL0009	0	GCA_001680605.1	
F. virguliforme	NRRL34551	0	GCA_001680525.1	
Fusarium sp.	FSSC6	1	GCA_001633045.1	
Fusarium sp.	JS1030	0	GCA_000966855.1	
Fusarium sp.	JS626	0	GCA_000966865.1	
Fusarium sp.	Na10	4	NJCS01000000	

Chapter 6

General discussion

General discussion

The presumed asexual, soil-borne fungus *Fusarium oxysporum* (Fo) causes vascular wilt disease in a large number of plants species, many of which are economically important crops. Though the host range of the species complex as a whole encompasses more than 120 different hosts, each *forma specialis* (f. sp.) is typically restricted to only one or a few plants (Michielse and Rep, 2009). Earlier research in our lab has focused on the group of isolates capable of infecting tomato: Fo f. sp. *lycopersici* (Fol).

At the start of this project, it was known that Fol secretes small proteins into the xylem sap of tomato plants that increase the virulence of the fungus. These Secreted in xylem (Six) effector proteins are encoded by genes that are concentrated on a single chromosome, the second smallest chromosome of strain Fol4287 (chr 14) (Ma *et al.*, 2010). Over time, 14 such Six proteins were identified in the xylem sap of infected tomato plants (Houterman *et al.*, 2007; Schmidt *et al.*, 2013). Several of these were shown to be essential for complete virulence on tomato plants, like Six1 (Rep *et al.*, 2004) and Avr2 (Six3) (Houterman *et al.*, 2009). Furthermore, it was discovered that a similar set of *SIX* genes is present in all Fol isolates (Lievens *et al.*, 2009; van Dam *et al.*, 2016). An exception to this rule is formed by *SIX* genes whose protein products are recognized by Resistance (R) genes in the plant: *AVR1 (SIX4)* has been lost in race 2 and 3 isolates and *AVR2 (SIX3)* has mutated in race 3 isolates to avoid recognition by the resistance protein I-2 (Takken and Rep, 2010).

The first twelve whole genome sequences generated by the Broad Institute of MIT and Harvard allowed the identification of *SIX* gene homologs in other *formae speciales* (ff. spp.) of Fo. It was deemed likely that each plant pathogenic strain of Fo possesses a set of effector genes that allow it to overcome the host's innate immune responses, but no comparative studies involving multiple genomes per f. sp. had been performed. We set out to generate whole genome sequences of many isolates of a selected set of *formae speciales* affecting a group of related plant species: the cucurbits. The primary goal of the project was to identify novel virulence genes in these strains and compare them with each other. In this way, we hoped to learn more about the genes necessary for the development of *Fusarium* wilt disease and root rot in cucurbits. Moreover, evaluating the phylogeny of these genes could potentially shed light on the evolution of host-specific pathogenicity in Fo in general. Both points will be further discussed in this chapter.

Which genetic factors define host specificity?

Exactly why a specific set of effectors, secreted enzymes and secondary metabolites allows pathogenic colonization of one plant species but not of the other is not quite clear. Possibly, there are key differences in the (importance of) molecular targets between plant species. If this were the case, incidences of cross-pathogenicity would occur more frequently between

formae speciales infecting genetically similar hosts. Incidental cross-pathogenicity between the ff. spp. *cucumerinum, melonis* and *niveum* has indeed been described (Mcmillan, 1986; Cafri *et al.*, 2005; Zhou and Everts, 2007), and is further explored in Chapter 2. Strains of these ff. spp. were tested for cross-pathogenicity towards the three respective plant species, cucumber, muskmelon and watermelon. Only mild cross-pathogenicity could be observed, even though these plant species – particularly cucumber and melon – are highly similar both morphologically and genetically. The mean nucleotide sequence similarity between the genomes of cucumber and melon over coding regions is 95% (Huang *et al.*, 2009). No cross-pathogenicity occurred from any of the cucurbit-infecting ff. spp. to tomato, or from Fol to any of the cucurbits, supporting the hypothesis that molecules targeted by the pathogen are too much diverged between these plant species.

Intriguingly, Fo could occasionally be recovered from stem pieces of inoculated non-host plants when placed on solid medium in most tested f. sp. - plant combinations. Fo f.sp. *cucumerinum* and *melonis* could even – albeit infrequently - be recovered from infected tomato plants, a species that belongs to an entirely different order than the natural hosts of both of these ff. spp. Experiments involving cross-pathogenicity testing are not frequently performed. For Fo47, a bioprotective strain that is known to prevent disease development caused by Fol, it was shown that colonization of tomato plants is generally limited to the root epidermis (Olivain and Alabouvette, 1997; Olivain *et al.*, 2006). This strain, like other non-specialized strains, has no *SIX* gene homologs and very few predicted effectors. The capability of growing into plant roots beyond the epidermis can therefore be considered to not be present in all Fo strains. Events of mild cross-pathogenicity or cross-colonization like the ones described above may be caused by (incidental) overcoming of the defense responses of the non-host plant.

A potential explanation for this could be that initial access to the vasculature is facilitated by a first wave of (general) effector proteins and enzymes. Successful subsequential colonization of the plant through the vessels, however, may need a second wave of (host-specialized) effectors that counteract pamp-triggered and effector-triggered immune responses (PTI and ETI, respectively). In line with this, Williams *et al.* (2016) reported differential expression onset (varying from 1 – 7 days post inoculation) of candidate effectors in Fo f. sp. *medicaginis*.

Occurrence of effectors in the Fo species complex

Comparison of the effector genes present in the genomes of strains belonging to the ff. spp. *cucumerinum, niveum, melonis* and *radicis-cucumerinum* showed that strains of each f. sp. possess similar sets of effector genes (Chapter 2). Moreover, the effector suites of cucurbit-infecting strains as a whole were found to be more similar to each other than to strains pathogenic to other plant species like tomato, banana or *Brassicaceae* spp. This is an interesting observation, since the investigated strains belong to many different clonal lines throughout

the three main phylogenetic clades of Fo.
Each pathogenic strain of Fo needs to 'conquer' the plant vasculature in a presumably similar way, a process that would require a similar genetic toolkit. The set of effectors encoded in the genome, however, varies widely from strain to strain, mostly depending on which plant can be infected. Although several virulence-associated genes are present in all plant-pathogenic Fo strains, effector genes typically occur in a 'patchy' distribution. What works in one plant species apparently does not work for another. This irregular distribution is not restricted to the *F. oxysporum* species complex (FOSC), but occurs in many fungal pathogens (van der Does and Rep, 2007).
I see two likely reasons for the patchy distribution. Firstly, effectors may at some point be recognized directly or indirectly by the plant as a result from an arms race, thus resulting in limiting of colonization (i.e. as an avirulence factor) (Jones and Dangl, 2006). This is an incentive to lose this gene, or if this is not an option, mutation in order to escape recognition whilst retaining functionality. Similarly, changes in the molecular targets of effectors in the host plant could be selected by disease pressure. In response, the pathogen needs to keep its virulence factors up-to-date through mutation to retain the capability of virulence on the host (van der Does and Rep, 2007).
Secondly, horizontal gene transfer (HGT) and horizontal chromosome transfer (HCT) are known to occur within and between *Fusarium* species (Ma *et al.*, 2010; Milani *et al.*, 2012; Vlaardingerbroek, Beerens, Rose, *et al.*, 2016). Because virulence genes are physically clustered together, sets of 'successful' effectors for a given host can be passed from strain to strain, even in predominantly or exclusively asexual species. Arguments supportive of this hypothesis are numerous and include identical *SIX* gene sequences between strains of the same f. sp. (Chapter 2), stretches of hundreds of kb of chromosomal sequence with 95-100% pairwise similarity between Fo f. sp. *radicis-cucumerinum* and *melonis* (Chapter 3) and the occurrence of identical *miniature impalas* (*mimps*) as well as effector genes between strains of *F. proliferatum* / *F. hostae* and Fo (Chapter 5).

Developing diagnostic markers

The polyphyletic nature of many *formae speciales* of *F. oxysporum* obstructs easy 'subspecies' diagnostics of newly encountered strains. Reliable molecular detection methods that could replace labour- and time-intensive disease assays are therefore highly desired. The very limited sequence diversity of effector genes between members of the same *forma specialis*, plus the fact that they are functional elements in the pathogen-host interaction makes them prime targets for diagnostic markers (as demonstrated in Chapter 4). As described in the previous paragraph, there is a high selection pressure on effector genes resulting in accelerated evolution. It is not likely that this will negatively influence the reliability of using them for marker development though, as this type of diversification is likely to occur on much longer

time frames. It is probable that this elevated selection pressure has predominantly resulted in a high level of sequence difference between *formae speciales* allowing their molecular differentiation.

However, up to the present the most prevalent method of f. sp. marker development in Fo has been based on Random Amplified Polymorphic DNA (RAPD) fragments, resulting in Sequence Characterized Amplified Regions (SCARs). By definition, unique SCAR marker sequences are recovered from genomic regions other than the core. In fact, many of these markers were later found to target (parts of) transposable elements (TEs) (Lievens *et al.*, 2008). These mobile sequences are strongly enriched on pathogenicity chromosomes of Fo (Schmidt *et al.*, 2013; Chapter 3) and are therefore likely to co-migrate in the event of HCT resulting in the transmission of host-specificity into another Fo lineage. Diagnostics based on mobile elements that can actively move through the genome is likely less reliable than using virulence-related sequences. Indeed, several of the previously described RAPD sequences later proved to be aspecific when they were compared to other Fo genomes with BLAST (personal observations).

An alternative strategy that is becoming possible with the accumulating number of available genome sequences is the alignment of whole genomes followed by selection of regions that are specific to the target group of isolates. This approach was described for Fo f. sp. *conglutinans* (Focon), where the genome sequence of Focon strain Foc2 was compared to the other 11 Broad assemblies (Ling *et al.*, 2016). This resulted in a total of 355 candidate Foc2-specific genome fragments with lengths larger than 1 kb (accounting for 817 kb). One of these regions turned out to be present in 23 tested Focon strains but not in any of the other 19 tested ff. spp., providing a region suitable for PCR-based marker design. We also attempted this method with little success in the *Cucurbitaceae*-infecting ff. spp. Since the accessory genetic material of these ff. spp. probably has a shared evolutionary origin (shown by the effector profiles in Chapter 2 and marker profiles in Chapter 4), there was simply too much overlap between these sequences. Therefore, 'unbiased' comparison of DNA sequences to find markers seems to be a viable method to discriminate ff. spp. infecting more distantly related plant species, but for differential identification of ff. spp. affecting similar plants, virulence-associated genes probably provide the most solid basis. Also discrimination of physiological races, that can be determined by the presence/absence of a single gene or even by Single Nucleotide Polymorphisms (SNPs) within a virulence gene (Takken and Rep, 2010), is more easily accomplished through such markers. This was recently demonstrated for the races of Fol by Ayukawa and colleagues (Ayukawa *et al.*, 2017).

Presence of effector genes is not always an accurate predictor of pathogenicity

Although the data presented in Chapter 2 show that the effector suite present in the genome of plant-pathogenic Fo strains is associated with the strain's host range, solely having effector

genes is not sufficient for virulence. This was illustrated in a recent study by Jelinski *et al* (2017), where 74 strains isolated from Fo-infested tomato fields were genotyped for the presence of *SIX1-7* (Jelinski *et al.*, 2017). 71 (96%) of the recovered strains possessed either all race 3-associated *SIX* genes (1, 2, 3, 5, 6, 7) or no *SIX* genes at all. Two strains were found that had only *SIX2* and *3*, while a third strain additionally had *SIX6*. These three strains, as well as the *SIX*-less strains were not pathogenic towards tomato plants. Interestingly, though most of the SIX_{123567} strains were virulent, two strains were not, showing that the presence of effector genes alone is not sufficient for pathogenicity. In our own experiments, we also identified a strain of Fo f.sp. *melonis* (Fomln002, Chapter 4) that has a marker profile identical to other *melonis* strains, but when tested on susceptible muskmelon plants did not induce symptom development. The association of a select set of effector genes with host specific pathogenicity is, however, not contradicted by these observations. A case which is the other way around – a plant-pathogenic strain lacking known effectors – would be much more worrying and would indeed contradict the hypothesis presented in Chapter 2. So far, however, such a strain has not been identified.

Multiple factors besides effector genes are known to influence plant pathogenicity. Mutation or inhibition of transcription factors (TFs) orchestrating the expression levels of effector genes (van der Does *et al.*, 2016), chromosomal rearrangements that are relatively common in Fo (Vlaardingerbroek, Beerens, Rose, *et al.*, 2016; Vlaardingerbroek, Beerens, Schmidt, *et al.*, 2016) and epigenetic organization of the genome (*i.e.* condensed DNA is less accessible and as a consequence transcription levels in these regions are lower) may all negatively affect the virulence of a strain. Moreover, genes that were shown to be involved in virulence but also in other processes (hyphal growth, sporulation, etc.) might be affected or mutated, with a similar effect on pathogenicity.

Spontaneous loss or reduction of pathogenicity is indeed frequently observed in the lab when Fo is transferred from plate to plate for some time (personal observations). One explanation for this could be that a pathogenicity chromosome is not required for vegetative growth and is subject to chromosome rearrangement and loss events (the latter is estimated to occur spontaneously once in every 35,000 spores *in vitro* (Vlaardingerbroek, Beerens, Schmidt, *et al.*, 2016)). Large-scale deletions were found to be more frequent in populations that were proliferated from plate to plate compared to strains that were passed through host plants in experimental evolution experiments (Christina Lopez-Diaz and Dilay Ayhan, unpublished data).

The influence of horizontal chromosome transfer on genome evolution in the genus *Fusarium*

Though pathogenicity chromosomes may pose costs to the fungus while growing vegetatively, the benefits of having them when an opportunity for plant infection arises are large. HCT

allows rapid adaptation by providing the necessary 'keys' (effectors) to surpass a novel host's defenses, similar to the spread of antibiotic resistance through horizontal transfer of plasmids in prokaryotic pathogens.

HCT (i.e. non-meiotic transfer of genetic material) has been described for Fol4287 chromosome 14 and the smallest chromosome of Fol007 (Ma *et al.*, 2010). The recent finding that HCT is not restricted to TE-rich accessory chromosomes but that Fol core chromosomes 7 and 8 can also (partially) be exchanged between otherwise incompatible strains (Vlaardingerbroek, Beerens, Rose, *et al.*, 2016) is an indication that size and repeat content are not direct prerequisites for transfer. The suggested mechanism by which the process occurs is nuclear fusion in a (unstable) heterokaryon followed by selective loss of chromosomes from one of the fusion partners, but exactly how this happens remains elusive (Vlaardingerbroek, Beerens, Rose, *et al.*, 2016). Hyphal fusion of genetically dissimilar strains normally leads to programmed cell death as a result from vegetative incompatibility responses (Glass and Dementhon, 2006).

HCT is believed to have contributed to genetic diversity and the generation of new pathogenic variants (Ma *et al.*, 2013; Kang *et al.*, 2014). The results described in Chapter 3 are the first report of HCT within the FOSC using another *forma specialis* than Fol as the chromosome donor (in this case, *radicis-cucumerinum*). HGT between *Fusarium* species has previously been suggested between *Nectria haematococca* (the teleomorph of *F. solani*) and Fo f. sp. *phaseoli* due to discordance between the gene genealogy of the virulence gene pisatin demethylase (*PDA*) and the organismal phylogeny (Milani *et al.*, 2012). Additionally, the spotty distribution of the *Fot1* transposon and the high degree of similarity (up to 98%) of this element between strains of Fo and *N. haematococca* is also suggestive of horizontal transfer between these species (Daboussi *et al.*, 2002). Similar evidence supportive of interspecies HCT (Fo to/from *F. proliferatum* and *F. hostae*) is presented in Chapter 5. These observations are indicative of the potentially large influence horizontal transfer has on genome evolution and particularly on (host specific) pathogenicity in the *Fusarium* genus and the FOSC in particular.

Nothing is known about the level of host specificity and whether a concept like *forma specialis* exists in *F. hostae* (Fh) because only one plant host has been described so far (*Hyacinthus* sp.) (Baayen *et al.*, 2001). In contrast, *F. proliferatum* (Fp) colonizes a wide variety of hosts, ranging from maize to banana, pine trees and asparagus (Jurado *et al.*, 2010; Stępień *et al.*, 2011). This species is known to produce a number of toxins, such as fumonisins, moniliformin, beauvericin, fusaric acid and fusaroproliferin (Jurado *et al.*, 2010). In contrast to Fo, both in Fh and Fp sexual forms (teleomorphs) have been described: *Gibberella hostae* and *G. intermedia*, respectively (O'Donnell *et al.*, 1998; Geiser *et al.*, 2001).

It would be very interesting to have higher quality genome assemblies generated from long-read sequencing technologies (e.g. PacBio) of *F. proliferatum* Fol3 and *F. hostae* Hy9/Hy14 as they may provide further proof of HCT between these species and Fo. These assemblies would allow the discerning of larger syntenic blocks between strains and thereby the reconstruction

of accessory genome evolution at the level of chromosomes (as applied in Chapter 3).
HCT has never been shown experimentally between *Fusarium* species. Cocultivation experiments of the described strains with strains of Fo could provide definitive evidence that interspecies HCT is possible. If such an experiment succeeded, this would potentially shed light on the mechanisms involved on a cellular level. A next step to further map the accessory material in common between Fh/Fp (and potentially other species) and Fo would be the comparison of these genomes with assemblies of Fo infecting flower bulbs (e.g. f.sp. *hyacinthi / lilii / gladioli / tulipae*) and closely related *Fusarium* species from the *F. fujikuroi* species complex (e.g. *F. foetens* NRRL31046 from *Begonia* sp. that also possesses a *SIX1* homolog (Laurence *et al.*, 2015)).

The role of TEs in the genome of *F. oxysporum*

The ease with which pathogenicity can be transferred between Fo strains is remarkable: strains of Fo47 that received the Forc016 chr^{RC} gained the full capacity of causing root and shoot rot disease in multiple cucurbit species (Chapter 3). In order for horizontal transfer of pathogenicity to be efficient, virulence genes must be clustered together. The presence of many TEs in the accessory chromosomes is interesting in this respect since they may in fact facilitate the clustered arrangement of these genes. Repetitive elements have been associated with genome rearrangements, gene duplication and gene capture events (Thon *et al.*, 2006; Feschotte and Pritham, 2007; Chellapan *et al.*, 2016). Additionally, they have been associated with rapid sequence diversification of effectors in *Leptosphaeria maculans* (Rouxel *et al.*, 2011).
BLAST queries performed with a variety of TE families revealed that many sequences within these families are almost identical within the Fo genome, implying that they are young and many are probably still active (Clutterbuck, 2011). A recent study by Chellapan *et al.* (2016) showed that *Helitrons* are still active and that one subgroup, FoHeli1, is also present in the core genome. This raises the question how the division into a core and accessory subgenome is maintained and the spread of TEs to the 'housekeeping' division of the genome is largely prevented. Possibly, (re)insertion of active transposable elements occurs preferentially in the vicinity of other repetitive elements on the accessory regions, which maintains the structure. Since there is a clear advantage for preserving effector genes in clusters in a fungus that can horizontally transfer whole chromosomes, accumulation of TEs that facilitate the organization of these genes will occur through natural selection.
An exceptional genome in this regard is that of a *Brassicaceae* infecting strain that we sequenced with PacBio: Fo5176. We identified a remarkably high number of transposons and repeats in its core genome (Fig. 1). This strain has retained its pathogenicity towards *Arabidopsis,* but what the influence of such a high TE content on the rest of the genome is remains to be seen. Subnuclear chromosome organization and the epigenetic landscape may

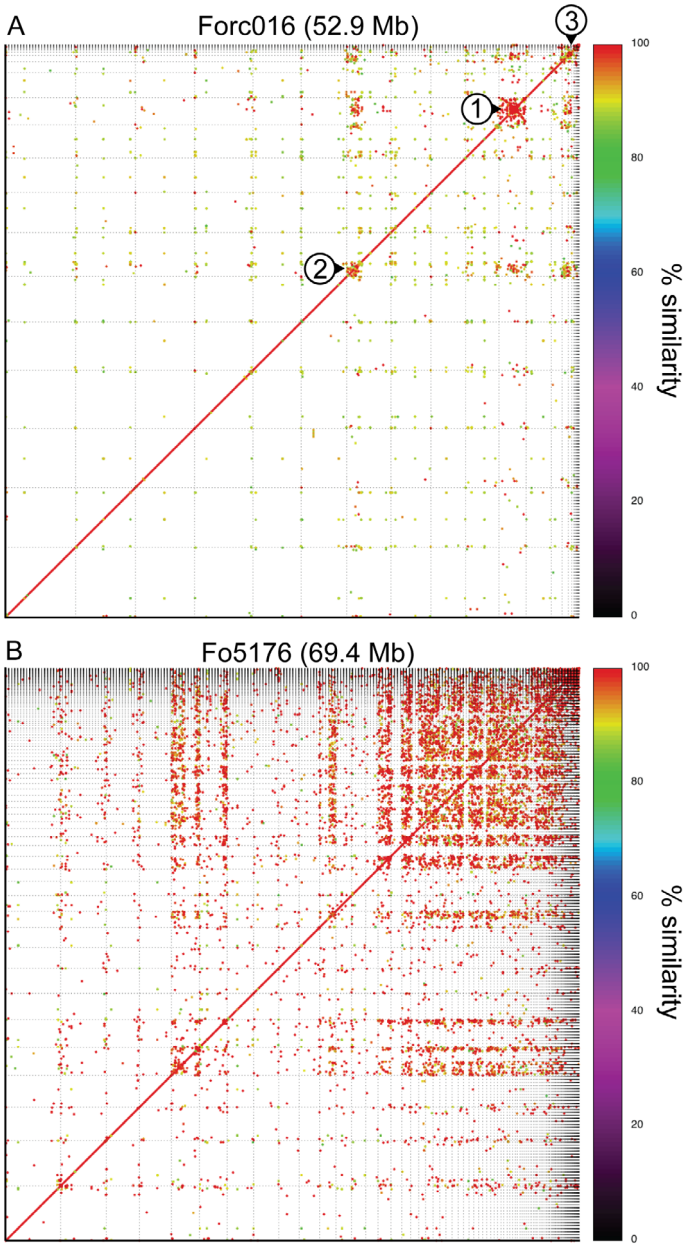


Fig. 1: Self-NUCmer alignments of PacBio assemblies of (A) Forc016 and (B) Fo5176 shows a large expansion in the number of repeats in the genome of Fo5176.
In Forc016, three predominant regions in the assembly contain TEs and other repetitive elements: (**1**) chr^{RC}, (**2**) a large part of one of the smaller (non-pathogenicity) accessory chromosomes attached to chr11 in this assembly and (**3**) the remaining unplaced contigs. In Fo5176, on the other hand, about a third of the genome is highly repetitive, while many alignments with high percentage of similarity also occur throughout core chromosomal regions. NUCmer v3.1 was run with default settings using the PacBio assembly of each genome as both the reference and the query.

play an important role in transposon activation and reinsertion preference. Investigation of epigenetic alterations of pathogenicity chromosomes, for example during plant colonization or other stressful growth conditions that may induce the 'opening' of the DNA structure (and allow TEs to be active) is recommendable.
For the TEs *mFot5, Helitron* and *mimp* a contextual association with effector genes has been described (Schmidt *et al.*, 2013; Chellapan *et al.*, 2016). *Mimps* do not seem to have a decisive function in influencing proximal gene expression, as was shown through knockout of a *mimp* element in the shared promoter region of *SIX3* and *SIX5* (Schmidt *et al.*, 2013). Rather they appear to be remnants of transposon activity and possibly gene capture. Virulence-related genes that are located in regions with higher mutation rates (like TE-rich regions) will accumulate more mutations, some of which may be beneficial non-synonymous changes that confer a fitness advantage. This produces a selection pressure that indirectly selects for the association of effector genes with TEs (Möller and Stukenbrock, 2017).
Karyotype instability, most often seen as rearrangements, duplications and deletions in the accessory chromosomes of Fo (Vlaardingerbroek, Beerens, Rose, *et al.*, 2016; Vlaardingerbroek, Beerens, Schmidt, *et al.*, 2016; Chapter 3), may also be induced by the abundance of TEs and can in this context be seen as positive (a feature) rather than negative (a bug) for the fungus. It allows rapid adaptation and generates variation that may be advantageous for an presumed asexual fungus like Fo, where mechanisms such as cross-over during meiosis apparently do not occur. Repeat-Induced Point mutation (RIP), a process that has been postulated to regulate TE activity, is probably absent in Fo (Clutterbuck, 2011). Further exploration of the effects of the presence of a pathogenicity chromosome in the genome of sexual *Fusarium* species like *F. hostae* and *F. proliferatum* would be very interesting in these respects.

The extended host range of Fo f. sp. *radicis-cucumerinum*

The chromosome transfer and loss experiments presented in Chapter 3 confirmed that chr^{RC} is the functional equivalent of Fol4287 chromosome 14; a pathogenicity 'module' that allows the receiving strain to be pathogenic to a new host. Contrary to Fol, only a single clonal line of Fo f. sp. *radicis-cucumerinum* (Forc) has been described to date, making the lab-generated Fo47 + chr^{RC} the second clonal lineage of this *forma specialis*. Ma *et al.* (2010) found a higher level of aggressiveness when another accessory Fol chromosome co-migrated with chromosome 14 to the Fo47 recipient, potentially due to the influence of transcription factors located on that chromosome (Ma *et al.*, 2010; van der Does *et al.*, 2016). In contrast, no difference in virulence between the HCT-derived strains and the Forc016 parent strain could be observed. Interestingly, two copies of *Fusarium Transcription Factor 1* (FTF1), associated with effector gene expression (Niño-Sánchez *et al.*, 2016; van der Does *et al.*, 2016), are present on chr^{RC} itself, which potentially indicates a level of transcriptional autonomy.
In the same chapter we describe that Six6 is an important virulence factor for Forc, particularly

for pathogenicity on cucumber. *SIX6* is not only present in the genomes of this *forma specialis*, but also in (among others) *niveum, cucumerinum, melonis, momordicae* and *luffae*. Interestingly, the 700 kb region in which *SIX6* is located on chr^{RC} is hyperdynamic, TE-rich and also contains almost all predicted effector genes in the genome of Forc016. Alignment of chr^{RC} with contig 127 in Fo f. sp. *melonis* (Fom) 001 showed that the chromosome is largely syntenic (with a pairwise nucleotide similarity close to 100%) to this contig, with the exception of the previously mentioned subregion. Several large-scale (estimated 20-80 kb) rearrangements have taken place here, and a ~300 kb region is present in chr^{RC} but absent in Fom001 contig 127. The *SIX6* open reading frame and the surrounding sequence is identical between the two. Forc is pathogenic towards several cucurbit species whereas Fom only infects melon plants.

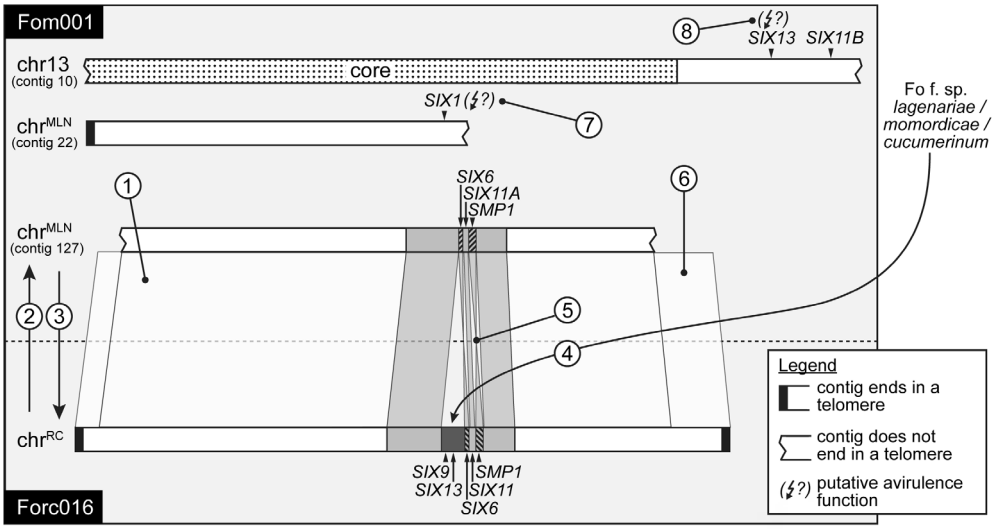


Fig. 2: Possible explanations for how the difference in phenotype and host range between *radicis-cucumerinum* and *melonis* evolved

(1) The pathogenicity chromosome of Forc016 (chr^{RC}) is largely syntenic with high pairwise sequence similarity to Fom001 contig 127, one of the latter strain's pathogenicity chromosomes (chr^{MLN}). The following scenarios are considered: **(2)** chr^{MLN} evolved from an ancestral chr^{RC} through a large deletion or rearrangement that caused *SIX9* to be lost in *melonis* and *SIX13* to be moved to another location within the Fom001 genome; or **(3)** chr^{RC} evolved from chr^{MLN} through **(4)** a large scale (~300 kb) insertion from an unknown source, possibly Fo f. sp. *cucumerinum, lagenariae* or *momordicae*. After this process, numerous rearrangement events occurred **(5)** resulting in multiple 20-80 kb inversions, including the regions containing *SIX6*, *SIX11* and *SMP1*. Genetic factors that could cause Forc to be virulent on multiple hosts include ten genes in the inserted region that are not found in Fom001, as well as **(6)** genes located in the subtelomeric region on the short arm of chr^{RC} for which no alignment could be identified in Fom001 (though this is less likely, since no *mimps* and very few homologs of predicted effectors were identified here). The alternative to Forc having something 'extra' over Fom that allows it to have a wider host range is an absence of one or more avirulence factor(s) encoded in its genome. One likely location for such a factor could be the accessory chromosome largely contained in contig 22 of Fom001, where **(7)** a homolog of *SIX1* is positioned. Six1 (alternatively known as Avr3) produced by Fo f.sp. *lycopersici* (Rep *et al.*, 2004) as well as other *formae speciales* (Sri Widinugraheni, personal communication) is recognized by the tomato R protein I-3. Six1 may act as an avirulence factor in cucumber and watermelon. No *SIX1* homolog is present in any of the sequenced *cucumerinum*, *niveum* or *radicis-cucumerinum* strains. Another possibility could be that a gene with a different sequence type, for example **(8)** *SIX13* encoded on an accessory extension of chromosome 13 (contig 10) in Fom001, results in an avirulence reaction in other cucurbit plants than melon.

Potential ways of how this situation may have arisen and how chr^{RC} may have come into existence are depicted in Fig. 2. The finding that chr^{RC} and contig 127 are so similar to each other indicates a recent shared origin of these chromosomes or chromosome parts. One possibility is that the difference in host range and symptom development (root rot vs wilt disease) between these *formae speciales* is determined in the dynamic central region of Forc's pathogenicity chromosome. Potentially, a large insertion happened in an ancestral *melonis* chromosome, conferring the novel traits to *radicis-cucumerinum*. Alternatively, a deletion of this region in Forc occurred, resulting in a more restricted *forma specialis (melonis)* in terms of host range because of this. However, since a contig similar to chr^{RC} is not present in all strains of *melonis* (e.g. Fom010, that lacks such a contig), the former hypothesis seems more likely.

Only 10 out of 195 genes predicted in the 700 kb region of chr^{RC} that is different in Fom001 contig 127 could not be identified with BLASTN in Fom001. Interestingly, three of these genes (all involved in *beta*-lactam biosynthesis) solely returned BLAST hits from other cucurbit infecting ff. spp., in particular f. sp. *lagenariae, momordicae* and *cucumerinum*. Of three other genes (including a gene encoding a presumed NADH-flavin oxidoreductase and *SIX9*), homologs were found to be present in these same ff. spp. with high similarity to the Forc016 genes. One or several of these genes may thus contribute to the ability of Forc to cause root rot in several cucurbit species (but not *SIX9*, because knockout of this gene did not result in reduction of virulence). In order to conclude that the above scenario is indeed true, additional experiments would be necessary involving HCT of contig 127 into a Fo47 background. If the resulting strain has a Fo f. sp. *melonis* phenotype, it means that the chromosome containing contig 127 is the major determining factor for melon infection by Fom and that the differences between these two *formae speciales* are probably encoded in the dynamic middle region. It should be noted that in the Fom001 assembly, other *SIX* gene-containing contigs were found and that multiple chromosomes involved in *melonis* pathogenicity could be present. Alternatively, HCT of chr^{RC} from *radicis-cucumerinum* into a *melonis* background could quickly illustrate what the effect of the addition of the Forc-variant of the dynamic central region of this chromosome to a *melonis* genome would be. The expected outcome of this experiment would be a strain that is now able to colonize and cause root rot in multiple cucurbits.

Further recommendations for continuation of this exciting line of research include the use of Forc chromosome loss strain #2 (that has no accessory chromosomes anymore) as a Fom chromosome acceptor. Another interesting experiment that could be performed to further understand whether the subregion in chr^{RC} is in itself sufficient for Forc-like pathogenicity would be to generate partial chromosome deletion strains. A similar study has recently been published involving partial deletions of Fol4287 chromosome 14, showing the (limited) effect of gene deletion of multiple *SIX* genes at once on pathogenicity (Vlaardingerbroek, Beerens, Schmidt, *et al.*, 2016).

Pathogenicity chromosomes in *melonis, cucumerinum* and *niveum*

For the cucurbit infecting ff. spp. other than radicis-cucumerinum, less detailed comparative analyses have been performed. Preliminary analysis of *cucumerinum* shows that multiple evolutionary paths have led to the ability to infect cucumber plants. Though a significant part of the accessory genome of Foc strains is shared (exemplified by the comparison of effector presence), Foc015 possesses 10 Mb of accessory sequence of which most is strain-specific when compared to Foc013 and Foc021. Foc013 has about 2 Mb of sequence that is not found in Foc015, but particularly Foc021 is rich in strain-specific accessory DNA: ~20 Mb of sequence (29% of its total genome size) is specific compared to the other two *cucumerinum* genomes. The separate effector clustering of Foc021, Foc018 and Foc030 from the other *cucumerinum* strains (Chapter 2) and the fact that these strains all belong to a single vegetative incompatibility group (VCG) are indications that f. sp. *cucumerinum* is not a one-trick pony.
A similar conclusion can be drawn for Fo f.sp. *melonis*. A large genetic diversity was found during the process of designing and testing Fom markers. The three strains that were sequenced with PacBio each possess multiple Mb of strain-specific sequence that is not found in the others. Additionally, a strain capable of infecting both muskmelon and watermelon was identified that also reacted with markers of both ff. spp.
Accessory material of Fo f. sp. *niveum* strains seems to be more similar to each other, but here, too, many strain-specific sequences were discovered with sometimes high similarity to other cucurbit infecting ff. spp. This is in line with the fact that several effector candidates are 100% identical between *melonis, niveum, radicis-cucumerinum, cucumerinum* and the other cucurbit-infecting *formae speciales*.
Though it is clear that the *formae speciales* affecting cucurbits share part of their evolutionary trajectory, much more detailed comparative analysis of good quality genome assemblies is needed to reconstruct their evolution.

Final remarks

The results presented in this thesis have increased our understanding of the concept of host specificity in *F. oxysporum*. Additionally, the many genome assemblies generated in this project (accumulating to 74 novel wild-type strains sequenced as well as resequencing of about 30 more strains) have given much insight into just how common horizontal transfer of genes and chromosomes is within *F. oxysporum* as well as with species outside the FOSC. Additionally, they have allowed the design of f. sp.-specific primers and probes that can be used in practice for speedy and specific detection of pathogenic Fo strains.
Taken together, they have shown the enormous potential of genomics for understanding complex fungal population dynamics. In only a few years' time, we have seen the price for

a 100× covered Fo Illumina assembly drop more than fivefold. Initial PacBio assembly of Fo genomes, though much improved compared to Illumina-generated assemblies, still results in fragmentation into 32 to 183 contigs (principally due to repeat-rich accessory chromosomes). For further study of chromosome-level genome comparison, the quality of *de novo* genome assemblies could be further elevated with optical mapping and/or by using even longer reads. New developments in the field of DNA sequencing that can generate much longer sequencing reads at lower costs are a big promise for the future of fungal (patho)genomics.

References

Ayukawa, Y., Hanyuda, S., Fujita, N., Komatsu, K., and Arie, T. (2017) Novel loop-mediated isothermal amplification (LAMP) assay with a universal QProbe can detect SNPs determining races in plant pathogenic fungi. 1–9.

Baayen, R.P., O'Donnell, K., Breeuwsma, S., Geiser, D.M., and Waalwijk, C. (2001) Molecular Relationships of Fungi Within the *Fusarium redolens-F. hostae* Clade. *Phytopathology* **91**: 1037–44.

Cafri, D., Katan, J., and Katan, T. (2005) Cross-pathogenicity between *formae speciales* of *Fusarium oxysporum*, the pathogens of cucumber and melon. *J. Phytopathol.* **153**: 615–622.

Chellapan, B.V., van Dam, P., Rep, M., Cornelissen, B.J.C., and Fokkens, L. (2016) Non-canonical Helitrons in *Fusarium oxysporum*. *Mob. DNA* **7**: 27.

Clutterbuck, A. (2011) Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. *Fungal Genet. Biol.* **48**: 306–326.

Daboussi, M.-J., Davière, J.-M., Graziani, S., and Langin, T. (2002) Evolution of the *Fot1* transposons in the genus *Fusarium*: discontinuous distribution and epigenetic inactivation. *Mol. Biol. Evol.* **19**: 510–520.

van Dam, P., Fokkens, L., Schmidt, S.M., Linmans, J.H.J., Kistler, H.C., Ma, L.-J., and Rep, M. (2016) Effector profiles distinguish *formae speciales* of *Fusarium oxysporum*. *Environ. Microbiol.* **18**: 4087–4102.

van der Does, H.C., Fokkens, L., Yang, A., Schmidt, S.M., Langereis, L., Lukasiewicz, J.M., et al. (2016) Transcription Factors Encoded on Core and Accessory Chromosomes of *Fusarium oxysporum* Induce Expression of Effector Genes. *PLOS Genet.* **12**: e1006401.

van der Does, H.C. and Rep, M. (2007) Virulence genes and the evolution of host specificity in plant-pathogenic fungi. *Mol. Plant. Microbe. Interact.* **20**: 1175–1182.

Feschotte, C. and Pritham, E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**: 331–368.

Geiser, D.M., Juba, J.H., Wang, B., and Jeffers, S.N. (2001) *Fusarium hostae* sp. nov., a Relative of *F. redolens* with a *Gibberella* Teleomorph. *Mycologia* **93**: 670–678.

Glass, N.L. and Dementhon, K. (2006) Non-self recognition and programmed cell death in filamentous fungi. *Curr. Opin. Microbiol.* **9**: 553–558.

Houterman, P.M., Ma, L., Van Ooijen, G., De Vroomen, M.J., Cornelissen, B.J.C., Takken, F.L.W., and Rep, M. (2009) The effector protein Avr2 of the xylem-colonizing fungus *Fusarium oxysporum* activates the tomato resistance protein I-2 intracellularly. *Plant J.* **58**: 970–978.

Houterman, P.M., Speijer, D., Dekker, H.L., De Koster, C.G., Cornelissen, B.J.C., and Rep, M. (2007) The mixed xylem sap proteome of *Fusarium oxysporum*-infected tomato plants. *Mol. Plant Pathol.* **8**: 215–221.

Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., and Fan, W. (2009) The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* **41**: 1275–1281.

Jelinski, N., Broz, K., Jonkers, W., Ma, L.-J., and Kistler, H.C. (2017) Effector Gene Suites in Some Soil Isolates of *Fusarium oxysporum* Are not Sufficient Predictors of Vascular Wilt in Tomato. *Phytopathology* **17**: 842–851.

Jones, J.D.G. and Dangl, J.L. (2006) The plant immune system. *Nature* **444**: 323–329.

Jurado, M., Marín, P., Callejas, C., Moretti, A., Vázquez, C., and González-Jaén, M.T. (2010) Genetic variability and Fumonisin production by Fusarium proliferatum. *Food Microbiol.* **27**: 50–57.

Kang, S., Demers, J., del Mar Jimenez-Gasco, M., and Rep, M. (2014) *Fusarium oxysporum*. In, *Genomics of Plant-Associated Fungi and Oomycetes: Dicot Pathogens*. Springer, pp. 99–119.

Laurence, M.H., Summerell, B.A., and Liew, E.C.Y. (2015) *Fusarium oxysporum* f. sp. *canariensis*: evidence for horizontal gene transfer of putative pathogenicity genes. *Plant Pathol.* **64**: 1068–1075.

Lievens, B., Houterman, P.M., and Rep, M. (2009) Effector gene screening allows unambiguous identification of *Fusarium oxysporum* f. sp. *lycopersici* races and discrimination from other *formae speciales*. *FEMS Microbiol. Lett.* **300**: 201–215.

Lievens, B., Rep, M., and Thomma, B.P.H.J. (2008) Recent developments in the molecular discrimination of *formae speciales* of *Fusarium oxysporum*. *Pest Manag. Sci.* **64**: 781–788.

Ling, J., Zhang, J., Zeng, F., Cao, Y., Xie, B., and Yang, Y. (2016) Comparative genomics provide a rapid detection of *Fusarium oxysporum* f. sp. *conglutinans*. *J. Integr. Agric.* **15**: 822–831.

Ma, L.-J., van der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.-J., Di Pietro, A., et al. (2010) Comparative analysis reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367–373.

Ma, L.-J., Geiser, D.M., Proctor, R.H., Rooney, A.P., O'Donnell, K., Trail, F., et al. (2013) *Fusarium* pathogenomics. *Annu. Rev. Microbiol.* **67**: 399–416.

Mcmillan, T. (1986) Cross pathogenicity studies with isolates of *Fusarium oxysporum* from either cucumber or watermelon pathogenic to both crop species. *Ann. Appl. Biol.* **109**: 101–105.

Michielse, C.B. and Rep, M. (2009) Pathogen profile update: *Fusarium oxysporum*. *Mol. Plant Pathol.* **10**: 311–324.

Milani, N.A., Lawrence, D.P., Elizabeth Arnold, A., and Van Etten, H.D. (2012) Origin of pisatin demethylase (PDA) in the genus *Fusarium*. *Fungal Genet. Biol.* **49**: 933–942.

Möller, M. and Stukenbrock, E.H. (2017) Evolution and genome architecture in fungal plant pathogens. *Nat. Rev. Microbiol.* doi: 10.1038/nrmicro.2017.76.

Niño-Sánchez, J., Casado-Del Castillo, V., Tello, V., de Vega-Bartol, J.J., Ramos, B., Sukno, S.A., et al. (2016) The *FTF* gene family regulates virulence and expression of *SIX* effectors in *Fusarium oxysporum*. *Mol. Plant Pathol.* 1–18.

O'Donnell, K., Cigelnik, E., and Nirenberg, H.I. (1998) Molecular Systematics and Phylogeography of the *Gibberella fujikuroi* Species Complex. *Mycologia* **90**: 465–493.

Olivain, C. and Alabouvette, C. (1997) Colonization of tomato root by a non-pathogenic strain of *Fusarium oxysporum*. *New Phytol.* **137**: 481–494.

Olivain, C., Humbert, C., Nahalkova, J., Fatehi, J., L'Haridon, F., and Alabouvette, C. (2006) Colonization of tomato root by pathogenic and nonpathogenic *Fusarium oxysporum* strains inoculated together and separately into the soil. *Appl. Environ. Microbiol.* **72**: 1523–1531.

Rep, M., Van Der Does, H.C., Meijer, M., Van Wijk, R., Houterman, P.M., Dekker, H.L., et al. (2004) 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.

Rouxel, T., Grandaubert, J., Hane, J.K., Hoede, C., van de Wouw, A.P., Couloux, A., et al. (2011) Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nat. Commun.* **2**: 202.

Schmidt, S.M., Houterman, P.M., Schreiver, I., Ma, L., Amyotte, S., Chellappan, B., et al. (2013) MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genomics* **14**: 119.

Stępień, L., Koczyk, G., and Waśkiewicz, A. (2011) Genetic and phenotypic variation of *Fusarium proliferatum* isolates from different host species. *J. Appl. Genet.* **52**: 487–96.

Takken, F. and Rep, M. (2010) The arms race between tomato and *Fusarium oxysporum*. *Mol. Plant Pathol.* **11**: 309–314.

Thon, M.R., Pan, H., Diener, S., Papalas, J., Taro, A., Mitchell, T.K., and Dean, R. a (2006) The role of transposable element clusters in genome evolution and loss of synteny in the rice blast fungus *Magnaporthe oryzae*. *Genome Biol.* **7**: R16.

Vlaardingerbroek, I., Beerens, B., Rose, L., Fokkens, L., Cornelissen, B.J.C., and Rep, M. (2016) Exchange of core chromosomes and horizontal transfer of lineage-specific chromosomes in *Fusarium oxysporum*. *Environ. Microbiol.* **18**: 3702–3713.

Vlaardingerbroek, I., Beerens, B., Schmidt, S.M., Cornelissen, B.J.C., and Rep, M. (2016) Dispensable chromosomes in *Fusarium oxysporum* f. sp *lycopersici*. *Mol. Plant Pathol.* **17**: 1455–1466.

Williams, A.H., Sharma, M., Thatcher, L.F., Azam, S., Hane, J.K., Sperschneider, J., et al. (2016) Comparative genomics and prediction of conditionally dispensable sequences in legume–infecting *Fusarium oxysporum formae speciales* facilitates identification of candidate effectors. *BMC Genomics* **17**: 191.

Zhou, X.G. and Everts, K.L. (2007) Characterization of a Regional Population of *Fusarium oxysporum* f. sp. *niveum* by Race, Cross Pathogenicity, and Vegetative Compatibility. *Phytopathology* **97**: 461–469.

Summary

Fusarium oxysporum is a fungus that occurs frequently in soils worldwide, usually without negative consequences for other organisms. In some cases, however, it can cause wilt disease and root rot in various plant species. Because of this, it is listed in the top 10 most important plant-pathogenic fungi. An interesting aspect of the biology of *F. oxysporum* is the fact that every individual strain is restricted to causing disease in only one or a few plant species. When it is confronted with another plant, it will try to start the colonization process, but usually it cannot proceed further than the outer cell layers of the roots. The central question addressed in this thesis targets this host specificity: what determines the fact that one strain is capable of infecting tomato plants, whereas another strain specifically causes disease in melon plants. The answer is not only fundamentally interesting, but is also relevant for growers and plant breeders. The understanding of what makes it so the fungus can or cannot infect particular plants will help us in developing plant resistance against *Fusarium*.

To anwer this question, we looked at the DNA of many different fungal strains in order to compare them to each other. The DNA contains the information that determined how living organisms will look. It is the biological blueprint that is stored in the nucleus of each and every cell, including the cells of fungi. This information is arranged in small packages on the DNA: the genes. Each gene results in a protein; a product or building block that has a specific function for the organism.

From earlier research, it is known that the fungus produces small proteins that are secreted during plant infection and can act as a 'key' in this process. These proteins (called 'effectors') can deregulate the plant's defence responses, allowing the fungus to penetrate the root tissue, enter the vasculature of the plant and further colonize it. By carefully examining the DNA regions on which these previously described effectors are located, we found a way of predicting novel effectors.

In **chapter 2** we search the DNA of 59 individual *F. oxysporum* strains, most of which cause disease in cucumber, melon, watermelon and tomato, for the presence of new effector genes. The most important finding described in this chapter is that the strains affecting the same host plants also share a similar set of effectors. Moreover, it turned out that the genes encoding these effectors typically also share an identical DNA sequence between strains that cause disease in the same plant. This is surprising because normally spontaneous mutations in the DNA occur over time. These mutations are found back in the other genes of these fungi, which is an indication that the effector genes have probably been 'horizontally transmitted'. What this means will be explained in the next section.

In 2010, a strain causing tomato wilt disease, *F. oxysporum* f. sp. *lycopersici* (Fol) 4287, was found to have 15 chromosomes. The second smallest chromosome contains all of the effector genes of this strain. Moreover, under lab conditions it could be shown that this chromosome could be horizontally 'moved' (without the interference of sex) to a strain that previously

did not cause any disease in tomato. Upon transfer, the recipient strain turned out to have gained that capability of infecting tomato plants. The hypothesis was thus that this process of horizontal chromosome transfer may also occur frequently in the 'wild', since the recipient strain gains a huge evolutionary advantage.

In **chapter 3** we describe the chromosome composition of an extraordinary strain of *F. oxysporum:* Forc016, that is capable of infecting cucumber, melon *and* watermelon. It turned out that this strain, too, has a single chromosome that resembles the 'pathogenicity'-chromosome of Fol4287. By growing this strain in the presence of the non-pathogenic strain Fo47, we succeeded in transferring the chromosome to Fo47. The new strains turned out to be capable of the same as the Forc parent strain: all three tested plant species became diseased. In another experiment, we induced the loss of this chromosome in Forc016, resulting in strains that were no longer capable of causing disease.

We sequenced the Forc016 DNA with a relatively new sequencing technique ('PacBio'), that produces long stretches of DNA sequence. This allowed us to examine this chromosome as a whole in stead of in small parts, as is often the case with other sequencing methods. We compared the pathogenicity chromosome of Forc016 with a highly similar chromosome, that of Fom001 (only pathogenic towards melon plants). Large parts of the chromosome proved to be 100% identical to each other. However, we also found a region that was highly distinct between both strains. In this particular region, we identified almost all of the predicted effector genes. We therefore believe that this is an important region determining that one strain is restricted to only infecting melon plants while the other can additionally cause disease in cucumber and watermelon.

The fact that *Fusarium oxysporum* can so 'easily' pass on entire chromosomes to others explains why strains affecting the same plant species also have similar (identical) effectors. This knowledge has been applied in **chapter 4** by designing molecular markers based on these effector genes. At the moment, diagnostics is still often performed through time consuming disease assays to identify whether a newly found fungal strain can cause disease in the suspected target plant. With molecular (q)PCR markers, that can determine whether a particular DNA sequence is present in the found strain, this process can be performed many times faster. One of the conclusions drawn from the results presented in this chapter is that the chromosomes that allow pathogenic infection of cucumber, melon, watermelon and several other cucurbits probably have a shared evolutionary ancestry.

This chapter illustrates the step towards application in practice using results obtained in fundamental scientific research. For companies that e.g. produce cucumber seeds, it can be of imporant to have a quick and reliable detection method that allows them to check if their seed batches are not contaminated with pathogens towards cucumber plants.

Finally, in **chapter 5**, we examine three strains that were initially thought to also be *F. oxysporum,* but later turned out to belong to other *Fusarium* species. In the genomes of these strains, belonging to the species *Fusarium proliferatum* and *Fusarium hostae,* we find

indications of horizontal transfer of genetic material (maybe even parts of chromosomes) between *F. oxysporum* and other *Fusarium* species. Two exciting findings support this hypothesis: the presence of : i) so-called '*mimps*' (transposable elements) and ii) effector genes in the identified strains. The fact that these *mimps* and effectors were sometimes identical to DNA fragments of *F. oxysporum* was convincing for the hypothesis of horizontal transfer. These three strains were isolated from diseased lily and hyacinth flower bulbs. Thus, horizontal transfer of (part of) a chromosome potentially occurred in flower bulb fields.

All in all, the results in this thesis have brought us a little bit closer to understanding how host-specific pathogenic interactions work in *Fusarium oxysporum* and the evolutionary mechanisms behind them.

Nederlandse samenvatting

Fusarium oxysporum is een schimmel die veelvuldig in de bodem voorkomt, meestal zonder negatieve gevolgen voor andere organismen. In sommige gevallen kan hij echter verwelkingsziekte of wortelrot in verschillende plantensoorten veroorzaken. Hierdoor is hij opgenomen in de top 10 meest belangrijke plant-pathogene (ziekmakende) schimmels in een onderzoek uitgevoerd door het vakblad Molecular Plant Pathology in 2012. Een interessant aspect aan de biologie van *F. oxysporum* is het feit dat elke individuele schimmelstam beperkt is tot het ziek maken van slechts één of enkele plantensoort(en). Wordt de schimmel geconfronteerd met een andere plant, dan zal hij wel proberen binnen te dringen, maar lukt het meestal niet om verder te komen dan de buitenste cellagen van de wortels. De centrale vraag in dit proefschrift richt zich op deze 'host specificity', zoals het in de titel wordt genoemd: wat bepaalt nu dat de ene stam alleen tomatenplanten kan infecteren, terwijl de andere stam specifiek meloenplanten ziek maakt? Het antwoord is niet alleen wetenschappelijk interessant, maar is ook relevant voor kwekers en vooral plantenveredelaars. Immers, als we begrijpen wat de schimmel in staat stelt om bepaalde planten wel of niet te infecteren kan er ook gerichter gewerkt worden aan ziekteresistentie tegen *Fusarium*.

Om deze vraag te beantwoorden bepaalden we de DNA sequentie (lettervolgorde) van de genomen van verschillende schimmelstammen en vergeleken deze met elkaar. Het genoom bevat de informatie die bepaalt hoe levende organismen eruit komen te zien. Het is een biologische blauwdruk die ligt opgeslagen in de kern van elke cel, dus ook in de cellen van schimmels. Deze informatie ligt in pakketjes: de genen. Elk gen resulteert in een eiwit; een molecuul dat een bepaalde functie heeft voor het organisme.

Uit eerder onderzoek is gebleken dat *F. oxysporum* kleine eiwitjes aanmaakt die worden uitgescheiden tijdens het infectieproces en hierbij als 'sleutels' kunnen dienen. Deze eiwitten, genaamd 'effectoren', ontregelen bijvoorbeeld de verdedigingsreacties van de plant, zodat de schimmel de kans heeft om via de wortels de watervaten van de plant in te groeien en zo de plant verder te koloniseren. Door zorgvuldig te kijken naar in welke delen van het genoom de eerder gevonden effectoren zich bevinden, hebben we een manier gevonden om nieuwe effectoren te voorspellen.

In **hoofdstuk 2** doorzoeken we het DNA van 59 *F. oxysporum* stammen waarvan de meeste ziekte veroorzaken komkommer, meloen, watermeloen of tomaat. We zoeken hierbij naar nieuwe effectoren. De belangrijkste nieuwe vinding in dit hoofdstuk is dat de schimmels die dezelfde plantensoort ziek maken, ook een vergelijkbare set effectoren hebben. Daarnaast bleek dat de genen van deze effectoren over het algemeen ook precies dezelfde DNA sequentie hebben in stammen die dezelfde plant ziek maken. Dit is verrassend, omdat normaal gesproken spontane mutaties (veranderingen) optreden in het DNA gedurende de tijd. Deze mutaties zijn inderdaad wel terug te vinden in de andere genen van deze stammen. Dit is een indicatie dat de effectorgenen waarschijnlijk 'horizontaal' zijn overgedragen. Dit werd voor het

eerst gedemonstreerd in 2010, toen het genoom van een stam die tomatenverwelkingsziekte veroorzaakt, *F. oxysporum* f. sp. *lycopersici* (Fol) 4287, werd gesequenced. Het bleek dat het 4287 genoom bestaat uit 15 chromosomen, waarvan het één na kleinste chromosoom alle effectorgenen bevat. Bovendien bleek dat dit chromosoom onder lab-condities horizontaal overgedragen kon worden naar een stam die voorheen geen ziekte veroorzaakte in tomaten, zonder de tussenkomst van paring en meiose. Na ontvangst bleek deze ontvangende stam ook in staat tot het infecteren van tomatenplanten. De hypothese was dat dit proces van horizontale overdracht van genenpakketjes ook in het 'wild' veelvuldig voorkomt aangezien dit de overeenkomst in effectorgenen tussen stammen die dezelfde plant infecteren kan verklaren. In **hoofdstuk 3** beschrijven we de chromosomensamenstelling van een bijzondere *F. oxysporum* stam: Forc016, die in staat is komkommer, meloen én watermeloen te infecteren. Ook in deze stam blijkt een enkel chromosoom aanwezig te zijn die veel lijkt op het eerder beschreven 'ziekteverwekkende'-chromosoom van Fol4287. Door deze stam samen te laten groeien met de niet-pathogene stam Fo47, lukte het ons om stammen te selecteren waarbij dit chromosoom was overgedragen van Forc016 naar Fo47. De nieuwe stammen bleken ook in staat om alle drie de geteste plantensoorten ziek te maken, net als Forc016 zelf. Bij een experiment waarin we Forc dit chromosoom hebben laten verliezen, bleek dat die stammen geen ziekte meer konden veroorzaken.

Het genoom van Forc016 hebben we gesequenced met een nieuwe techniek ('PacBio'), die relatief lange stukken sequentie genereert. Hierdoor konden we het 'ziekteverwekkende' chromosoom van Forc016 in zijn geheel bekijken in plaats van in kleine stukjes, zoals vaak gebeurt met andere sequencing technieken. We vergeleken het 'ziekteverwekkende' chromosoom van Forc016 met een bijna gelijk chromosoom; die van Fom001, die alleen meloen kan ziek maken. Grote delen van deze chromosomen bleken 100% identiek aan elkaar te zijn. We vonden echter ook een regio die juist erg verschillend is tussen beide stammen. Daarop bevinden zich bijna alle effectorgenen. We vermoeden daarom dat daar de belangrijkste genen op liggen die 'host-specificity' bepalen.

Het feit dat *Fusarium oxysporum* zo 'gemakkelijk' hele chromosomen kan overdragen tussen stammen verklaart waarom stammen die dezelfde plant ziek maken ook dezelfde (identieke) effectoren hebben. Deze kennis hebben we in **hoofdstuk 4** toegepast door moleculaire merkers te maken op basis van effectorgenen. Momenteel worden er voor diagnostiek nog vaak tijdrovende ziektetoetsen gedaan om te testen of een nieuw gevonden schimmelstam ook daadwerkelijk ziekte veroorzaakt bij een bepaalde plant. Dat kan met moleculaire (q)PCR merkers veel sneller. Met deze technieken kan worden gekeken of een stukje DNA met een bepaalde sequentie voorkomt in de gevonden schimmel. Eén van de conclusies die we konden trekken uit de resultaten in dit hoofdstuk is dat de chromosomen die het mogelijk maken om komkommer, meloen, watermeloen en de andere komkommerachtigen te infecteren waarschijnlijk een gemeenschappelijke evolutionaire oorsprong hebben. Dit hoofdstuk illustreert de stap naar toepassing in de praktijk vanuit fundamenteel wetenschappelijk

onderzoek. Voor bedrijven die bijvoorbeeld komkommerzaden produceren kan het van belang zijn om een snelle en betrouwbare test te doen om te beoordelen of de zaden helemaal schoon zijn van komkommer-ziekteverwekkers.

Ten slotte bekijken we in **hoofdstuk 5** een drietal stammen waarvan werd aangenomen dat het *F. oxysporum* stammen waren, maar andere *Fusarium* soorten bleken te zijn. In het DNA van deze stammen, behorend tot de soorten *Fusarium proliferatum* en *Fusarium hostae*, vinden we aanwijzingen dat er horizontale overdracht van genetisch materiaal (misschien zelfs wel delen van chromosomen) heeft plaatsgevonden tussen *F. oxysporum* en andere *Fusarium* soorten. Twee opvallende bevindingen waren daarin ondersteunend: de aanwezigheid van (i) zogeheten '*mimps*' ('springende DNA elementen') en (ii) effectorgenen in de gevonden stammen. Het feit dat deze *mimps* en effectoren soms identiek zijn aan stukjes DNA van *F. oxysporum* ondersteunt de hypothese van horizontale overdracht. Deze drie stammen zijn geïsoleerd uit zieke lelie en hyacint bollen. Mogelijk heeft deze horizontale overdracht van een (stuk van een) chromosoom dus plaatsgevonden in een bloembollenveld.

Al met al begrijpen we dankzij de resultaten in dit proefschrift iets beter hoe de specifieke interactie tussen de ziekmakende schimmel *Fusarium oxysporum* en de plant plaatsvindt, en hoe deze evolutionair gezien tot stand is gekomen.

Thank you!

The moment is finally there to thank the people that helped me along the road to getting my PhD. There are many people that deserve my thanks and helped me with bigger or smaller things during the last 5 years at the molecular plant pathology department (de fyto's). Looking back, I can safely say that I am really happy that I chose to start a PhD. I have learned a lot about many things related to cloning, transformation, genome sequencing, (lots of) bioassays, python and R but importantly I also learned many things about myself. The people that I met along the way have been a huge help in getting there and I am thoroughly grateful for their support during these years.

Martijn, jou wil ik graag als eerste bedanken. Toen ik in 2012 begon aan het project had ik vrij weinig lab-ervaring maar je heb me alle ruimte gegeven om me te ontwikkelen in het lab en daarna een nieuwe weg in te slaan met de bioinformatica. Als ik soms moedeloos werd van tegenvallende resultaten had ik na ons wekelijkse overleg telkens weer een positief gevoel en zat ik vol nieuwe plannen. Dankjewel voor je begeleiding, mentorschap, luisterend oor en algehele enthousiasme!
Ben, je straalt altijd rust uit en je gaf me het vertrouwen dat het allemaal wel goed met me komt. Af en toe kwam je langs om het te hebben over of een effector wel of niet voorkomt in Fo5176 of een ander genoom en die discussies vond ik ook prettig! Bedankt ook voor je begeleiding van praktische zaken tegen het einde.
In het rijtje begeleiders horen natuurlijk nog **Frank** en **Harrold**. Bedankt voor jullie kritische input op woensdagochtend en de inhoudelijke discussies, maar ook voor de random facts en gezelligheid tijdens koffie en borrels!
Eén van de belangrijkste personen voor elke fyto-AIO is natuurlijk **Petra**. Petra, ik bewonder je people-skills en je organisatietalent waarmee je de labs waarin zoveel verschillende mensen werken zo goed draaiende houdt (is de oplossing overal labels op plakken?). Dankjewel voor je ervaren blik tijdens mijn experimenten, het kunnen delen van persoonlijke twijfels en de gezelligheid op congressen!

Some people were really closely involved in my project and I would like to thank them for the collaboration and the help they provided.
Like, jij had altijd nog wel "een scriptje liggen" voor iets dat ik probeerde te doen in python. We zijn samen aan dit project begonnen en hebben gaandeweg met zijn tweeën de meeste *Fusarium oxysporum* genomen ter wereld gegenereerd en geanalyseerd schat ik. Je hebt me ook op weg geholpen met mijn eerste python probeersels (omdat jij nu eenmaal in python programmeerde) en daar ben ik ook heel blij mee! Dankjewel voor je hulp en succes met je fantastische Veni project de komende jaren!
Mara, ik ben zo blij dat je het marker project hebt voortgezet. Het is sowieso ontzettend

fijn om een project uit wetenschappelijke interesse door te zien groeien tot een praktisch toepasbare methode en dankzij de samenwerking met NAKtuinbouw (en de inzet van oa. **Fleur**) is dit mogelijk geworden. Ik had me geen prettiger persoon kunnen voorstellen voor een directe samenwerking en ken ook niemand die (q)PCR's netter kan uitvoeren dan jij. Dankjewel!
De studenten die bij mijn promotietraject betrokken zijn geweest, **Demelza, Jasper, Michelle** en **Anneliek** wil ik allemaal bedanken voor hun inzet en de gezellige samenwerking. Jullie bijdragen (kloneringen, honderden/duizenden *Fusarium* transformaties, marker-PCRs, genoom-assemblies, om er een paar te noemen) zijn voor mij van onschatbare waarde geweest, thanks!
Sarah, thank you for helping me start up my project in the wonderful world of *Fusarium* genomics, effector prediction and miniature *impalas*. Good luck with your banana - *cubense* project!
Dear **Yu**, I would like to express my thanks for your help in the chromosome loss experiments that would not have been included to this thesis if it weren't for you and the fact that you refuse to give up when experiments don't work. Good luck with finishing your own PhD in Japan.
Dear **Ruth**, thanks for your help and collaboration when you were visiting our lab in 2013. I can clearly remember the endless bioassays with *Arabidopsis* and other experiments we did together. Good memories!
Beste **Balázs**, dankjewel voor de fijne samenwerking om het mitochondrieel DNA van het FOSC beter in kaart te brengen en het pacbio-puzzelen dat we samen deden voor hoofdstuk 3! Ook bedankt voor je fasta_tools, ik gebruik ze nog regelmatig.
Over the last couple of years, I have added 356 new wildtype *Fusarium* isolates to our -80°C freezer. I would like to thank the scientists that were willing to send me a selection of their strains. Their help has allowed me to perform this research in the first place. Thanks **Bart Lievens, D.J. Vakalounakis, Kathryne Everts, Takashi Tsuge, Corby Kistler**, **Kerry O'Donnell** and **employees of Naktuinbouw**. Dear **Li-Jun Ma**, thanks for your contribution of genome sequences to the effector profile chapter (chapter 2).
Ronald Wilterdink en **Karin Posthuma** van ENZA en **Eelco Gilijamse** van Rijk Zwaan wil ik graag bedanken voor de komkommer-, meloen- en watermeloenzaadjes die ze regelmatig naar me opstuurden.

I believe that we have a great atmosphere in our department that stimulates personal development and co-operation and I'm proud to have been part of this group of people.
Dear **Heni**, you are a very special person to me in the lab and I am really happy that you joined us a few years ago. Thanks for your companionship, regular coffee breaks and friendship and of course for the nice springrolls, dumplings and other Indonesian food now and then. Hang in there for your own project, good luck and go for it! **Xiaotang,** thank you for your help

with the frequent bioassays and the Chinese hotpot. **Lotje** and **Nico**, first of all thanks for your invaluable input during the design of new experiments that have helped me the last couple of years. I think you are both great scientists and I look up to you both. Also, I have fond memories of our trip before Asilomar 2015 to Sequoia National Park! **Ido**, bedankt dat je mijn meest luidruchtige, tonijn-uit-blik etende collega was. We hebben samen veel beleefd en je hebt me daarnaast wetenschappelijk geïnspireerd (bijvoorbeeld HCT & chr loss)! Onze trips naar Yosemite en natuurlijk Andalusië (oa. die vierpersoons hamburger die jij alleen opat) zijn onvergetelijk! **Shermineh**, thanks for some great trips to conferences, company in the lab, help with huge bioassays and that one poker night! **Valentin**, you're a cool dude and I loved your company and some parties we shared together, thanks a lot for that! I'll remember the boat ride! **Manon** and **Marijn**, thanks for cheering up the department, I've loved your company.

To all the people that were here when I started or that joined during my PhD, thanks for the awesome time! Dear **Magda M., Francesca, Biju, Lingxue, Hanna, Patrick, Magda J., Tieme, Mark, Sayantani, Bas, Manolo, Marieke, Till, Blaise, Alessandra, Bora, Benjamin, Ringo, Michel de V., Petra B.**, thanks so much for the great atmosphere in our group(s). I think we can be proud of what we have here and I look back on a lot of fun activities with you, like escape rooms, drinks, parties! Members of the running club, **Marijn, Mark, Marc, Georgios, Chris, Petra H.** and **Sayantani**, I enjoyed our regular runs on Wednesday afternoon, thanks! **Maria** (uhuu), **Jim** and **Fran**, the youngest *Fusari's* in our group, I wish you all the luck and success with your projects! Thanks for the great time together in the lab and I have faith in you!

Maaike, samen hebben we veel activiteiten georganiseerd op en rond het Science Park, we zaten vaak echt op dezelfde golflengte! Bedankt daarvoor! Ook de andere leden van de FPPF (inmiddels FPPFPCB) feestcommissie, **Ruy, Like, Selena**, **Ola, Marijn** en **Manon** wil ik bedanken voor het samen organiseren van gezellige borrels, ontbijtjes, Sinterklaasactiviteiten en labuitjes!

Christa, Teun, Rob, Marcel en **Michel H.**, bedankt voor jullie bijdragen tijdens PSM presentaties, discussies en gezelligheid op de afdeling.

Harold, Ludek en **Thijs,** mannen van de kas, bedankt voor het verzorgen van mijn plantjes en het gezellige contact! Dankzij jullie stond altijd alles netjes klaar voor mijn biotoetsen zodat ik gelijk de PKM3 kas kon induiken om komkommers en meloenen te infecteren, zelfs bij (regelmatige) last-minute aanvragen.

Machiel, ontzettend bedank voor het ontwerpen van de prachtige omslag! Ik ben er heel blij mee.

Manon en **Joël** wil ik bedanken voor de overleggen die we een paar keer hadden en waarin we gezelligheid en herkenning vonden. **Guido**, bedankt dat je mijn externe begeleider was.

Dan resteren natuurlijk nog de mensen die in de niet-werksfeer hun bijdragen geleverd hebben.

Mijn biologievrienden en vrienden van de middelbare school wil ik heel erg bedanken voor hun interesse en afleiding van het hele *Fusarium* gebeuren met leuke dingen zoals spelletjes, weekendjes weg en biertjes.
Lieve **mama**, jou wil ik speciaal bedanken voor je steun en eeuwige bezorgdheid de afgelopen jaren. Het is verdrietig dat papa er niet bij kan zijn nu, hij zou vast trots zijn. **Daan & Willemijn** (en de meisjes), **Loes & Fab** en **Hanneke**, bedankt voor jullie steun en het zijn van een hechte familie, dat is voor mij heel belangrijk. Op naar de volgende promotie! Ook **Wim, Els, Penny, Mieke** en **Craig** wil ik bedanken voor hun belangstelling en de ontzettend fijne, relaxte sfeer bij familiegelegenheden.
Tot slot, bedankt lieve **Kitty**. Bedankt voor het aanhoren van soms onbegrijpelijke verhalen. Bedankt dat je ervoor zorgde dat ik er vertrouwen in hield. Je hielp me bij de keuze om überhaupt in Amsterdam te beginnen en je steun heeft me ontzettend geholpen tijdens de daarop volgende jaren om dit boekje af te krijgen. Ik ben zo ontzettend blij en trots dat je mijn partner bent!

It is difficult to find exactly those words that reflect my feelings. I'm just really excited and happy to be finished and move on towards the next step. Cheers!

Peter.

List of publications

(newest at the top)

van Dam, P., de Sain, M., ter Horst, A., van der Gragt, M., and Rep, M. (submitted) Comparative genomics-based markers: discrimination of host-specificity in *Fusarium oxysporum*.

Brankovics, B., **van Dam, P.,** Rep, M., de Hoog, S.G., van der Lee, T.A.J., Waalwijk, C., and van Diepeningen, A.D. (2017) Mitochondrial genomes reveal recombination in the presumed asexual *Fusarium oxysporum* species complex. *BMC Genomics*. doi: 10.1186/s12864-017-4116-5

van Dam, P., Fokkens, L., Ayukawa, Y., van der Gragt, M., ter Horst, A., Brankovics, B., Houterman, P.M., Arie, T., and Rep, M. (2017) A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of multiple cucurbit species. *Scientific Reports* **7**: 9042. doi: 10.1038/s41598-017-07995-y

van Dam, P. and Rep, M. (2017) The Distribution of *Miniature Impala* Elements and *SIX* Genes in the *Fusarium* Genus is Suggestive of Horizontal Gene Transfer. *J. Mol. Evol.* doi: 10.1007/s00239-017-9801-0

Chellapan, B.V., **van Dam, P.**, Rep, M., Cornelissen, B.J.C., and Fokkens, L. (2016) Non-canonical Helitrons in *Fusarium oxysporum*. *Mob. DNA* **7**: 27. doi: 10.1186/s13100-016-0083-7

Fransz, P., Linc, G., Lee, C.-R., Aflitos, S.A., Lasky, J.R., Toomajian, C., Hoda, A., Peters, J., **van Dam, P.**, *et al.* (2016) Molecular, genetic and evolutionary analysis of a paracentric inversion in *Arabidopsis thaliana*. *Plant J.* **88**(2): 159-178. doi: 10.1111/tpj.13262.

van Dam, P., Fokkens, L., Schmidt, S.M., Linmans, J.H.J., Kistler, H.C., Ma, L.-J., and Rep, M. (2016) Effector profiles distinguish *formae speciales* of *Fusarium oxysporum*. *Environ. Microbiol.* **18**: 4087–4102. doi: 10.1111/1462-2920.13445

Schmidt, S.M., Lukasiewicz, J., Farrer, R., **van Dam, P.,** Bertoldo, C., and Rep, M. (2016) Comparative genomics of *Fusarium oxysporum* f. sp. *melonis* reveals the secreted protein recognized by the *Fom-2* resistance gene in melon. *New Phytol.* **209**: 307–318. doi: 10.1111/nph.13584