Segmental Duplications Drive the Evolution of Accessory Regions in a Major Crop Pathogen

A.C. van Westerhoven[1,2], C. Aguilera-Galvez[1], G. Nakasato-Tagami [1], X. Shi-Kunne [1], J. Dijkstra[1], E. Martinez de la Parte[1], E. Chavarro Carero[1], H.J.G. Meijer[1,3], A. Feurtey[4,5], N. Maryani[6], N. Ordóñez[1], R. Hofstede[7], A. H. J. Wittenberg[7], E.H. Sørensen[7], F. García-Bastidas[7], H. Schneiders[7], K. Nijbroek[7], E.H. Stukenbrock[4], G.H.J. Kema[1*#], M.F. Seidl[2*#]

[1]Wageningen University, Laboratory of Phytopathology, Wageningen, the Netherlands

[2]Utrecht University, Department of Biology, Theoretical Biology & Bioinformatics, Utrecht, the Netherlands

[3]Wageningen University, Department Biointeractions and Plant Health, Wageningen, the Netherlands

[4]Christian-Albrechts University of Kiel, Germany, and Max Planck Institute for Evolutionary Biology, Plön, Germany

[5]Eidgenössische Technische Hochschule Zürich, Plant Pathology, D-USYS, CH-8092, Zurich, Switzerland

[6]Biology Education, Universitas Sultan Ageng Tirtayasa, Banten, Indonesia

[7]KeyGene, Wageningen, the Netherlands

*Authors contributed equally

[#]Corresponding authors: Gert H.J. Kema (gert.kema@wur.nl) and Michael F. Seidl (m.f.seidl@uu.nl)

Abstract

Many pathogens evolved compartmentalized genomes with conserved core and variable accessory regions which carry effector genes mediating virulence. The fungal plant pathogen *Fusarium oxysporum* has such accessory regions, often spanning entire chromosomes, yet understanding of their evolution in strains that infect the same host remains limited. Here, we define the pan-genome of 69 diverse *Fusarium* strains that cause Fusarium wilt of banana, a significant constraint to global banana production, and demonstrate that segmental duplications drive the evolution of accessory regions. Accessory regions are highly variable and encode distinct gene repertories, with evolutionarily less constrained, younger genes and *in planta*-induced effectors, which differ between strains infecting the same banana cultivars. We show that recent segmental duplications and aneuploidy of accessory chromosomes cause the expansion of accessory regions in some *Fusarium* species. Taken together we conclude that extensive recent duplications drive the evolution of accessory regions in *Fusarium*, which contribute to the evolution of virulence.

Introduction

The interaction between fungal plant pathogens and their hosts drives rapid diversification[1]. Plant hosts have evolved sophisticated immune systems to detect pathogens, while adapted pathogens, in turn, secrete effector molecules to deregulate immune responses and to support host colonization[2,3]. To facilitate this co-evolutionary cycle, many filamentous plant pathogens evolved compartmentalized genomes where effector genes are localized in distinct genomic regions[4,5]. These effector-rich compartments show extensive genetic variation and are enriched for repetitive sequences such as transposable elements[1,4–7]. This spatial separation is often referred to as the 'two-speed' genome[4] and is thought to facilitates the rapid diversification of effector gene repertoires to enable pathogens to evade host resistances and to enable host jumps[8].

Fusarium oxysporum is a genetically diverse fungal species complex that consists of three clades[9]. Members of this species complex can collectively infect more than 100 economically important crops. In contrast, individual strains typically only infect a single host plant[10]. *Fusarium oxysporum* evolved a two-speed genome organization where specific genomic regions are characterized by extensive presence-absence polymorphisms between strains[11–14]. These variable accessory regions (ARs) can be embedded in core chromosomes or can encompass entire accessory chromosomes[11,12]. Interestingly, shared ARs in otherwise genetically diverse strains have been linked to the adaptation towards the same host[11,15]. Moreover, ARs can transfer the capacity to infect a specific host between isolates[12,16]. For example, transfer of accessory chromosome 14 from a tomato infecting *Fusarium* (*Fol*4287) to a non-

pathogenic *Fusarium* strain converts the non-pathogenic isolate into a tomato infecting pathogen[12]. This transfer of pathogenicity is likely linked to effector genes that are localized in ARs such as the 14 well-studied *Secreted In Xylem* (*SIX*) effectors[17]. Consequently, the presence and absence of effectors can group genetically diverse *F. oxysporum* strains based on their host range[18]. Despite the importance of ARs for host adaptation, little is known about the origin and evolutionary dynamics of these genomic regions in genetically diverse species that cause disease in the same host.

Fusarium wilt of banana (FWB), which is caused by a suite of *F. oxysporum* species[9], is a major constraint for banana production and threatens food security in tropical and subtropical countries where >400 million people depend on banana cultivation[19]. In *F. oxysporum* species causing FWB, three different races can be distinguished based on their pathogenicity towards subsets of banana varieties; Race 1 (R1) strains cause FWB in Gros Michel, Race 2 (R2) strains infect Bluggoe, and Tropical Race 4 (TR4) strains infects Gros Michel, Bluggoe, and Cavendish[20]. Additionally, all races infect other locally important banana varieties. TR4 has developed into a pandemic over the last 60 years[21,22] and is of particular concern as it causes FWB in Cavendish, the banana variety that dominates global productions (>50%) and export trade (>95%)[19]. Despite the impact of FWB on banana production, we know little about the ARs in *Fusarium* strains that infecting banana and their relation to pathogenicity.

Here we analyze the dynamics of ARs in a suite of *Fusarium* strains causing FWB. To gain insights into the genomic structure, we generate seven chromosome-level reference genome assemblies for R1, R2 and TR4 strains, and use a pan-genomic

framework together with a global panel of 69 strains to gain insights into the evolution of ARs in genetically diverse *Fusarium* species

Results

Fusarium strains *infecting banana are genetically diverse*

To study the occurrence and evolution of ARs[12,13,23] in *Fusarium* strains infecting banana, we sequenced and assembled 69 strains (**Fig. 1a**, **Supplementary table 1**) that were isolated from all major banana growing regions and classified into R1 (39 strains), R2 (two strains) and TR4 (28 strains) (**Fig. 1a,b**, **Supplementary table 1**). *De novo* assembly of seven long-read sequencing strains (three TR4, two R1, two R2) achieved chromosome-level contiguity containing 12 to 15 contigs, most of which represent complete chromosomes flanked by telomeric repeats (**Supplementary table 2,3**). While genome assemblies obtained solely from short-read data yielded more fragmented assemblies, all assemblies contained at least 96.9% of the single-copy BUSCO genes (**Supplementary table 1**). The assemblies range from 43 to 51 Mb in size (**Fig. 1b**) and encode between 15,664 and 17,865 predicted proteins, of which 551 to 671 are effector candidates, and consist for 2.6 – 8.9% of repeats (**Table S1**, **Fig. 1e,f**), similar to previous *Fusarium* genome assemblies[14,18,24].

To assess the genetic diversity of *Fusarium* strains infecting banana strains in relation to 55 *Fusarium* strains infecting eight different host plants[18] (**Supplementary table 1**), we conducted a phylogenetic analysis based on 3,811 single-copy orthologous genes (2,226,902 amino acid positions). As expected, *Fusarium* strains infecting banana are polyphyletic and associated with *F. oxysporum* clades 1, 2, and 3 (**Fig. 1a**)[9,25]. We observed low nucleotide diversity between TR4 strains (pi-nucleotide diversity 0.0018) compared to a 100x higher diversity between R1 strains (pi-nucleotide diversity 0.185), corroborating that TR4 strains evolved from a single recent clonal

origin[9,21]. Based on their genetic diversity, *Fusarium* strains infecting banana have recently been separated into different species, and the lineage that encompasses TR4 strains is now referred to as *Fusarium odoratissimum* (**Fig. 1d**)[24,26,27].

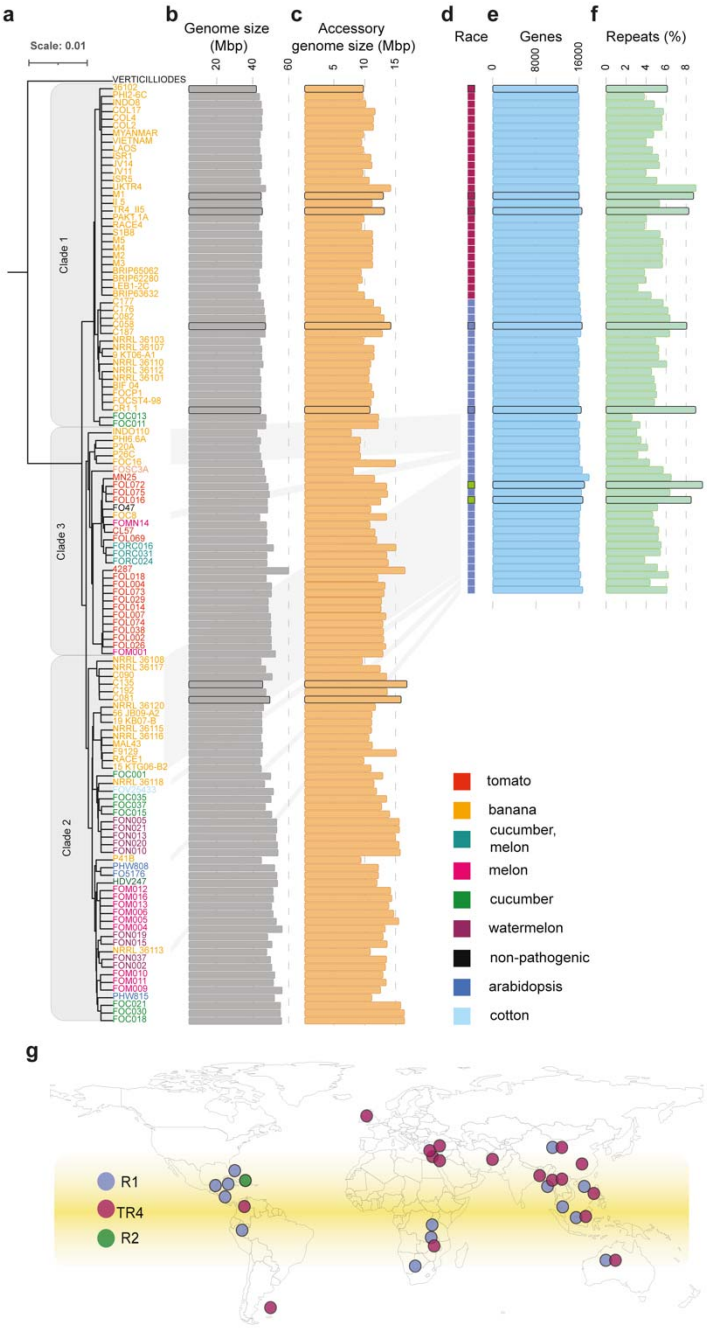


Figure 1 – *Fusarium* strains infecting banana are genetically diverse. a) *Fusarium* strains that infect banana are polyphyletic (yellow labels). The relationship between *Fusarium* infecting banana and *Fusarium* strains that infect other plant hosts (see label color) is estimated using maximum-likelihood phylogenetic analysis based on 3,811 concatenated conserved single-copy BUSCO genes. **b)** Genome sizes of *Fusarium* strains range between 43 to 61 Mb. **c**) The sizes of accessory regions in *Fusarium* strains range from 6.7 to 16.5 Mb. **d**) TR4 strains (purple squares) occur in clade 1 and

are genetically similar. By contrast, the R1 strains (blue squares) are polyphyletic. **e**) *Fusarium* strains that infect banana encode between 15,664 and 17,865 genes and **f**) between 2.9 and 8.9% repeats. **g)** The collection of 69 *Fusarium* strains originates from different banana growing regions worldwide; the approximate origin of isolation is indicated by dots, colored according to the race designation (R1, R2, and TR4). Strains sequenced by long-read sequencing technology are highlighted in black.

Fusarium *strains causing Fusarium wilt of banana have a compartmentalized genome*

To identify ARs in *Fusarium* strains infecting banana, we performed whole-genome alignments of the seven chromosome-level genome assemblies to the reference genome assembly of the tomato infecting *F. oxysporum* strain 4287 (*Fol*4287)[12]. We observed eleven homologous chromosomes between the TR4 reference strain II5 and *Fol*4287 (chromosome 1-11; **Fig. 2a**), as well as two large ARs specific for II5. The first spans 1.8 Mb occupying 27% of chromosome 1 in proximity to one telomere and the second region is 1.1 Mb in size and constitutes the complete chromosome 12 (**Fig. 2a**).

Compared to the core chromosomes, the two ARs have an increased repeat content (28.30% vs 6.41%) and a decreased gene density (32.15% vs 59.71%) and GC content (46.4% vs 47.5%). Importantly, the ARs in II5 encode 463 genes, nine are homologous to *SIX* genes and 22 are predicted effectors (out of 629 predicted effectors). Although this suggests that the ARs are important for pathogenicity on banana, they are not co-linearity with other *Fusarium* strains infecting banana (**Supplementary Fig. 2**).

To further analyze the diversity of ARs, we identified ARs in our global collection using a pan-genomic approach based on all-vs-all whole genome alignments. ARs were defined as regions longer than 5 kb that are absent in more than 55 of 69 strains (80%).

We identified a varying amount of ARs in *Fusarium* strains infecting banana (**Fig. 1c**), ranging from 6.7 Mb (15% of the genome size) in strain Indo110 (R1) to 15.5 Mb (30%) in strain C135 (R2), in line with the 19 Mb (29%) of ARs in the reference strain *Fol*4287[12]. Next to the two ARs in chromosomes 1 and 12, our pan-genomic approach identified additional 8 Mb of ARs in II5, typically localized at sub-telomeric regions (defined as the first and last 10% of each chromosome) (**Fig. 2b**). These sub-telomeric regions contain 1,287 genes, of which 80 are predicted effector genes (6.6%). We also identified that chromosomes 9 (42.6% ARs, 87 SNPs per kb), 10 (27% ARs, 80 SNPs per kb), and 11 (36% ARs, 92 SNPs per kb) are less conserved than the other core chromosomes (on average 20% ARs, 59 SNPs per kb) **(Supplementary Fig. 1)**, which resembles the situation of the previously identified 'fast-core' chromosomes in *Fol*4287[11].

We show that ARs can be defined based on sequence conservation and differ from the core genome in gene, repeat and GC-content. Consequently, principal component analyses distinguished ARs from the core genome based on these four features (**Fig. 2d**). Unanticipatedly, this analysis also separated the large ARs on chromosome 1 and 12 from the ARs localized at the sub-telomeres (**Fig. 2d**), and this separation is driven by the slightly lower GC content in sub-telomeric regions (44% vs 47%). A reduced GC-content can be caused by repeat induced polymorphisms (RIP), a fungal defense mechanisms that introduces C to T mutations in repetitive regions[28]. In the sub-telomeric region, 25.4% of the repetitive elements are likely subjected to RIP (i.e., composite RIP index > 0), in contrast to only 7.7% in the other ARs, suggesting

that repetitive elements at the sub-telomeres are older, or alternatively, that RIP has been more active.

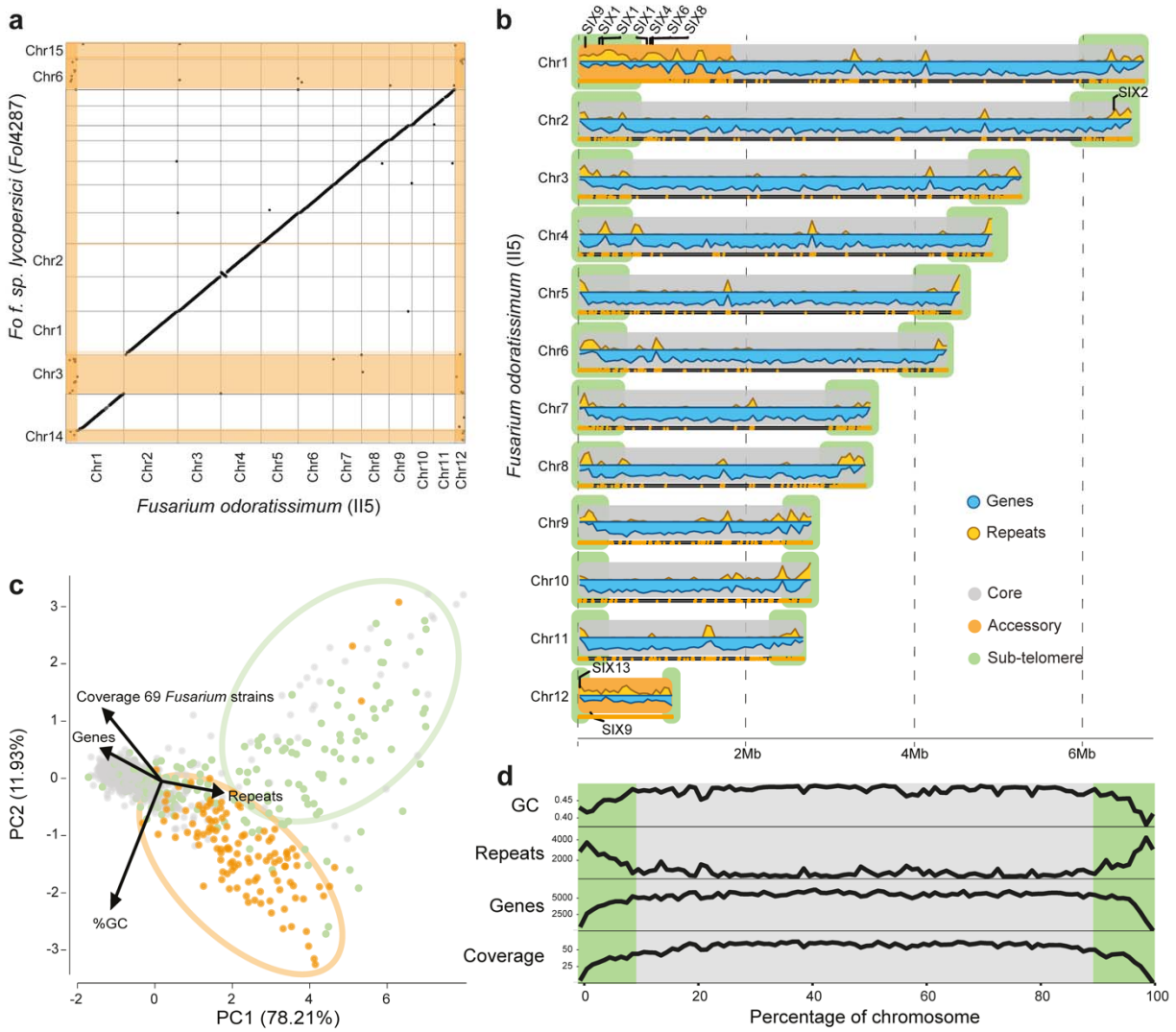


Figure 2 - *Fusarium* strains infecting banana carry accessory regions (ARs). a) The TR4 reference strain II5 contains eleven core chromosomes (chromosome 1-11) as well as two ARs (1,8 Mb on chromosome 1 and entire chromosome 12). Core chromosomes align to *Fol*4287 in contrast to the ARs that are highlighted in orange. **b**) The ARs (orange) in II5 carry most *Secreted In Xylem* (*SIX*) genes and show an increased repeat and lower gene content. A similar pattern is found in the sub-telomeric regions (green). **c**) Principal component analysis on gene-, repeat-, GC-content, and coverage of 69 *Fusarium* strains separates the genomic regions of II5 into three different clusters. ARs (orange and green) are separated from the core region (grey) and further separated into those ARs localized in sub-telomeric regions (green) and those on chromosome 1 and 12 (orange). **d**) Sub-telomeric regions (the first and last 10% of the chromosome) have different genomic features than the remainder of the chromosome. Sub-telomeric regions in II5 show a lower gene content, higher repeat

content, less similarity to other banana infecting strains (lower sequence coverage), and fewer genes.

Highly variable accessory regions in* Fusarium *strains infecting banana

To characterize the similarity of ARs in *Fusarium* strains infecting banana, we compared the ARs from the seven chromosome-level genome assemblies. Notably, the ARs are highly diverse with little sequence and structural similarity (**Fig. 3**, **Supplementary Fig. 2**). In TR4, the ARs from strain II5 on chromosomes 1 and 12 show extensive similarity to the corresponding chromosomes in strain M1 (**Fig. 3a**, **Supplementary Fig. 2**). Similarly, two accessory chromosomes of the R2 strains correspond to each other (**Supplementary Fig 2**), but no shared ARs are found between R1 strains. Moreover, we observed that no ARs are conserved between different races, for example only limited genetic material is shared between ARs of TR4 and R2 (S**upplementary Fig 2**). The diversity of ARs was further corroborated by pairwise analyses of all 69-strains, which revealed that, on average, TR4 shared 91% of the ARs, whereas R1 shared only 24% (**Fig. 3b**). In line with the high number of shared ARs in TR4 (*F. odoratissimum*, **Fig 3b**), the amount of shared ARs was highest between R1 strains belonging to the same *Fusarium* species (60% shared ARs**, Fig 3b**, **Supplementary Fig. 3**), however R1 strains from different species share very few ARs (**Fig. 3b**).

To quantify to which extent *Fusarium* strains infecting banana share ARs with *Fusarium* strains infecting different plant hosts, we identified and compared ARs in a set of 55 additional *Fusarium* strains that are pathogenic to different plant hosts[18]. We observed that genetically related strains share most ARs (**Supplementary Fig. 4**), however we also noticed that *Fusarium* strains infecting banana encode more diverse

ARs than for instance the tomato-infecting strains; on average, *Fol* strains share 62% ARs, with a minimal 30% of the ARs shared between any two *Fol* strains (**Supplementary Fig. 4**). In contrast, banana-infecting strains on average share 44% of the ARs and a pair of strains can share as little as 12% (**Supplementary Fig. 4**). Thus, we demonstrate that ARs are highly variable between *Fusarium* strains infecting banana and, importantly, strains do not share one host-associated chromosome in contrast to what was previously reported for *Fusarium* infecting tomato[11].

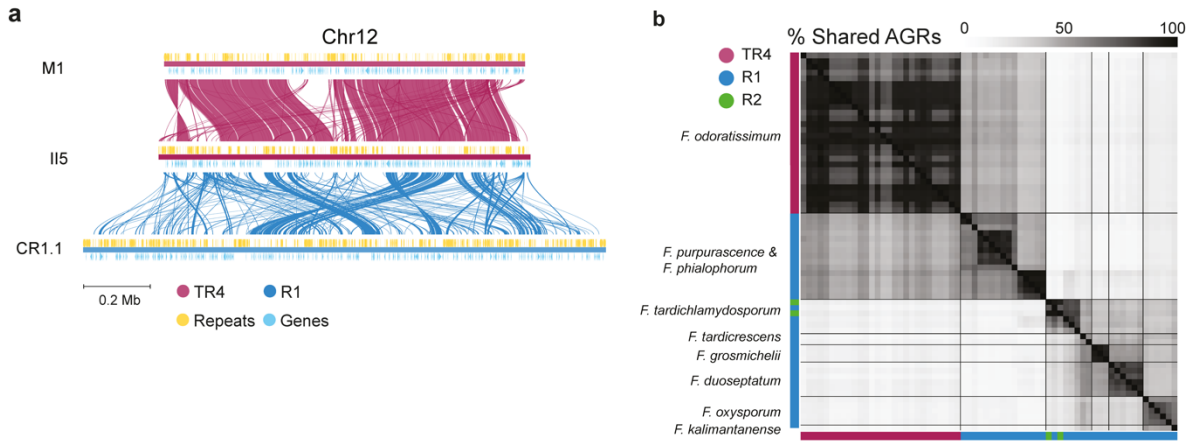


Figure 3 - Accessory regions in *Fusarium* strains infecting banana differ significantly. a) Accessory chromosome 12 of the TR4 reference strain II5 is shared with TR4 strain M1, however this chromosome shows little similarity to R1 strain CR1.1. Lines (red and blue) indicate aligned regions between strains. Small alignments are present between R1 and TR4 (blue lines), however these are rearranged and often involve repetitive sequences (yellow blocks). **b**) R1 strains (39 strains, highlighted in blue) have a diverse accessory content sharing on average only 20% of the ARs, whereas TR4 strains (28, highlighted in purple) share most of their ARs (91%). Phylogenetically related strains, here indicated by their species name[9], share more ARs than distantly related strains.

Evolutionary *dynamics of the accessory genome*

Fusarium strains infecting banana do not share large ARs (**Fig. 3b**), but we reasoned that ARs nevertheless might share genes that contribute to the pathogenicity towards banana. We therefore grouped 1.1 million protein-coding genes predicted in 69 strains into 22,612 orthologous groups, and subsequently analyzed the gene-content diversity. The pan-genome consisted of 12,101 core groups (53%) that are present in all 69 strains, 2,395 softcore groups present in at least 80% of the strains, 5,595 (25%) accessory groups present in less than 80% of the strains, and 2,521 unique genes (**Fig. 4a**). Importantly, the pan-genome based on our collection captures the diversity of protein-coding genes in *Fusarium* strain infecting banana as we did not observe an increase in the pan-genome size after including more than 40 strains (**Fig. 4a**). The conserved core genes are, as expected, enriched with genes with housekeeping functions, whilst accessory genes are enriched with genes encoding secondary metabolites (**Supplementary Fig. 5**). No enrichment of effector genes is found in any of the three categories (core, accessory, or softcore); the pan-genome consists of 739 gene families encoding predicted effectors that are evenly distributed over the different gene categories (3.7% of accessory genes and 3.5% of the core genes). The role of accessory genes in host infection is suggested by gene expression, as accessory genes are upregulated eight days after banana infection, with a median $\log_2$fold change of 1.59 (compared with the *in vitro* control) (**Fig. 4d**), comparable to the upregulation of effector genes (median $\log_2$fold change of 2.4). The core genes, however, show significantly less increase of gene expression upon plant infection (median $\log_2$fold change = 0.25, $P < 0.05$, two-sided Mann-Whitney U test) (**Fig. 4d)**. Although this suggests a role of

accessory genes in host-pathogenicity, we observed a highly diverse gene content in R1 strains with only 246 out of 7,832 accessory genes (3%) being shared among all R1 strains (**Fig. 4f**). TR4 strains, in contrast, have a more similar gene profile and share 2,493 of the 4,406 accessory genes (57%).

In II5, ARs have a different gene composition than the core regions, with ARs consisting mostly of accessory genes (405 out of 463; 87%), however four core and 38 soft core genes are also found in accessory regions (**Fig 4e**), indicating that some shared genes can be found between otherwise non-conserved regions. ARs seem to contain recently evolved genes, as most AR genes (71%) have homologs in closely related fungi (**Fig. 4 g,h**) and only 29% are considered “old”, i.e., with homologs in all fungi, which contrasts to the 53% old genes in the core genome (7,764, out of 14,774 genes) (**Fig. 4g,h**). Additionally, accessory genes are significantly shorter (315 aa) and evolve under relaxed selective pressure (dN/dS = 0.20), compared to the core genes (469 aa, dN/dS = 0.089; $P < 0.05$, two-sided Mann-Whitney U test) (**Fig. 4b,c),** both hallmarks of more recent genes[29]. This suggests that most genes in the ARs evolved recently, and that these regions serve as cradles for the evolution of novel genes. In addition, many genes on the ARs (177 out of 463) belong to expanded gene families (**Fig. 4e**), suggesting that gene duplications play a role in the evolution of ARs in *Fusarium* strains infecting banana.

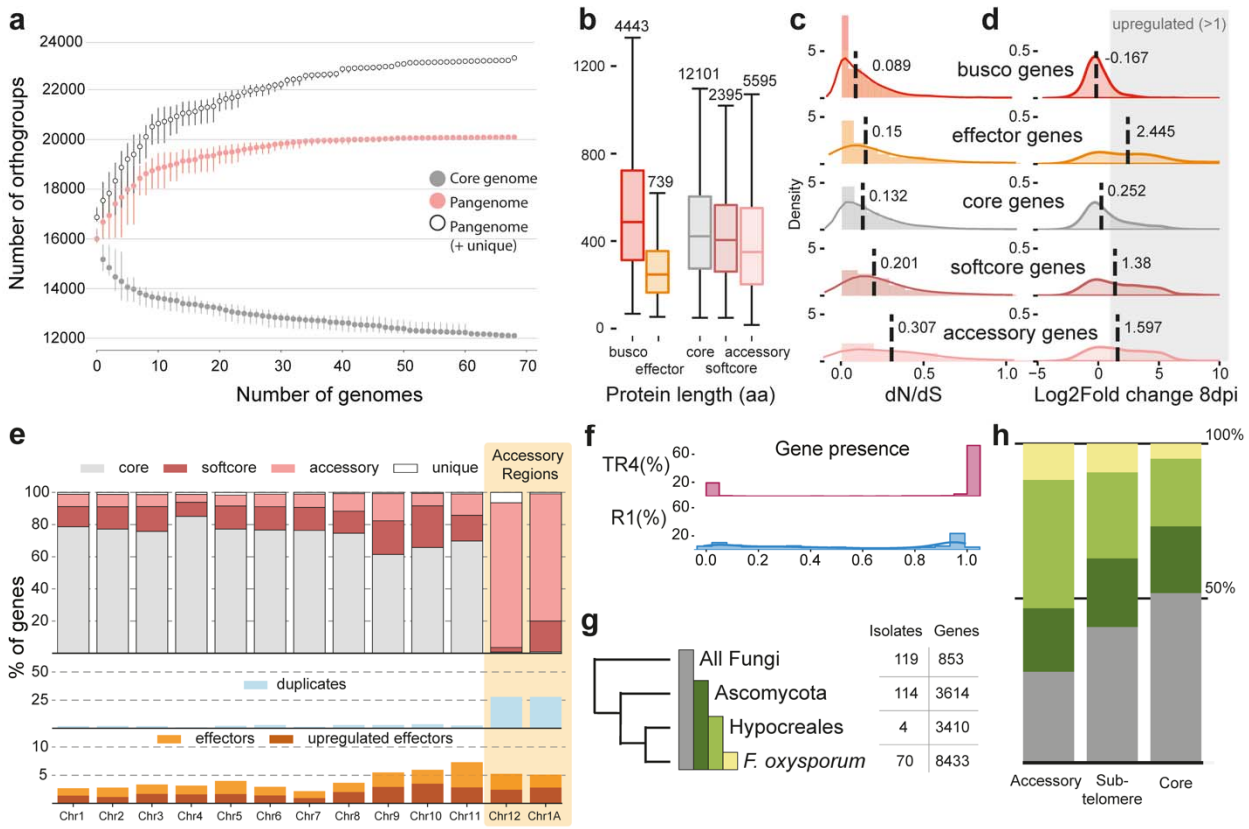


Figure 4 - Evolutionary dynamics of genes in the different genomic compartments in *Fusarium* strains infecting banana. a) Global collection of *Fusarium* infecting banana has a pan-genome size of 20,091 orthologous groups (OG, 22,612 when unique genes are included), of which 12,101 are present in all strains (grey). The pan-genome is saturated after adding 40 genomes. **b)** Distribution of protein length separated by different gene categories shows that accessory gene families (5,595 OGs, pink) are shorter (median length of 315 aa per OG) than core gene families (median length 469 aa per OG, 12,101 OGs, grey). All gene categories differed significantly from each other ($p<0.05$, Mann-Whitney U test). **c**) Sequence conservation estimated by dN/dS values. BUSCO gene families (4,443 OGs) and core gene families are most conserved, whereas the accessory genes evolve under more relaxed selective pressure. All gene categories show significant variation in dN/dS value ($p<0.05$, Mann-Whitney U test). **d**) Accessory genes are upregulated *in planta* at 8 days post inoculation. Average log2fold change of accessory genes (2,416 in II5) and effector genes (629 in II5) shows these genes are typically upregulated upon plant infection, which contrasts the core (11,740 genes, log_2-fold change = 0.252) and BUSCO genes (4,364 genes, log_2-fold change = -0.167). The median log2fold change of all gene categories is significantly different ($p<0.05$, two sided Mann-Whitney U test). **e**) ARs in TR4 strain II5 (orange) consist mostly of accessory genes (pink, 405 out of 463, 87%) and 28% of these genes is part of expanded orthogroups (129 out of 463). Accessory genes can also be found in core chromosomes. **f**) Gene content is variable between R1

strains, and only 20% of the genes is present in 80-90% of the R1 strains. Most genes (60%) in TR4 are present in all TR4 strains (90-100%). **g)** Gene families in TR4 strain II5 differ to which extend they have homologs in 308 fungal phyla; most genes are present in related *Fusarium* strains, but only a few gene families (853) are conserved among all considered fungi. **h**) Accessory regions contain most recent genes, present in *Fusarium* (51) and Hypocreales (160). Genes in core regions are evolutionarily older with more genes that have homologs in 'all fungi' (7,764 out of 11,740 genes).

Variable effector repertoire in* Fusarium oxysporum *genomes

Effector gene repertoires define host range[18,30], and we anticipated that effector presence-absence profiles would distinguish *Fusarium* strains infecting banana from *Fusarium* strains infecting other hosts. We predicted candidate effector genes in all 124 *Fusarium* isolates based on their proximity to Miniature IMPala elements (MIMP) that are known to co-localize with some effector genes in *F. oxysporum*[18,30], which yielded 398 MIMP-associated effector gene families (160 of these were predicted in the complete set of 739 candidate effectors). The predicted MIMP-associated effector repertoires clustered *F. oxysporum* isolates based on the host range[18,30] (**Fig. 5)**. While banana-infecting strains form a clear cluster, only four MIMP-associated effectors were identified that occur in most (i.e., >80%) of banana-infecting strains but are consistently absent (i.e., present in less than 20%) in other *F. oxysporum* strains. Moreover, races within *Fusarium* strains infecting banana are not clearly separated; TR4 strains have very similar effector profiles, yet R1 strains differ considerably and do not encode shared effectors. The dissimilarity of the effector repertoire in R1, together with the variable ARs and gene content, suggests that the species encompassing R1 utilize varying molecular mechanisms to support banana infection.

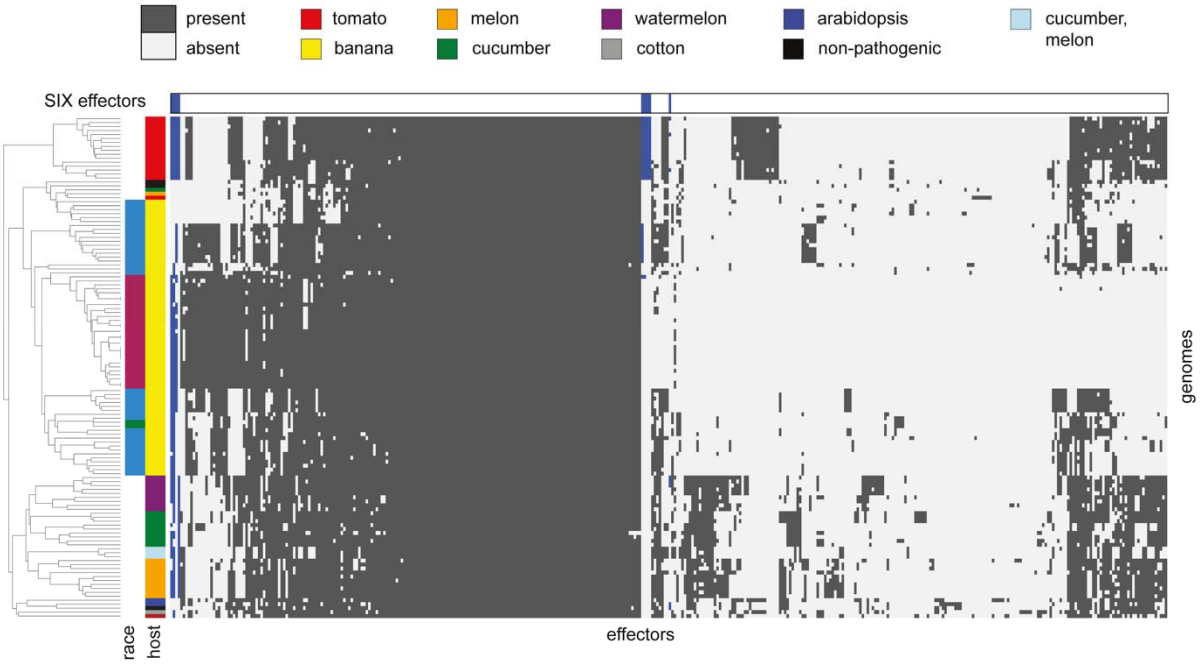


Figure 5. *Fusarium* strains infecting the same host cluster on predicted effector profile *Fusarium* strains infecting banana (yellow) are separated from *Fusaria* infecting other hosts. In the set of 124 *Fusarium* isolates, 398 MIMP-associated effector families are predicted. The R1 strains (blue) have diverse effector profiles and do not cluster together, whereas the effector profiles of TR4 strains (purple) are highly similar. Nine *SIX* effectors (blue) are identified in the 398 MIMP-associated effector families and are present in *Fusarium* strains infecting tomato, from which they are originally identified, as well as in other *Fusarium* genomes.

***Recent segmental duplications drive accessory genome expansion in* Fusarium oxysporum.**

We observed diverse ARs in *Fusarium* strains infecting banana with a high abundance of evolutionary young genes and genes that evolved via gene duplications (**Fig. 4e**). Duplications have been previously reported in *F. oxysporum* and also in R1 *Fusarium* strains infecting banana[12,31–33]. To better understand the role of gene duplications in the origin and diversification of ARs, we characterized gene duplications in *F. oxysporum* and classified them into four categories (dispersed, proximal, tandem, and segmental)

(**Fig. 6a**). Remarkably, most genes in strain II5 (8,831 genes) have been affected by gene duplication during their evolution (**Fig. 5b, Supplementary Fig. 6**). We observed similar results for other *F. oxypsorum* strains as well as for other *Fusarium* species; *F. verticillioides*, *F. graminearum,* and *F. solani* (syn. *Neocosmospora solani*), suggesting that duplications played a major role in *Fusarium* evolution (**Fig 6b**). Interestingly, we observed that segmental duplications occur specifically in ARs (72 out of 124 in II5) and sub-telomeres (48 out of 124 in II5) in all chromosome-level *F. oxysporum* assemblies (**Fig 6b**, **Supplementary Fig. 6**). In contrast to *F. oxysporum*, we did not identify segmental duplications in the species *F. verticillioides* and *F. graminearum* (**Fig. 6f**). Only in *F. solani,* which contains accessory chromosomes analogous to those observed in *F. oxysporum*[34], 139 segmental duplications were identified. The highest number of segmental duplications (2,887) is observed in *Fol*4287, carrying the largest accessory genome (19 Mb, 30% of the total genome size) (**Fig. 6g**) and 2,366 (82%) of the segmental duplications are in ARs. Interestingly, many segmental duplications are shared between accessory chromosomes 3 and 6 (**Supplementary Fig. 7**), as previously observed[12], indicating that accessory chromosomes evolved by inter-chromosomal segmental duplication that drive the expansion of ARs in *Fusarium*.

Duplications of AR genes significantly effect effector repertoires as 336 out of 669 predicted effectors in II5 evolved via duplications, nine by segmental duplications. Among the segmentally duplicated effectors is *SIX1*, an effector that is essential for full virulence to banana[35]. Three copies of *SIX1* (*SIX1a*,*b*,*c*) are present in the AR on chromosome 1 of the TR4 strains M1, 36102 and II5. *SIX1a* and *SIX1b* are part of a segmentally duplicated block (**Fig. 6c**), whereas *SIX1c* is a proximal duplication that

shares 82% amino acid sequence identity with *SIX1a* and 73% sequence identity with *SIX1b*. In contrast, the R1 strain CR1.1 contains only one copy of *SIX1*, resembling *SIX1c*, on accessory chromosome 12. The duplicated blocks with *SIX1* are interspersed with non-duplicated genes (**Fig. 6c**), indicating that over time these blocks further diverged by gene gains and losses. Interestingly, the segmentally duplicated *SIX1* blocks also share similarity to a region on chromosome 12 (**Fig. 6c**), yet no copy of *SIX1* is present, and thus *SIX1* is either lost or has been gained on chromosome 1 after the segmental duplication between chromosome 1 and 12 occurred. Segmental duplications therefore contribute to the evolution of virulence factors such as *SIX1*, which is essential for successful banana infection[35].

To estimate the relative timing of the duplication events, we used the synonymous substitution rate (dS) between duplicated gene pairs as a proxy of time. Dispersed duplicates had the highest average dS value (3.08), suggesting that these are the most ancient duplicates, while genes pairs that arose via segmental duplication evolved more recently, with a significantly lower average dS value of 0.61 (**Fig. 6d**). To assess if even more recent large-scale duplications are present in our global panel, we determined the read depth for all isolates sequenced with short reads against II5. Based on the genome-wide average read depth (~40x coverage), we determined that chromosome 12 has been entirely duplicated (read depth of ~80x coverage), while a section (position ~64-116 kb; ~160x coverage) occurs four times (**Fig. 6h**). This demonstrates that the II5 carries two (nearly) identical copies of chromosome 12 (i.e., aneuploidy), in addition to the previously observed segmental duplications (**Fig. 6h**). Moreover, the four partial copies of chromosome 12 possibly originate from an

additional duplication, underscoring the highly dynamic accessory chromosomes. It remains unclear if this duplicated stretch is attached to chromosome 12, or if it possibly gave rise to a small variant of the accessory chromosome. Remarkably, we didn't observe a threefold increase on chromosome 12, suggesting that the duplication happened multiple times. Interestingly, aneuploidy of chromosome 12 can also be observed in three additional TR4 strains from the Middle East, but not in the other 26 TR4 strains (**Fig. 6h**). Collectively, we demonstrate that the evolutionary dynamics of ARs in *F. oxysporum* is driven by recent segmental duplications and aneuploidy.

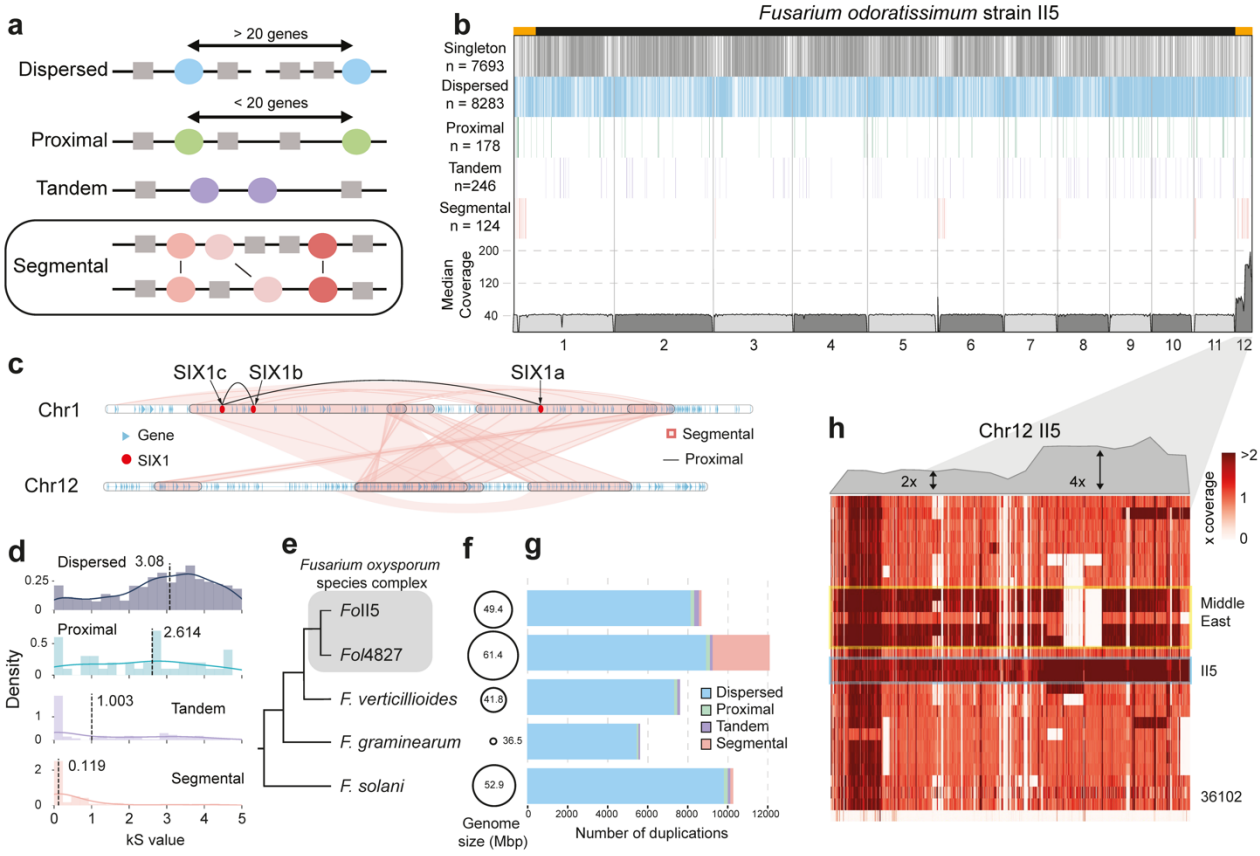


Figure 6 - Segmental duplications are involved in the expansion of accessory regions in *F. oxysporum*. a) Four different types of duplications are distinguished. **b**) Segmental duplications (pink) occur mainly in the accessory regions in strain II5 (TR4). Sub-telomeric regions have more dispersed duplications (blue) than the chromosomal arms. On average, the read coverage of II5 mapped onto the II5 reference genome is 40x, whereas chromosome 12 has a coverage between 80x and 160x, suggesting copy-

number variations. **c**) Genes encoding the virulence factor SIX1 (red) are part of a segmental duplication within chromosome 1. *SIX1a* and *SIX1b* are segmentally duplicated, and *SIX1c* is a proximal duplication sharing 82% amino acid sequence similarity to *SIX1a*. Light red blocks indicate segmental duplications, light red lines between genes (blue arrows) highlight those involved in the segment duplication. A region on chromosome 12 is similar to the *SIX1* regions on chromosome 1, yet *SIX1* is absent on chromosome 12. **d**) The age of duplications was estimated from the number of synonymous substitutions between duplicated gene pairs. Segmental duplications Segmental duplications occurred most recently (kS = 0.119). kS distributions differ significantly between all duplication types ($p < 0.05$, two sided Mann-Whitney U test). **e-f**) The schematic tree depicts the relationship of *Fusarium* species with their genome sizes. **g**) Extensive segmental duplications are only found in *F. oxysporum* and *F. solani* and are virtually absent in *F. verticillioides* and *F. graminearum*. **h)** Sequencing read coverage of strain II5 and three strains from the Middle East (JV11, ISR5, and JV14) show a two-fold increase relative to the genome-wide coverage. The other 26 strains do not show an increase in coverage, although smaller duplications are found throughout chromosome 12.

Discussion

Many plant pathogens evolved compartmentalized genomes with conserved core regions and variable accessory regions (ARs)[4,5]. Some members of the *Fusarium oxysporum* species complex are important plant pathogens and are well-known to have extensive ARs that can encompass entire chromosomes, which have been linked to pathogenicity towards specific hosts[8]. Here we use a pan-genomic framework to study the occurrence, composition, and evolution of ARs in a global collection of banana infecting *Fusarium* species. Strains carry up to 15 Mb of ARs that contain *in planta* induced effector genes as well as most *SIX* effectors. ARs are highly divergent between strains and races, and consequently, we cannot identify shared ARs or accessory chromosomes that can be linked to pathogenicity towards banana. We furthermore demonstrate that *F. oxysporum*, but also *F. solani*, evolve by extensive duplications and that ARs specifically are shaped by recent segmental duplications. Lastly, we uncover that an accessory chromosome in II5 underwent recent aneuploidy, suggesting that these processes drive the emergence and evolution of ARs in *F. oxysporum*.

Fusarium wilt of banana is a major threat to food security as banana is a major staple crop in tropical and subtropical regions[22,36]. Compared to R1 strains, which caused epidemics in Gros Michel bananas in the 1900s, TR4 strains emerged more recently (1967) and cause the ongoing FWB pandemic in Cavendish plantations[22]. TR4 strains are genetically highly similar, corroborating that a single clone that circumvents Cavendish resistance drives the ongoing pandemic[21,22]. By contrast, R1 strains are genetically diverse, which is possibly due to prolonged co-evolution of R1 strains with a plethora of genetically diverse banana varieties in the center of origin in Southeast

Asia[37]. Extensive genetic diversity among *Fusarium* strains that infect the same host is not uncommon[15,18,38], yet their ARs and effector profiles are typically remarkably similar[11,18,30]. We, however, observe that the ARs and effector profile of R1 strains show extensive variation, underscoring their diversity and possibly suggesting that different mechanisms contribute to disease in banana[15,30]. *Fusarium* strains that infect strawberry have two distinct ARs, one that contributes to yellowing and the other to wilting[15]. Quantitative differences in banana corm-discoloration have been noted between R1 strains[39], and our results suggest that this could be caused differences in effector profiles, yet further functional assays are needed to elucidate the link between specific effectors and phenotypic variation.

We identify abundant and recent segmental duplications in *F. oxysporum* ARs, as well as aneuploidy of chromosome 12 in II5. Previous studies have observed large-scale duplications and deletions in the ARs of *F. oxysporum* during *in vitro* chromosome transfer[32,33], yet our results show that duplication of ARs also occur in *Fusarium* strains sampled directly from infected banana plants. Although variations in chromosome number typically comes with a fitness cost[40], it can also provide genetic variation necessary for adaptation, for example increasing virulence or fungicide resistance[41,42]. We speculate that the observed large-scale duplications play a major role in the origin and evolution of ARs in *Fusarium*. In the wheat pathogen *Zymoseptoria tritici*, chromosomal duplication, together with breakage and fusion, shape the evolutionary dynamics of accessory regions[43,44]. Similar dynamics of chromosome duplication, breakage, and fusion may underly the recent segmental duplications observed in *Fusarium*. For example, the fusion of a (partially) duplicated chromosome 12 with the

arm of chromosome 1 in II5 could explain the observed sequence similarity between these chromosomes. Chromosome dynamics in *F. oxysporum* require further elucidation and, so far, the mechanisms leading to chromosomal duplications are unclear; they might arise from non-disjunction during mitosis, from incomplete chromosome loss following heterokaryon formation[45,46], or through consecutive horizontal chromosome transfers when strains acquire the identical chromosome several times.

In contrast to the structural variation in the ARs, the core chromosomes in *Fusarium* are remarkably stable. Although ARs are likely more tolerant to structural variation as they encode fewer essential genes, additional processes can influence genome stability. For example, the presence of histone modifications such as trimethylation of lysine 27 on histone 3 (H3K27me3), a histone modifications often enriched in ARs[5,11,47] in some fungal plant pathogens, has been linked to genome instability[47,48]. H3K27me3 also occurs in the ARs of *F. oxysporum* infecting tomato (*Fol*4287), but can also be found in smaller core chromosomes ('fast-core' chromosomes, 9 – 11[11]) that do not seem to undergo segmental duplications or aneuploidy, and thus additional factors likely influence chromosome stability in *F. oxysporum.*

Extensive duplications similarly occur in other plant pathogens[49–52], and are thought to be important drives in co-evolutionary arms races with their hosts. Understanding the evolution of ARs in *Fusarium* can facilitate the discovery of new effector genes and provide insights into effector diversification. Knowledge of effector profiles is crucial for designing effective disease control strategies[1], and supports the identification of durable crop resistances[2].

479

Methods

Collection of *Fusarium* strains infecting banana

We analyzed a diverse set of 69 *Fusarium* strains infecting banana (**Supplementary Table 1**). *Fusarium* strains were isolated from different banana growing regions worldwide directly from the pseudostem of symptomatic banana plants. Here, 35 *Fusarium* strains were newly sequenced and we included 34 publicly available whole-genome sequencing datasets (based on short-read sequencing approaches)[24,26,55–60]. To obtain high-quality reference assemblies, seven strains were sequenced with Oxford Nanopore Technology.

Whole-genome sequencing

Fusarium isolates were cultured on potato dextrose agar (PDA) and conidia were transferred to liquid media for incubation at 25°C. Fungal material was isolated from liquid culture for genomic DNA extraction. The exact DNA isolation protocol per strain can be found in Supplementary Material 1.1 to 1.3. To obtain long-read, whole-genome sequencing data, seven strain (36102, II5, CR1.1, C058, C135, C081, and M1) were sequenced using Oxford Nanopore Technologies (Oxford, UK). An R9.4.1 flow cell was loaded and run for 24h. Base calling was performed using Guppy (version 3.1.5; C058, C135, C081, and M1) and MiniKNOW (version 3.4.6; II5, 36102, and CR1.1). The other *Fusarium* strains were sequenced by Beijing Genomic Institute (BGI) using Illumina paired-end sequencing, 31 were sequenced during this study and 38 were obtained during other studies [21,24,26,55,57–60].

Gene expression analysis

Cavendish 'Grand Naine' banana plants were grown in a greenhouse compartment (28 ± 2°C, 16 h light, and ~85% relativity humidity) and acclimatized under plastic for 2 weeks to maintain high humidity. The roots of ~2.5 months old plants were dip inoculated with 10ˆ6 spores per ml. RNA was isolated from the roots of inoculated plants at 8 days post inoculation and from mycelium grown on PDA medium. Samples were ground and RNA extraction was performed using the Maxwell Plant RNA kit (Promega) following manufacturer's instructions. The quality of the RNA samples was tested by agarose electrophoresis and quantified by Nanodrop (ThermoFisher), and subsequently sent to KeyGene (The Netherlands) for sequencing. Quantifications of gene transcripts from sequenced RNA-seq reads was performed using Kallisto[61].

Genome assembly and annotation

We obtained chromosome-level *de novo* genome assemblies for seven Fusarium isolates sequenced by ONT long-read technology (Oxford Nanopore Technologies, Oxford, UK). Adapter sequences were removed with Porechop (version 0.2.4, default settings; https://github.com/rrwick/Porechop/tree/master) and reads were self-corrected, trimmed, and assembled using Canu[62] (version 1.8). Genome assemblies were polished using ONT raw reads with Racon[63] (version 1.5.0) and with high-quality Illumina short read using two rounds of Pilon[64] (version 1.24). Contigs were scaffolded, if needed, using longstitch[65] (version 1.0.2).

Fusarium strains sequenced solely with short-read sequencing technology were *de novo* assembled with Spades[66] (version 3.13.0, with default settings). Contigs shorter than 500 bp were subsequently removed. To assess the assembly quality, the genome assemblies were analyzed with QUAST[67] (version 5.0.2), and genome completeness was estimated based on the presence of conserved single-copy genes using BUSCO[68] (version 5.1.2) with the *hypocreales* odb10 database.

We performed gene annotation of 69 *Fusarium* genome assemblies using Funannotate[69] (version 1.8.9) with the *Fusarium oxysporum* f. sp. *lycopersici* 4287 Uniprot proteins, uniport database, and previously predicted effectors[18] as external evidence. Genomes were masked using RepeatMasker (version 4.1.1) following repeat prediction with RepeatModeler (http://www.repeatmasker.org). For *Fusarium* strains II5, CR1.1, and 36102, RNA-seq data was used to guide gene model prediction, and *ab initio* training parameters estimated using RNA-seq data of II5 were re-used to support the annotation in other genome assemblies. Effectors genes were predicted based on their association with Miniature IMPala (MIMP) elements in *Fusarium*, as described by van Dam et al 2016[18]. Briefly, ORFs were predicted 5 kb up- and downstream to MIMPs (‘AGT[GA][GA]G[GAT][TGC]GCAA[TAG]AA’) and these were refined using AUGUSTUS (version 3.3.3)[70], with the *Fusarium* species specified. In the set of predicted MIMP-genes, secreted proteins were predicted using SignalP (version 5) and subsequently effectors were predicted using EffectorP (version 3.0). The predicted effectors were added to the Funannotate annotation using AGAT (version 0.8.0.)

To recover potentially missing genes in the gene annotations, we used the comparative annotation toolkit[71] (version 2). Cactus[72] (version 2.0.1) was used to create

whole-genome alignments of all 69 genomes, using the Minigraph-Cactus approach[73]. Consecutively, the CAT was used to carry-over gene annotations based pn funannotate gene predictions with the RNA-seq data (with Augustus settings pre trained on II5, --global-near-best 1, --augustus-utr-off, --augustus-cgp, --maf-chunksize 550000 --maf-overlap 50000), which added approximately 200 genes per genome assembly.

Core gene phylogeny

The phylogenetic relationship between *Fusarium* strains infecting banana together with 55 publicly available non-banana infecting *Fusarium* strains was inferred from 3,811 conserved single-copy BUSCO genes. Protein sequences were aligned using maftf[74], (version 7.453) and a maximum-likelihood phylogeny was determine using IQtree[75] (version 1.6; -m MFP+MERGE, -B 1000 with 1,000 ultra-fast bootstrap), *Fusarium verticilliodes* (ASM14955v1[76]) was included as an outgroup.

Pangenome analysis

We detected conserved and accessory regions in the set of *Fusarium* isolates by performing all-vs-all whole-genome alignments with nucmer[77] (version 3.1; --max-match). We removed small alignments with delta-filter (-l 5000) and only retained the best match per alignment (-1). Genomic regions present in all 69 *Fusarium* isolates were considered core, regions covered by less than 80% were considered accessory, and regions between these values were considered softcore. Pairwise similarity

between ARs was calculated using custom python scripts and alignments were visualized with PyGenomeViz (https://github.com/moshi4/pyGenomeViz) and Circlize[78].

To identify orthologous groups, we used Broccoli[79] (version 1.1) on the predicted protein-coding genes from the 69 *Fusarium* strains infecting banana; per gene, only the longest transcript was included. OGs present in all isolates were considered core, OGs found in more than 55 genomes (80%) were considered softcore, and genes found in less than 55 of the genomes were considered accessory.

We determined the number of non-synonymous and synonymous substitutions (dN/dS) per OG. All genes in an OG were aligned using mafft[74] (version 7.490) and non-synonymous and synonymous substitutions per gene pair in the OG were inferred based on a codon-guided nucleotide alignment, created by pal2nal[80] (version 14), with codeML (from paml[81], version 4.9). Differences between groups of genes were further analyzed using custom python scripts and compared using a two-sided Mann-Whitney U test in SciPy[82] (version 1.10.1).

Estimation of gene age

Orthofinder[83] (version 2.5.4) was used to detect the presence of homologs in 308 phylogenetically distinct fungal species from the joint genome institute; per fungal family, one fungal genome assembly was randomly selected (**Supplementary table 4**). Homologs that occur at least once outside the phylum Ascomycota are considered *old* homologs, homologs occurring only within Ascomycota are further subdivided into *Hypocreales* and *Fusarium*.

Detection of duplication events

We detected gene duplication events by self-comparisons of the translated protein sequences (only the longest isoform) of the chromosome-level genome assemblies of *Fusarium* infecting banana, as well as *Fusarium oxysporum* f. sp. *lycopersici* (strain *Fol*4287[12]; GCF_000149955.1, *Fusarium* graminearum[76] (GCF_000240135.3), *Fusarium verticilliodes*[76] (GCF_000149555.1), and *Fusarium solani*[84] (GCF_020744495.1). Per *Fusarium* strain, the predicted protein sequences were compared with BLAST and the best five hits were retained after filtering for an e-value $1e^{-10}$ and query coverage larger than 50%. Then, BLAST alignments were used to detect collinear duplicated blocks with MCscanX[85]; blocks of 5 genes were considered co-linear and matches were filtered with the parameters match-score 50 and e-value 10^{-5}. Subsequently, we classified duplicated genes into dispersed, proximal, tandem, or segmental duplications with MCscanX's 'duplicate_gene_classifier' script.

Copy number variations were identified by mapping short-read, whole-genome sequencing datasets against the chromosome-level genome assembly of *Fusarium* strain II5 using BWA-mem[86] (version 0.7.17). We determined the sequencing coverage over the reference genome with bedtools (version 2.30), which was visualized Wgscoverageplotter from jvarkit[87] (version 1.1.0). Single nucleotide polymorphisms were called using GATK (version 4.0.2; -ploidy 1) and filtered according to GATK best practices[88].

Acknowledgements

ACW, GHJK, CAG, GNT, RS and JD were supported by the Bill and Melinda Gates Foundation, grant number AG - 4425. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank the Research Support Group (RSG) at KeyGene (Wageningen, The Netherlands), which provided DNA isolation, library preparation and long-read sequencing for three of the strains. Banana research at Wageningen University has been supported by the Dutch Dioraphte Foundation, grant number 20 04 04 02.

Author Contribution

ACW collected data and performed the analyses. EMP, CAG, ECC, RH, AW, HS, FGB, NM, NO, EHS and GNT collected data. JD, EHS, AF, KN and XSK contributed to the analyses. ACW and wrote the manuscript with input from HJGM, GHJK, and MFS. All authors contributed to writing and editing the manuscript. GHJK, RS, and MFS conceived and supervised the project.

Data Availability Statement

The availability of strains analyzed in this study is reported in **Supplementary table 1**. All strains sequenced during this study are deposited on NCBI PRJNA979975.

References

1. Möller, M. & Stukenbrock, E. H. Evolution and Genome Architecture in Fungal Plant Pathogens. *Nat Rev Microbiol* **15**, 756–771 (2017).
2. Cook, D. E., Mesarich, C. H. & Thomma, B. P. H. J. Understanding Plant Immunity as a Surveillance System to Detect Invasion. *Annu Rev Phytopathol* **53**, 541–563 (2015).
3. Rovenich, H., Boshoven, J. C. & Thomma, B. P. H. J. Filamentous Pathogen Effector Functions: Of Pathogens, Hosts and Microbiomes. *Curr Opin Plant Biol* **20**, 96–103 (2014).
4. Dong, S., Raffaele, S. & Kamoun, S. The Two-Speed Genomes of Filamentous Pathogens: Waltz with Plants. *Curr Opin in Genetics & Development* **35**, 57–65 (2015).
5. Torres, D. E., Oggenfuss, U., Croll, D. & Seidl, M. F. Genome Evolution in Fungal Plant Pathogens: Looking Beyond the Two-Speed Genome Model. *Fungal Biology Reviews* **34**, 136–143 (2020).
6. Raffaele, S. & Kamoun, S. Genome Evolution in Filamentous Plant Pathogens: Why Bigger Can Be Better. *Nat Rev Microbiol* **10**, 417–430 (2012).
7. Seidl, M. F. & Thomma, B. P. H. J. Transposable Elements Direct The Coevolution between Plants and Microbes. *Trends Genet* **33**, 842–851 (2017).
8. Sánchez-Vallet, A. *et al.* The Genome Biology of Effector Gene Evolution in Filamentous Plant Pathogens. *Annu Rev Phytopathol* **56**, 21–40 (2018).

9. Maryani, N. *et al.* Phylogeny and Genetic Diversity of the Banana Fusarium Wilt Pathogen *Fusarium oxysporum* f. sp. *cubense* in the Indonesian Centre of Origin. *Studies in Mycology* **92**, 155–194 (2019).
10. Edel-Hermann, V. & Lecomte, C. Current Status of *Fusarium oxysporum* Formae Speciales and Races. *Phytopathology* **109**, 512–530 (2019).
11. Fokkens, L. *et al.* The Multi-Speed Genome of *Fusarium oxysporum* Reveals Association of Histone Modifications with Sequence Divergence and Footprints of Past Horizontal Chromosome Transfer Events. 465070 Preprint at https://doi.org/10.1101/465070 (2018).
12. Ma, L.-J. *et al.* Comparative Genomics Reveals Mobile Pathogenicity Chromosomes in Fusarium. *Nature* **464**, 367–373 (2010).
13. van Dam, P. *et al.* A Mobile Pathogenicity Chromosome in *Fusarium oxysporum* for Infection of Multiple Cucurbit Species. *Sci Rep* **7**, 9042 (2017).
14. Zhang, Y. *et al.* The Genome of Opportunistic Fungal Pathogen *Fusarium oxysporum* Carries a Unique Set of Lineage-Specific Chromosomes. *Commun Biol* **3**, 1–12 (2020).
15. Henry, P. M. *et al.* Horizontal Chromosome Transfer and Independent Evolution Drive Diversification in *Fusarium oxysporum* f. sp. *Fragariae*. *New Phytologist* **230**, 327–340 (2021).
16. Ayukawa, Y. *et al.* A Pair of Effectors Encoded on a Conditionally Dispensable Chromosome of *Fusarium oxysporum* Suppress Host-Specific Immunity. *Commun Biol* **4**, 1–12 (2021).

17. Rep, M. *et al.* Mass Spectrometric Identification of Isoforms of Pr Proteins in Xylem Sap of Fungus-Infected Tomato. *Plant Physiol* **130**, 904–917 (2002).
18. van Dam, P. *et al.* Effector Profiles Distinguish *formae speciales* of *Fusarium oxysporum*. *Environmental Microbiology* **18**, 4087–4102 (2016).
19. FAO. *Fruit and Vegetables – Your Dietary Essentials: The International Year of Fruits and Vegetables, 2021, Background Paper*. (FAO, 2020). doi:10.4060/cb2395en.
20. Ploetz, R. C. Management of Fusarium Wilt of Banana: A Review with Special Reference to Tropical Race 4. *Crop Protection* **73**, 7–15 (2015).
21. Ordonez, N. *et al.* Worse Comes to Worst: Bananas and Panama Disease—When Plant and Pathogen Clones Meet. *PLOS Pathogens* **11**, e1005197 (2015).
22. van Westerhoven, A. C. van, Meijer, H. J. G., Seidl, M. F. & Kema, G. H. J. Uncontained Spread of Fusarium Wilt of Banana Threatens African Food Security. *PLOS Pathogens* **18**, e1010769 (2022).
23. Yang, H., Yu, H. & Ma, L.-J. Accessory Chromosomes in *Fusarium oxysporum*. *Phytopathology®* **110**, 1488–1496 (2020).
24. Warmington, R. J. *et al.* High-Quality Draft Genome Sequence of the Causal Agent of the Current Panama Disease Epidemic. *Microbiol Resour Announc* **8**, e00904-19 (2019).
25. O'Donnell, K., Kistler, H. C., Cigelnik, E. & Ploetz, R. C. Multiple Evolutionary Origins of the Fungus Causing Panama Disease of Banana: Concordant Evidence from Nuclear and Mitochondrial Gene Genealogies. *Proceedings of the National Academy of Sciences* **95**, 2044–2049 (1998).

26. Martinez de la Parte, E. Genetic Diversity of Fusarium Infecting Banana in Cuba. *In Progress*.
27. Maryani, N. Global Diversity of Fusarium Wilt Infecting Banana. *In Progress*.
28. Galagan, J. E. & Selker, E. U. RIP: The Evolutionary Cost of Genome Defense. *Trends in Genetics* **20**, 417–423 (2004).
29. Wolf, Y. I., Novichkov, P. S., Karev, G. P., Koonin, E. V. & Lipman, D. J. The Universal Distribution of Evolutionary Rates of Genes and Distinct Characteristics of Eukaryotic Genes of Different Apparent Ages. *Proceedings of the National Academy of Sciences* **106**, 7273–7280 (2009).
30. Brenes Guallar, M. A., Fokkens, L., Rep, M., Berke, L. & van Dam, P. *Fusarium oxysporum* Effector Clustering Version 2: An Updated Pipeline to Infer Host Range. *Frontiers in Plant Science* **13**, (2022).
31. Kistler, H. C., Benny, U., Boehm, E. W. & Katan, T. Genetic Duplication in *Fusarium oxysporum*. *Curr Genet* **28**, 173–176 (1995).
32. Li, J., Fokkens, L., Conneely, L. J. & Rep, M. Partial Pathogenicity Chromosomes in *Fusarium oxysporum* Are Sufficient to Cause Disease and Can Be Horizontally Transferred. *Environmental Microbiology* **22**, 4985–5004 (2020).
33. Vlaardingerbroek, I., Beerens, B., Schmidt, S. M., Cornelissen, B. J. C. & Rep, M. Dispensable Chromosomes in *Fusarium oxysporum* f. sp. *lycopersici*. *Molecular Plant Pathology* **17**, 1455–1466 (2016).
34. Coleman, J. J. *et al.* The Genome of *Nectria haematococca*: Contribution of Supernumerary Chromosomes to Gene Expansion. *PLOS Genetics* **5**, e1000618 (2009).

35. Widinugraheni, S. *et al.* A SIX1 homolog in Fusarium oxysporum f.sp. cubense tropical race 4 contributes to virulence towards Cavendish banana. *PLOS ONE* **13**, e0205896 (2018).
36. Drenth, A. & Kema, G. The Vulnerability of Bananas to Globally Emerging Disease Threats. *Phytopathology®* **111**, 2146–2161 (2021).
37. Perrier, X. *et al.* Multidisciplinary Perspectives on Banana (Musa spp.) Domestication. *Proceedings of the National Academy of Sciences* **108**, 11311–11318 (2011).
38. Baayen, R. P. *et al.* 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 (2000).
39. Garcia-Bastidas, F. Panama Disease in Banana: Spread, Screens and Genes. (Wageningen University, 2019).
40. Todd, R. T., Forche, A. & Selmecki, A. Ploidy Variation in Fungi: Polyploidy, Aneuploidy, and Genome Evolution. *Microbiology Spectrum* **5**, 5.4.09 (2017).
41. Sionov, E., Lee, H., Chang, Y. C. & Kwon-Chung, K. J. *Cryptococcus neoformans* Overcomes Stress of Azole Drugs by Formation of Disomy in Specific Multiple Chromosomes. *PLOS Pathogens* **6**, e1000848 (2010).
42. Ropars, J. *et al.* Gene Flow Contributes to Diversification of the Major Fungal Pathogen *Candida albicans*. *Nat Commun* **9**, 2253 (2018).
43. Croll, D., Zala, M. & McDonald, B. A. Breakage-fusion-bridge Cycles and Large Insertions Contribute to the Rapid Evolution of Accessory Chromosomes in a Fungal Pathogen. *PLOS Genetics* **9**, e1003567 (2013).

44. Möller, M., Habig, M., Freitag, M. & Stukenbrock, E. H. Extraordinary Genome Instability and Widespread Chromosome Rearrangements During Vegetative Growth. *Genetics* **210**, 517–529 (2018).
45. Harrison, B. D. *et al.* A Tetraploid Intermediate Precedes Aneuploid Formation in Yeasts Exposed to Fluconazole. *PLOS Biology* **12**, e1001815 (2014).
46. Shahi, S., Beerens, B., Bosch, M., Linmans, J. & Rep, M. Nuclear Dynamics and Genetic Rearrangement in Heterokaryotic Colonies of *Fusarium oxysporum*. *Fungal Genet Biol* **91**, 20–31 (2016).
47. Möller, M. *et al.* Destabilization of Chromosome Structure by Histone H3 Lysine 27 Methylation. *PLOS Genetics* **15**, e1008093 (2019).
48. Seidl, M. F., Cook, D. E. & Thomma, B. P. H. J. Chromatin Biology Impacts Adaptive Evolution of Filamentous Plant Pathogens. *PLOS Pathogens* **12**, e1005920 (2016).
49. Dutheil, J. Y. *et al.* A Tale of Genome Compartmentalization: The Evolution of Virulence Clusters in Smut Fungi. *Genome Biol Evol* **8**, 681–704 (2016).
50. Müller, M. C. *et al.* A Chromosome-Scale Genome Assembly Reveals a Highly Dynamic Effector Repertoire of Wheat Powdery Mildew. *New Phytologist* **221**, 2176–2189 (2019).
51. Wacker, T. *et al.* Two-Speed Genome Evolution Drives Pathogenicity in Fungal Pathogens of Animals. *Proceedings of the National Academy of Sciences* **120**, e2212633120 (2023).
52. Wyka, S. A. *et al.* Whole-Genome Comparisons of Ergot Fungi Reveals the Divergence and Evolution of Species within the Genus Claviceps Are the Result of

Varying Mechanisms Driving Genome Evolution and Host Range Expansion. *Genome Biology and Evolution* **13**, evaa267 (2021).

53. Latorre, S. M. *et al.* Genomic surveillance uncovers a pandemic clonal lineage of the wheat blast fungus. *PLOS Biology* **21**, e3002052 (2023).
54. Vleeshouwers, V. G. A. A. *et al.* Effector Genomics Accelerates Discovery and Functional Profiling of Potato Disease Resistance and Phytophthora Infestans Avirulence Genes. *PLOS ONE* **3**, e2875 (2008).
55. Garcia-Bastidas, F. *et al.* First Report of Fusarium Wilt Tropical Race 4 in Cavendish Bananas Caused by *Fusarium odoratissimum* in Colombia. *Plant Disease* **104**, 994–994 (2020).
56. García-Bastidas, F. A. *et al.* First Report of *Fusarium oxysporum* f. sp. *cubense* Tropical Race 4 Associated with Panama Disease of Banana outside Southeast Asia. *Plant Disease* **98**, 694–694 (2014).
57. Guo, L. *et al.* Genome and Transcriptome Analysis of the Fungal Pathogen *Fusarium oxysporum* f. sp. *cubense* Causing Banana Vascular Wilt Disease. *PLOS ONE* **9**, e95543 (2014).
58. Maymon, M., Sela, N., Shpatz, U., Galpaz, N. & Freeman, S. The Origin and Current Situation of *Fusarium oxysporum* f. sp.*cubense* Tropical Race 4 in Israel and the Middle East. *Sci Rep* **10**, 1590 (2020).
59. van Westerhoven, A. C. *et al.* Dissemination of Fusarium Wilt of Banana in Mozambique Caused by *Fusarium odoratissimum* Tropical Race 4. *Plant Disease* (2022) doi:10.1094/PDIS-07-22-1576-SC.

60. Zheng, S.-J. *et al.* New Geographical Insights of the Latest Expansion of *Fusarium oxysporum* f.sp. *cubense* Tropical Race 4 Into the Greater Mekong Subregion. *Frontiers in Plant Science* **9**, (2018).
61. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-Optimal Probabilistic RNA-Seq Quantification. *Nat Biotechnol* **34**, 525–527 (2016).
62. Koren, S. *et al.* Canu: Scalable and Accurate Long-Read Assembly Via Adaptive K-Mer Weighting and Repeat Separation. *Genome Res.* **27**, 722–736 (2017).
63. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and Accurate De Novo Genome Assembly from Long Uncorrected Reads. *Genome Res* **27**, 737–746 (2017).
64. Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE* **9**, e112963 (2014).
65. Coombe, L. *et al.* Longstitch: High-Quality Genome Assembly Correction and Scaffolding Using Long Reads. *BMC Bioinformatics* **22**, 534 (2021).
66. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* **19**, 455–477 (2012).
67. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
68. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
69. Palmer, J. & Stajich, J. nextgenusfs/funannotate: funannotate v1.5.3. (2019) doi:10.5281/zenodo.2604804.

70. Stanke, M. & Morgenstern, B. AUGUSTUS: A Web Server for Gene Prediction in Eukaryotes That Allows User-Defined Constraints. *Nucleic Acids Res* **33**, W465–W467 (2005).
71. Fiddes, I. T. *et al.* Comparative Annotation Toolkit (CAT)—Simultaneous Clade and Personal Genome Annotation. *Genome Res* **28**, 1029–1038 (2018).
72. Armstrong, J. *et al.* Progressive Cactus Is a Multiple-Genome Aligner for the Thousand-Genome Era. *Nature* **587**, 246–251 (2020).
73. Hickey, G. *et al.* Pangenome Graph Construction from Genome Alignments with Minigraph-Cactus. *Nat Biotechnol* 1–11 (2023) doi:10.1038/s41587-023-01793-w.
74. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. Mafft: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform. *Nucleic Acids Research* **30**, 3059–3066 (2002).
75. Chernomor, O., von Haeseler, A. & Minh, B. Q. Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Systematic Biology* **65**, 997–1008 (2016).
76. Cuomo, C. A. *et al.* The *Fusarium graminearum* Genome Reveals a Link Between Localized Polymorphism and Pathogen Specialization. *Science* **317**, 1400–1402 (2007).
77. Kurtz, S. *et al.* Versatile and Open Software for Comparing Large Genomes. *Genome Biol* **5**, 1–9 (2004).
78. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. Circlize Implements and Enhances Circular Visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).

79. Derelle, R., Philippe, H. & Colbourne, J. K. Broccoli: Combining Phylogenetic and Network Analyses for Orthology Assignment. *Molecular Biology and Evolution* **37**, 3389–3396 (2020).

80. Suyama, M., Torrents, D. & Bork, P. Pal2nal: Robust Conversion of Protein Sequence Alignments into the Corresponding Codon Alignments. *Nucleic Acids Res* **34**, W609-612 (2006).

81. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* **24**, 1586–1591 (2007).

82. Virtanen, P. *et al.* Scipy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat Methods* **17**, 261–272 (2020).

83. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics. *Genome Biol* **20**, 1–14 (2019).

84. Mesny, F. *et al.* Genetic Determinants of Endophytism in the Arabidopsis Root Mycobiome. *Nat Commun* **12**, 7227 (2021).

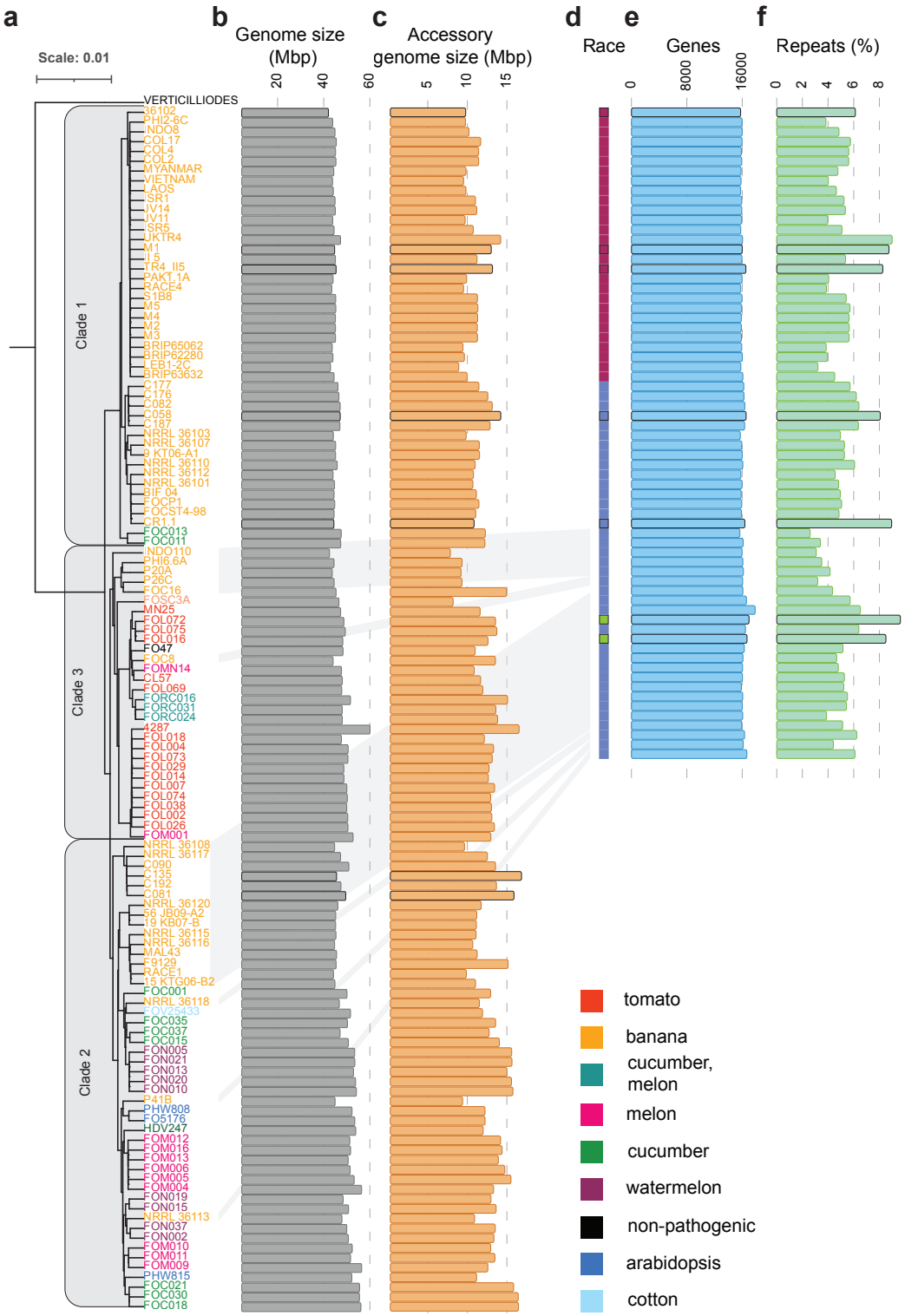
85. Wang, Y. *et al.* MCscanX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity. *Nucleic Acids Research* **40**, e49 (2012).

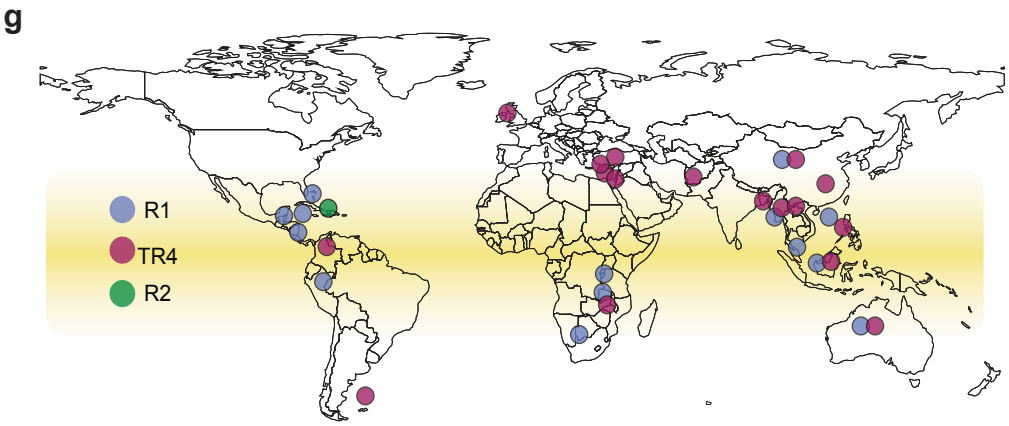
86. Li, H. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. (2013) doi:10.6084/M9.FIGSHARE.963153.V1.

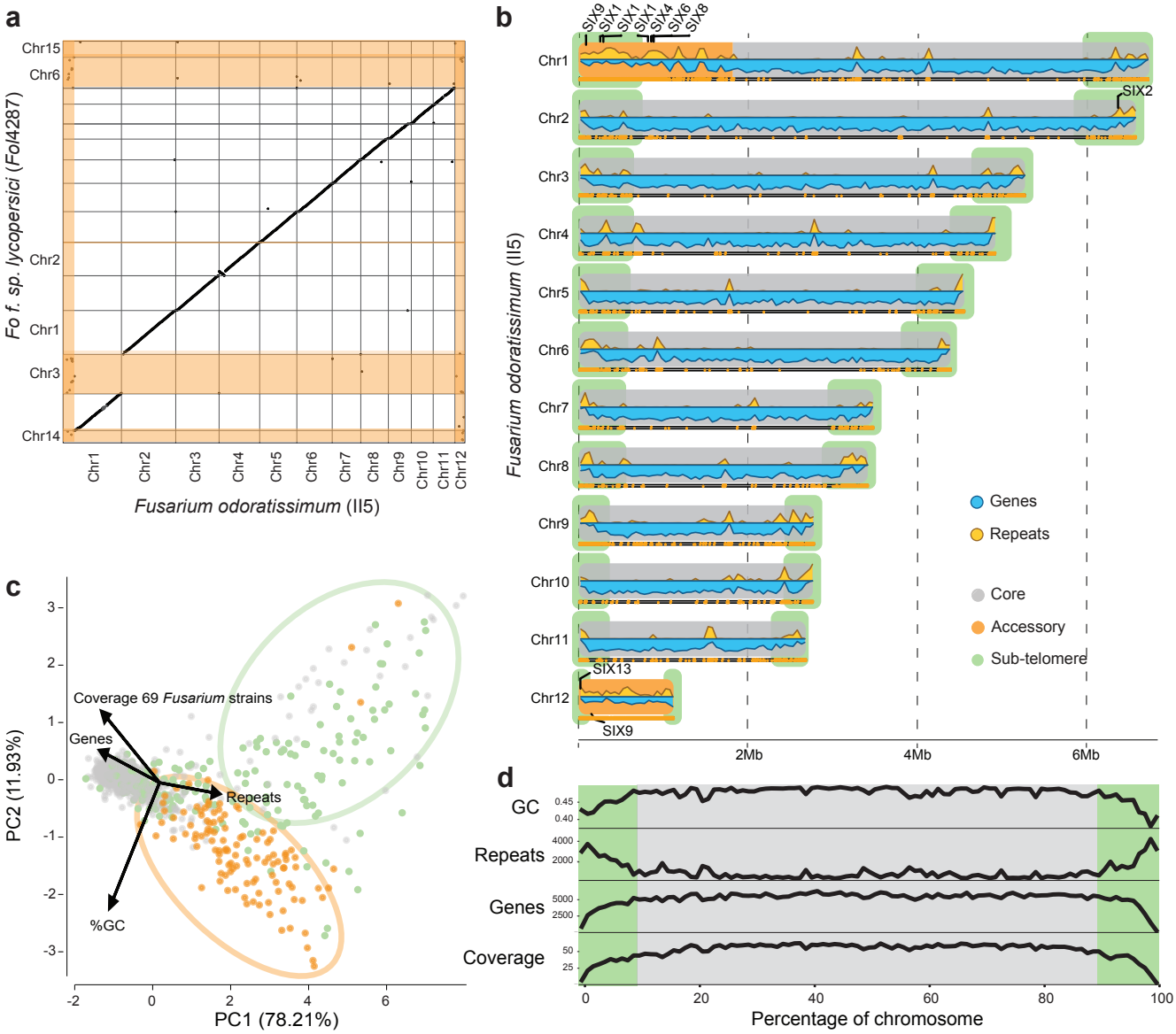
87. Lindenbaum, P. JVarkit: Java-Based Utilities for Bioinformatics. (2015) doi:10.6084/m9.figshare.1425030.v1.

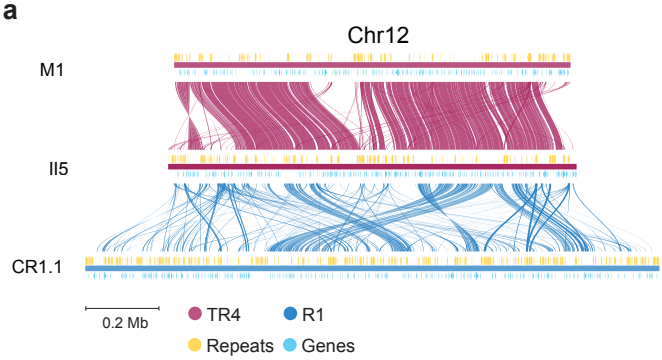
88. Van der Auwera, G. A. *et al.* From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics* **43**, 11.10.1-11.10.33 (2013).

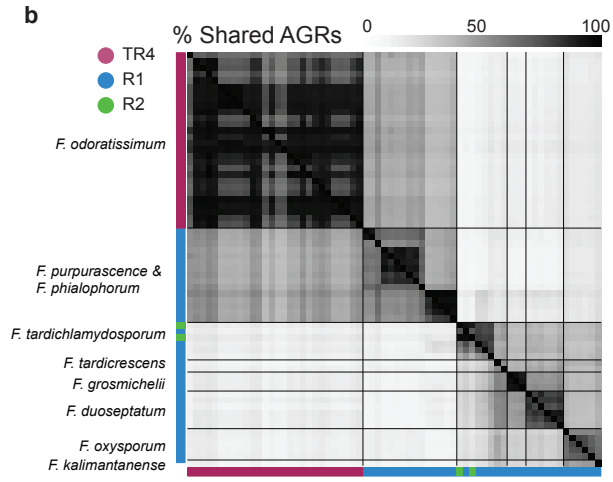
853

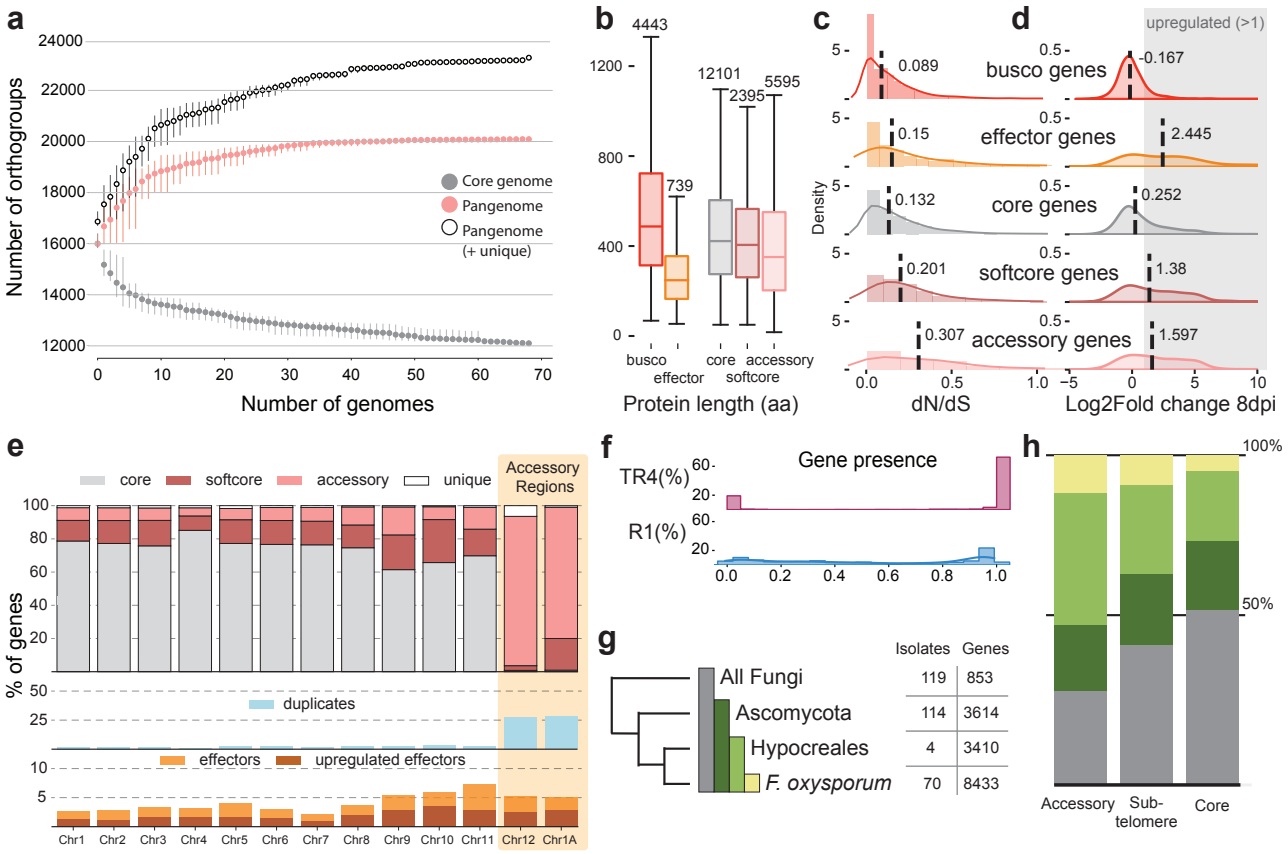

a
Scale: 0.01
b
Genome size (Mbp)
20
40
60
c
Accessory genome size (Mbp)
5
10
15
d
Race
e
Genes
0
8000
16000
f
Repeats (%)
0
2
4
6
8
Clade 1
Clade 3
Clade 2
tomato
banana
cucumber, melon
melon
cucumber
watermelon
non-pathogenic
arabidopsis
cotton

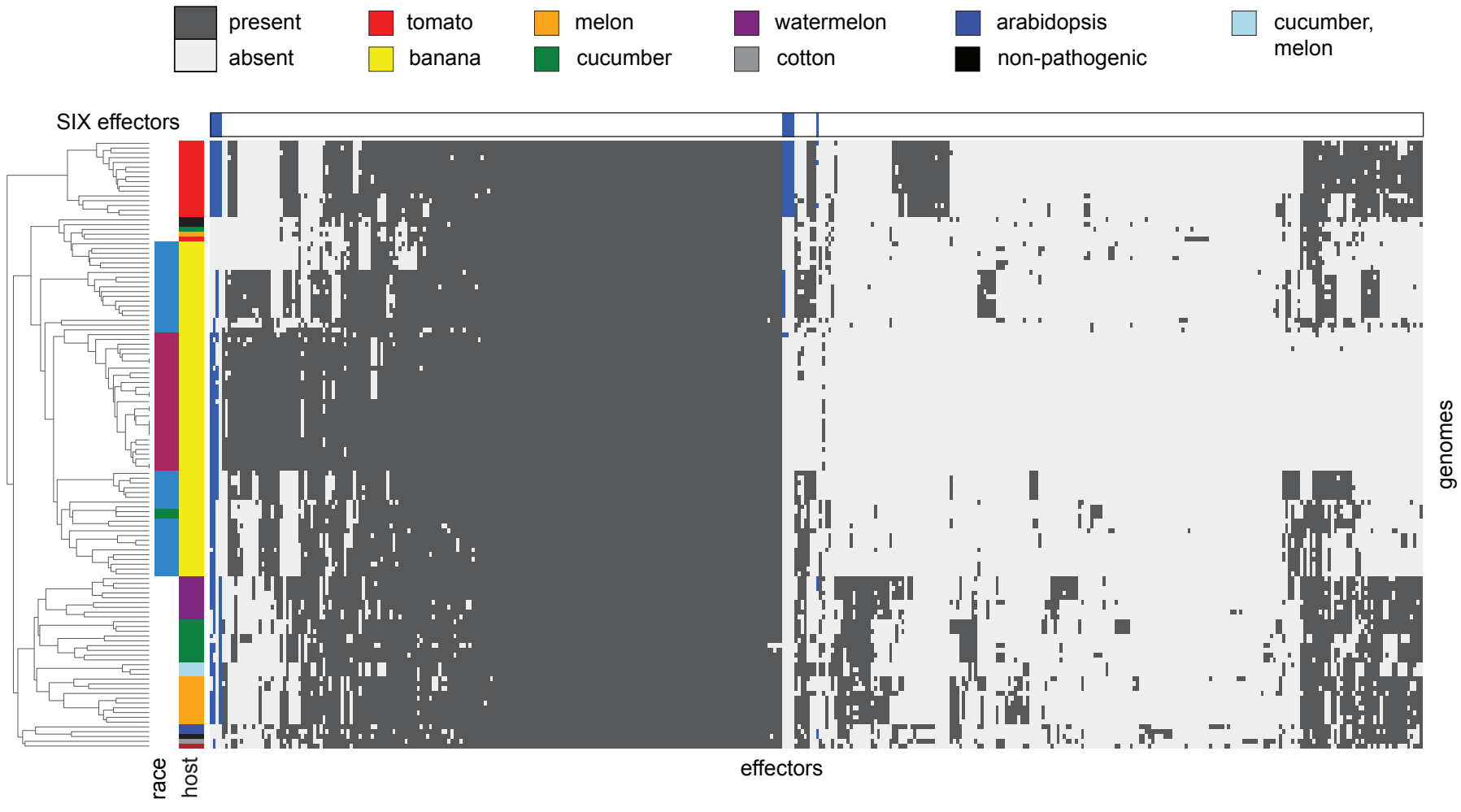

g
R1
TR4
R2

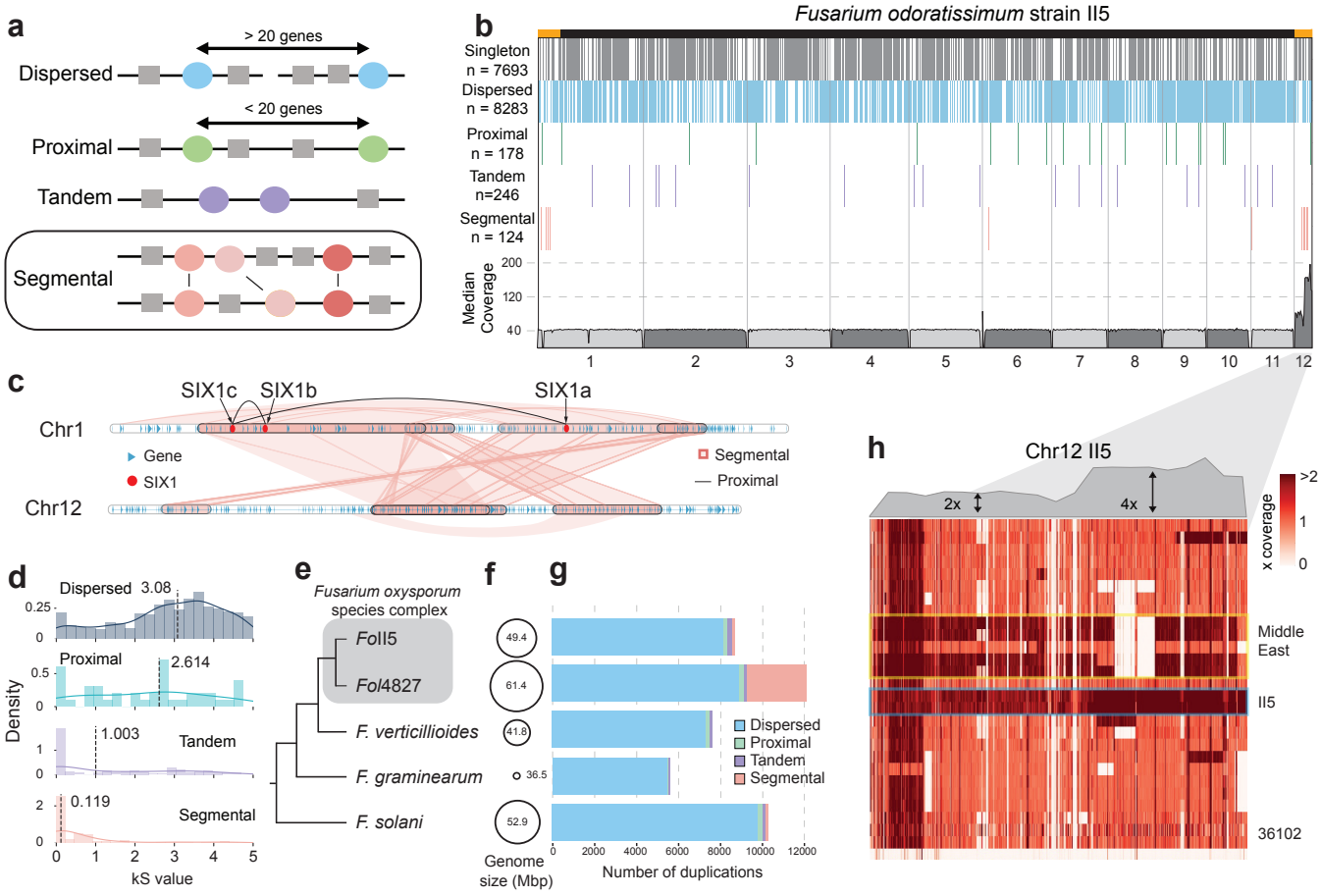

a
Chr15
Chr6
Chr2
Chr1
Chr3
Chr14
Fo f. sp. lycopersici (Fol4287)
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Fusarium odoratissimum (II5)
b
SIX9
SIX1
SIX1
SIX1
SIX4
SIX6
SIX8
SIX2
SIX13
SIX9
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Fusarium odoratissimum (II5)
2Mb
4Mb
6Mb
Genes
Repeats
Core
Accessory
Sub-telomere
c
Coverage 69 Fusarium strains
Genes
Repeats
%GC
PC2 (11.93%)
PC1 (78.21%)
d
GC
Repeats
Genes
Coverage
Percentage of chromosome


a
Chr12
M1
II5
CR1.1
0.2 Mb
TR4
R1
Repeats
Genes


b
% Shared AGRs
0
50
100
TR4
R1
R2
F. odoratissimum
F. purpurascence &
F. phialophorum
F. tardichlamydosporum
F. tardicrescens
F. grosmichelii
F. duoseptatum
F. oxysporum
F. kalimantanense


a
Number of orthogroups
Number of genomes
Core genome
Pangenome
Pangenome (+ unique)
b
4443
739
12101
2395
5595
busco
effector
core
accessory
softcore
Protein length (aa)
c
d
upregulated (>1)
busco genes
effector genes
core genes
softcore genes
accessory genes
0.089
-0.167
0.15
2.445
0.132
0.252
0.201
1.38
0.307
1.597
Density
dN/dS
Log2Fold change 8dpi
e
core
softcore
accessory
unique
Accessory Regions
% of genes
duplicates
effectors
upregulated effectors
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr1A
f
Gene presence
TR4(%)
R1(%)
g
All Fungi
Ascomycota
Hypocreales
F. oxysporum
Isolates
Genes
119
853
114
3614
4
3410
70
8433
h
100%
50%
Accessory
Sub-telomere
Core


present
absent
tomato
banana
melon
cucumber
watermelon
cotton
arabidopsis
non-pathogenic
cucumber, melon
SIX effectors
genomes
race
host
effectors


a
> 20 genes
Dispersed
< 20 genes
Proximal
Tandem
Segmental
b
Fusarium odoratissimum strain II5
Singleton n = 7693
Dispersed n = 8283
Proximal n = 178
Tandem n=246
Segmental n = 124
Median Coverage
200
120
40
1 2 3 4 5 6 7 8 9 10 11 12
c
SIX1c
SIX1b
SIX1a
Chr1
Chr12
Gene
SIX1
Segmental
Proximal
d
Dispersed 3.08
Proximal 2.614
1.003
Tandem
0.119
Segmental
Density
kS value
e
Fusarium oxysporum species complex
FoII5
Fol4827
F. verticillioides
F. graminearum
F. solani
f
49.4
61.4
41.8
36.5
52.9
Genome size (Mbp)
g
Dispersed
Proximal
Tandem
Segmental
Number of duplications
h
Chr12 II5
2x
4x
x coverage
>2
1
0
Middle East
II5
36102