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], E. Martinez de la Parte[1], E. Chavarro-Carrero[1], H.J.G. Meijer[1,3], A. Feurtey[4,5], N. Maryani[6], N. Ordóñez[1], H. Schneiders[7], K. Nijbroek[7], A. H. J. Wittenberg[7], R. Hofstede[7], F. García-Bastidas[7], E.H. Sørensen[7], R. Swennen[8,9], A. Drenth[10], 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

[8]Division of Crop Biotechnics, Laboratory of Tropical Crop Improvement, Catholic University of Leuven, Leuven, Belgium

[9]International Institute of Tropical Agriculture, Kampala, Uganda

[10]The University of Queensland, Brisbane, Australia

*Authors contributed equally

[#]Corresponding authors: Gert H.J. Kema and Michael F. Seidl

Email: gert.kema@wur.nl, m.f.seidl@uu.nl

Author Contributions: ACW collected data and performed the analyses. EMP, CAG, ECC, RH, AW, HS, FGB, NM, NO, EHS and GNT collected data. 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.

Competing Interest Statement: The authors declare no competing interests.

This manuscript includes:

- Main text (5655 words)
  - Introduction (634 words)
  - Material and Methods (1306 words)
  - Results (2682 words)
  - Discussion (1033 words)
- Figures 1 to 6 (colour)
- **Supplementary information:**
  - Supplementary table 1 - 4
  - Supplementary figure 1 – 8
  - Supplementary methods

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. The presence of specific accessory regions influences the host range, and horizontal transfer of some accessory regions can modify the pathogenicity of the receiving strain. However, understanding how these accessory regions evolve 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. In this diverse panel of *Fusarium* strains infecting banana, we analyzed the diversity and evolution of the accessory regions.
- Accessory regions in *Fusarium* strains infecting the same banana cultivar are highly diverse, and we could not identify any shared genomic regions and in planta induced effectors. We demonstrate that segmental duplications drive the evolution of accessory regions. Furthermore, we show that recent segmental duplications and aneuploidy occur specifically in accessory chromosomes and cause the expansion of accessory regions in *F. oxysporum*.
- Taken together we conclude that extensive recent duplications drive the evolution of accessory regions in *Fusarium*, which contribute to the evolution of virulence.

Keywords: Pathogen genome evolution, Genome compartmentalization, Accessory regions, *Fusarium oxysporum*, Segmental duplications

Main Text

Introduction

The interaction between fungal plant pathogens and their hosts drives rapid evolution (Möller & Stukenbrock, 2017). Plant hosts have evolved sophisticated immune systems to detect pathogens, while adapted pathogens, in turn, secrete effectors to deregulate immune responses and to support host colonization (Rovenich *et al.*, 2014; Cook *et al.*, 2015). To facilitate this co-evolutionary cycle, many filamentous plant pathogens evolved compartmentalized genomes where effector genes are localized in distinct genomic regions (Dong *et al.*, 2015; Torres *et al.*, 2020). These effector-rich compartments show extensive genetic variation and are enriched for repetitive sequences such as transposable elements (Raffaele & Kamoun, 2012; Dong *et al.*, 2015; Seidl & Thomma, 2017; Möller & Stukenbrock, 2017; Torres *et al.*, 2020). This spatial separation is often referred to as the 'two-speed' genome (Dong *et al.*, 2015) and is thought to facilitates the rapid diversification of effector gene repertoires to enable pathogens to evade host resistances and to enable host jumps (Sánchez-Vallet *et al.*, 2018).

Fusarium oxysporum is a genetically diverse fungal species complex that consists of three clades that can be further separated into different lineages, recently reclassified as separate species (Maryani et al., 2019). 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 (Edel-Hermann & Lecomte, 2019). *Fusarium oxysporum* evolved a two-speed genome organization where specific genomic regions are characterized by extensive presence-absence polymorphisms between strains (Ma *et al.*, 2010; van Dam *et al.*, 2017; Fokkens *et al.*, 2018; Zhang *et al.*, 2020). These variable accessory regions (ARs) can be embedded in core chromosomes or can encompass entire accessory chromosomes (Ma *et al.*, 2010; Fokkens *et al.*, 2018). Interestingly, shared ARs in otherwise genetically diverse strains have been linked to the adaptation towards the same host (Fokkens *et al.*, 2018; Henry *et al.*, 2021). Moreover, ARs can transfer the capacity to infect a specific host between isolates (Ma *et al.*, 2010; Ayukawa *et al.*, 2021). For example, transfer of accessory chromosome 14 from a tomato infecting *Fusarium* (*Fol*4287) to a

94 non-pathogenic *Fusarium* strain converts the non-pathogenic isolate into a tomato infecting
95 pathogen (Ma *et al.*, 2010). This transfer of pathogenicity is likely linked to effector genes that are
96 localized in ARs such as the 14 well-studied *Secreted In Xylem* (*SIX*) effectors (Rep *et al.*, 2002).
97 Consequently, the presence and absence of effectors can group genetically diverse *F. oxysporum*
98 strains based on their host range (van Dam *et al.*, 2016). Whole-genome sequencing of diverse *F.*
99 *oxysporum* can provide novel insights into the origin of host specificity (Fokkens *et al.*, 2018; Henry
100 *et al.*, 2021). A recent population study of *F. oxysporum* isolates infecting chickpea suggests that,
101 contrary to expectations (Gordon & Martyn, 1997; Henry *et al.*, 2021), *F. oxysporum* might undergo
102 sexual as well as clonal reproduction (Fayyaz *et al.*, 2023), which could impact the constitution and
103 the dynamics of ARs and, ultimately, influence pathogenicity. However, despite the importance of
104 ARs for host adaption, little is known about the constitution of ARs in *F. oxysporum* strains that
105 cause disease in the same host, and the origin and processes that shape the evolution of ARs in
106 *Fusarium*.

107 Fusarium wilt of banana (FWB), which is caused by a suite of *F. oxysporum* species
108 (Maryani *et al.*, 2019), is a major constraint for banana production and threatens food security in
109 tropical and subtropical countries where >400 million people depend on banana cultivation (FAO,
110 2020). In *F. oxysporum* species causing FWB, three different races can be distinguished based on
111 their pathogenicity towards subsets of banana varieties; Race 1 (R1) strains cause FWB in Gros
112 Michel, Race 2 (R2) strains infect Bluggoe, and Tropical Race 4 (TR4) strains infect Gros Michel,
113 Bluggoe, and Cavendish (Ploetz, 2015). In addition to TR4, subtropical race 4 (STR4) strains can
114 cause disease in Cavendish banana under environmental stress conditions, for example caused
115 by lower temperatures in subtropical regions (Ploetz, 2006). All races infect additional locally
116 important banana varieties. TR4 has developed into a pandemic over the last 60 years (Ordonez
117 *et al.*, 2015; van Westerhoven *et al.*, 2022b) and is of particular concern as it causes FWB in
118 Cavendish, the banana variety that dominates global productions (>50%) and export trade (>95%)
119 (FAO, 2020). Despite the impact of FWB on banana production, we know little about the ARs in
120 *Fusarium* strains that infecting banana and their relation to pathogenicity.

121 Here we analyze the dynamics of ARs in a suite of *Fusarium* strains causing FWB. To gain
122 insights into the genomic structure, we generate seven chromosome-level reference genome
123 assemblies for R1, R2 and TR4 strains, and use a pan-genomic framework together with a global
124 panel of 69 strains to gain insights into the evolution of ARs in genetically diverse *Fusarium* species.

125 **Materials and Methods**

126 **Collection of *Fusarium* strains infecting banana**

127 We analyzed a diverse set of 69 *Fusarium* strains infecting banana (**Supplementary Table 1**).
128 *Fusarium* strains were isolated from different banana growing regions worldwide directly from the
129 pseudostem of symptomatic banana plants. Here, 35 *Fusarium* strains were newly sequenced and
130 we included 34 publicly available whole-genome sequencing datasets (based on short-read
131 sequencing approaches) (Guo *et al.*, 2014; García-Bastidas *et al.*, 2014; Zheng *et al.*, 2018;
132 Warmington *et al.*, 2019; Maymon *et al.*, 2020; Garcia-Bastidas *et al.*, 2020; van Westerhoven *et*
133 *al.*, 2022a). To obtain high-quality reference assemblies, seven strains were sequenced with Oxford
134 Nanopore Technology (ONT).
135

136 **Whole-genome sequencing**

137 *Fusarium* isolates were cultured on potato dextrose agar (PDA) and conidia were transferred to
138 liquid media for incubation at 25°C. Fungal material was isolated from liquid culture for genomic
139 DNA extraction. The exact DNA isolation protocol per strain can be found in Supplementary
140 Methods. To obtain long-read, whole-genome sequencing data, seven strain (36102, II5, CR1.1,
141 C058, C135, C081, and M1) were sequenced using Oxford Nanopore Technologies (Oxford, UK).
142 An R9.4.1 flow cell was loaded and run for 24h. Base calling was performed using Guppy (version
143 3.1.5; C058, C135, C081, and M1) and MiniKNOW (version 3.4.6; II5, 36102, and CR1.1). The
144 other *Fusarium* strains were sequenced by Beijing Genomic Institute (BGI) using Illumina paired-
145 end sequencing, 31 were sequenced during this study and 38 were obtained during other studies
146 (Guo *et al.*, 2014; Ordonez *et al.*, 2015; Zheng *et al.*, 2018; Warmington *et al.*, 2019; Maymon *et*
147 *al.*, 2020; Garcia-Bastidas *et al.*, 2020; van Westerhoven *et al.*, 2022a).

148 **Gene expression analysis**

149 Gene expression patterns were compared between *in vitro* growth and 8 days post infection.
150 Cavendish 'Grand Naine' banana plants were grown in a greenhouse compartment (28 ± 2°C, 16
151 h light, and ~85% relativity humidity) and acclimatized under plastic for 2 weeks to maintain high
152 humidity. The roots of ~2.5 months old plants were dip inoculated with 10^6 spores per ml. RNA was
153 isolated from the roots of inoculated plants at 8 days post inoculation and from mycelium grown on
154 PDA medium. Samples were ground and RNA extraction was performed using the Maxwell Plant
155 RNA kit (Promega) following manufacturer's instructions. The quality of the RNA samples was
156 tested by agarose electrophoresis and quantified by Nanodrop (ThermoFisher), and subsequently
157 sent to KeyGene (The Netherlands) for sequencing. Quantifications of gene transcripts from
158 sequenced RNA-seq reads was performed using Kallisto (Bray *et al.*, 2016).

159

160 **Genome assembly and annotation**

161 We obtained chromosome-level *de novo* genome assemblies for seven Fusarium isolates
162 sequenced by ONT long-read technology. Adapter sequences were removed with Porechop
163 (version 0.2.4, default settings; https://github.com/rrwick/Porechop/tree/master) and reads were
164 self-corrected, trimmed, and assembled using Canu (Koren *et al.*, 2017) (version 1.8). Genome
165 assemblies were polished using ONT raw reads with Racon (Vaser *et al.*, 2017) (version 1.5.0) and
166 with high-quality Illumina short read using two rounds of Pilon(Walker *et al.*, 2014) (version 1.24).
167 Contigs were scaffolded, if needed, using longstitch (Coombe *et al.*, 2021) (version 1.0.2).

168 Fusarium strains sequenced solely with short-read sequencing technology were *de novo*
169 assembled with Spades (Bankevich *et al.*, 2012) (version 3.13.0, with default settings). Contigs
170 shorter than 500 bp were subsequently removed. To assess the assembly quality, the genome
171 assemblies were analyzed with QUAST (Gurevich *et al.*, 2013) (version 5.0.2), and genome
172 completeness was estimated based on the presence of conserved single-copy genes using
173 BUSCO (Simão *et al.*, 2015) (version 5.1.2) with the *hypocreales* odb10 database.

174 We performed gene annotation of 69 *Fusarium* genome assemblies using Funannotate
175 (Palmer & Stajich, 2019) (version 1.8.9) with the *Fusarium oxysporum* f. sp. *lycopersici* 4287
176 Uniprot proteins, uniport database, and previously predicted effectors (van Dam *et al.*, 2016) as
177 external evidence. Genomes were masked using RepeatMasker (version 4.1.1) following repeat
178 prediction with RepeatModeler (http://www.repeatmasker.org). For *Fusarium* strains II5, CR1.1,
179 and 36102, RNA-seq data was used to guide gene model prediction, and *ab initio* training
180 parameters estimated using RNA-seq data of II5 were re-used to support the annotation in other
181 genome assemblies. Effectors genes were predicted based on their association with Miniature
182 IMPala (MIMP) elements in *Fusarium* (van Dam *et al.*, 2016). Briefly, ORFs were predicted 5 kb
183 up- and downstream to MIMPs ('AGT[GA][GA]G[GAT][TGC]GCAA[TAG]AA') and these were
184 refined using AUGUSTUS (version 3.3.3)(Stanke & Morgenstern, 2005), with the *Fusarium* species
185 specified. In the set of predicted MIMP-genes, secreted proteins were predicted using SignalP
186 (version 5) and subsequently effectors were predicted using EffectorP (version 3.0). The predicted
187 effectors were added to the Funannotate annotation using AGAT (version 0.8.0.)

188 To recover potentially missing genes in the gene annotations, we used the comparative
189 annotation toolkit (Fiddes *et al.*, 2018) (version 2). Cactus (Armstrong *et al.*, 2020) (version 2.0.1)
190 was used to create whole-genome alignments of all 69 genomes, using the Minigraph-Cactus
191 approach (Hickey *et al.*, 2023). Consecutively, the CAT was used to carry-over gene annotations
192 based on Funannotate gene predictions with the RNA-seq data (with Augustus settings pre trained
193 on II5, --global-near-best 1, --augustus-utr-off, --augustus-cgp, --maf-chunksize 550000 --maf-
194 overlap 50000), which added approximately 200 genes per genome assembly.

195

196 **Core gene phylogeny**

197 The phylogenetic relationship between *Fusarium* strains infecting banana together with 55 publicly
198 available non-banana infecting *Fusarium* strains was inferred from 3,811 conserved single-copy
199 BUSCO genes. Protein sequences were aligned using Mafft (Katoh *et al.*, 2002), (version 7.453)

200 and a maximum-likelihood phylogeny was determine using IQtree (Chernomor *et al.*, 2016)
201 (version 1.6; -m MFP+MERGE, -B 1000 with 1,000 ultra-fast bootstrap), *Fusarium verticillioides*
202 (ASM14955v1 (Cuomo *et al.*, 2007)) was included as an outgroup.

203 **Pangenome analysis**

204 We detected conserved and accessory regions in the set of *Fusarium* isolates by performing all-
205 vs-all whole-genome alignments with nucmer (Kurtz *et al.*, 2004) (version 3.1; --max-match). We
206 removed small alignments with delta-filter (-l 5000) and only retained the best match per alignment
207 (-1). Genomic regions present in all 69 *Fusarium* isolates were considered core, regions covered
208 by less than 80% were considered accessory, and regions between these values were considered
209 softcore. Pairwise similarity between accessory regions was calculated using custom python scripts
210 (https://github.com/Anouk-vw/Fusarium_pg) and alignments were visualized with PyGenomeViz
211 (https://github.com/moshi4/pyGenomeViz) and Circlize (Gu *et al.*, 2014).

212 To identify orthologous groups, we used Broccoli (Derelle *et al.*, 2020) (version 1.1) on the
213 predicted protein-coding genes from the 69 *Fusarium* strains infecting banana; per gene, only the
214 longest transcript was included. OGs present in all isolates were considered core, OGs found in
215 more than 55 genomes (80%) were considered softcore, and genes found in less than 55 of the
216 genomes were considered accessory.

217 We determined the number of non-synonymous and synonymous substitutions (dN/dS)
218 per OG. All genes in an OG were aligned using mafft (Katoh *et al.*, 2002) (version 7.490) and non-
219 synonymous and synonymous substitutions per gene pair in the OG were inferred based on a
220 codon-guided nucleotide alignment, created by pal2nal (Suyama *et al.*, 2006) (version 14), with
221 codeML (from paml (Yang, 2007), version 4.9). Differences between groups of genes were further
222 analyzed using custom python scripts and compared using a two-sided Mann-Whitney U test in
223 SciPy (Virtanen *et al.*, 2020) (version 1.10.1).

224

225 **Estimation of gene age**

226 Orthofinder (Emms & Kelly, 2019) (version 2.5.4) was used to detect the presence of homologs in
227 308 phylogenetically distinct fungal species from the joint genome institute; per fungal family, one

228 fungal genome assembly was randomly selected (Supplementary table **5**). Homologs that occur at
229 least once outside the phylum Ascomycota are considered *old* homologs, homologs occurring only
230 within Ascomycota are further subdivided into *Hypocreales* and *Fusarium*.

231 **Detection of duplication events**

232 We detected gene duplication events by self-comparisons of the translated protein sequences (only
233 the longest isoform) of the chromosome-level genome assemblies of *Fusarium* infecting banana,
234 as well as *Fusarium oxysporum* f. sp. *lycopersici* (strain *Fol*4287 (Ma *et al.*, 2010);
235 GCF_000149955.1, *Fusarium* graminearum (Cuomo *et al.*, 2007) (GCF_000240135.3), *Fusarium*
236 *verticillioides* (Cuomo *et al.*, 2007) (GCF_000149555.1), and *Fusarium solani* (Mesny *et al.*, 2021)
237 (GCF_020744495.1). Per *Fusarium* strain, the predicted protein sequences were compared with
238 BLAST and the best five hits were retained after filtering for an e-value $1e^{-10}$ and query coverage
239 larger than 50%. Then, BLAST alignments were used to detect collinear duplicated blocks with
240 MCscanX (Wang *et al.*, 2012); blocks of five genes were considered collinear and matches were
241 filtered with the parameters match-score 50 and e-value 10^{-5}. Subsequently, we classified
242 duplicated genes into dispersed, proximal, tandem, or segmental duplications with MCscanX's
243 'duplicate_gene_classifier' script.

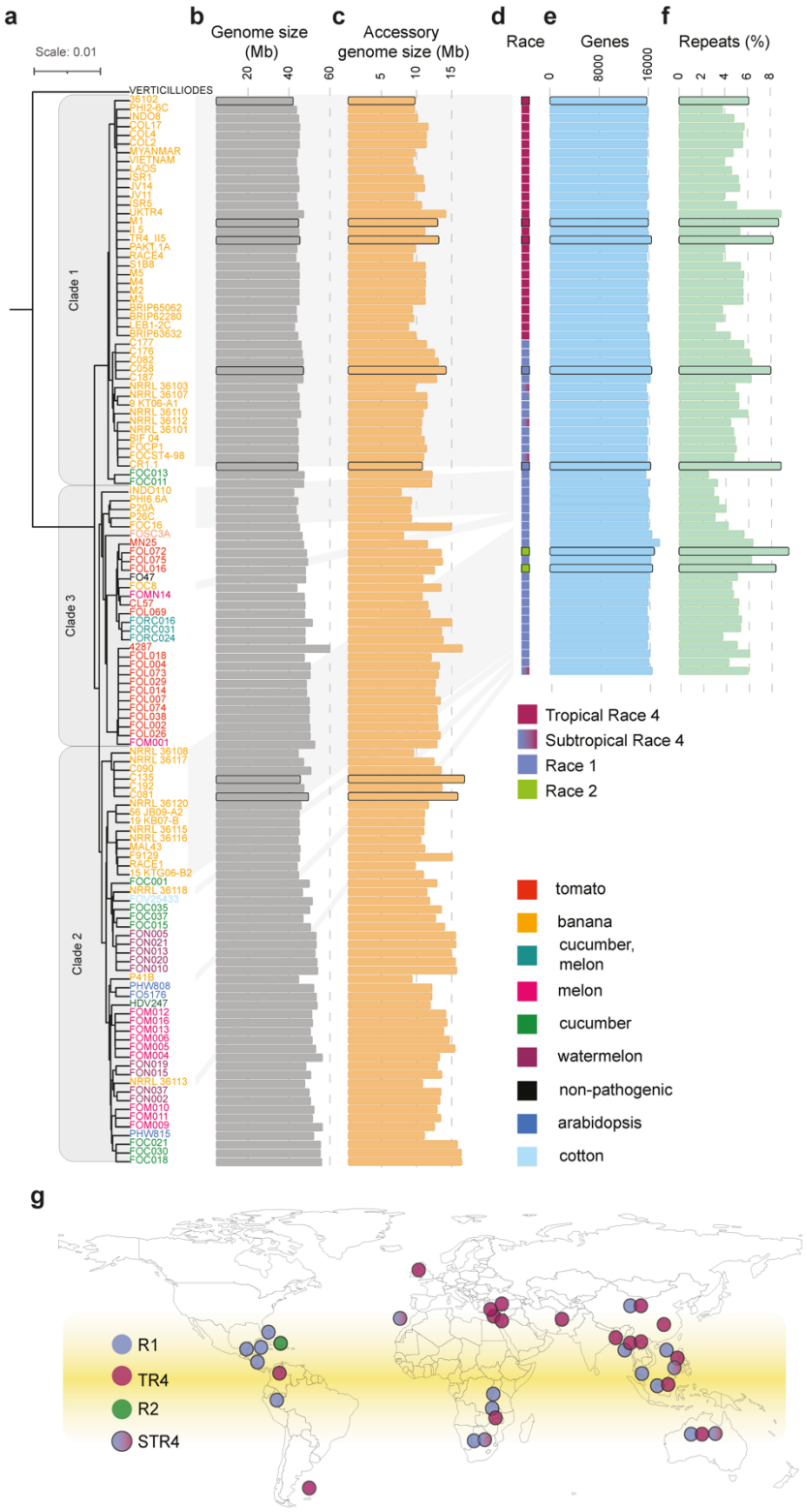
244 Copy number variations were identified by mapping short-read, whole-genome sequencing
245 datasets against the chromosome-level genome assembly of *Fusarium* strain II5 using BWA-mem
246 (Li, 2013) (version 0.7.17). We determined the sequencing coverage over the reference genome
247 with bedtools (version 2.30), which was visualized Wgscoverageplotter from jvarkit (Lindenbaum,
248 2015) (version 1.1.0), filtering reads with a mapping quality of 0. Single nucleotide polymorphisms
249 were called using GATK (version 4.0.2; -ploidy 1) and filtered according to GATK best practices
250 (Van der Auwera *et al.*, 2013).

251 **Results**

252 **Fusarium strains *infecting banana are genetically diverse***

253 To study the occurrence and evolution of ARs in *Fusarium* strains infecting banana, we sequenced
254 and assembled 69 strains (Fig. **1a**, Supplementary table **1**) that were isolated from all major banana
255 growing regions and classified into R1 (39 strains), R2 (two strains) and TR4 (28 strains) (Fig. **1a,b**,
256 Supplementary table **1**). *De novo* assembly of seven long-read sequencing strains infecting
257 different banana varieties (three TR4, two R1, two R2) achieved chromosome-level contiguity
258 containing 12 to 15 contigs, most of which represent complete chromosomes flanked by telomeric
259 repeats (Supplementary table **2,3**). While genome assemblies obtained solely from short-read data
260 yielded more fragmented assemblies, all assemblies contained at least 96.9% of the single-copy
261 BUSCO genes. The assemblies range from 43 to 51 Mb in size (Fig. **1b**) and encode between
262 15,664 and 17,865 predicted proteins, of which 551 to 671 are effector candidates, and consist for
263 2.6 – 8.9% of repeats (Fig. **1e,f**), similar to previous *Fusarium* genome assemblies (van Dam *et al.*,
264 2016; Warmington *et al.*, 2019; Zhang *et al.*, 2020).

265 To assess the genetic diversity of *Fusarium* strains infecting banana in relation to 55
266 publicly available *Fusarium* strains infecting eight different host plants (van Dam *et al.*, 2016)
267 (Supplementary table **1**), we conducted a phylogenetic analysis based on 3,811 single-copy
268 orthologous genes (2,226,902 amino acid positions). As expected, *Fusarium* strains infecting
269 banana are polyphyletic and associated with *F. oxysporum* clades 1, 2, and 3 (Fig. **1a**)(O'Donnell
270 *et al.*, 1998; Maryani *et al.*, 2019). We observed low nucleotide diversity between TR4 strains (pi-
271 nucleotide diversity 0.0018) compared to a 100x higher diversity between R1 strains (pi-nucleotide
272 diversity 0.185), corroborating that TR4 strains evolved from a single recent clonal origin (Ordonez
273 *et al.*, 2015; Maryani *et al.*, 2019). Based on their genetic diversity, *Fusarium* strains infecting
274 banana have recently been separated into different species, and the lineage that encompasses
275 TR4 strains (Fig. **1d**) is now referred to as *Fusarium odoratissimum* (Ordonez *et al.*, 2015; Maryani
276 *et al.*, 2019).


a
Scale: 0.01
b
Genome size (Mb)
c
Accessory genome size (Mb)
d
Race
e
Genes
f
Repeats (%)
Clade 1
Clade 3
Clade 2
Tropical Race 4
Subtropical Race 4
Race 1
Race 2
tomato
banana
cucumber, melon
melon
cucumber
watermelon
non-pathogenic
arabidopsis
cotton
g
R1
TR4
R2
STR4

277

278 **Figure 1. *Fusarium* strains infecting banana are genetically diverse. a**) *Fusarium* strains that
279 infect banana are polyphyletic (yellow labels). The relationship between *Fusarium* infecting banana
280 and *Fusarium* strains that infect other plant hosts (see label color) is estimated using maximum-
281 likelihood phylogenetic analysis based on 3,811 concatenated conserved single-copy BUSCO
282 genes. Black outline indicates high quality genome assemblies. **b)** Genome sizes of *Fusarium*
283 strains range between 43 to 61 Mb. **c**) The sizes of accessory regions in *Fusarium* strains range
284 from 6.7 to 16.5 Mb. **d**) TR4 strains (purple squares) occur in clade 1 and are genetically similar,
285 as well as the two R2 strains (green). By contrast, the R1 strains (blue squares) are polyphyletic.
286 **e**) *Fusarium* strains that infect banana encode between 15,664 and 17,865 genes and **f**) between
287 2.9 and 8.9% repeats. **g)** The collection of 69 *Fusarium* strains originates from different banana
288 growing regions worldwide; the approximate origin of isolation is indicated by dots, colored
289 according to the race designation (R1, R2, STR4, and TR4). Strains sequenced by long-read
290 sequencing technology are highlighted in black.

291

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

293 To identify ARs in *Fusarium* strains infecting banana, we performed whole-genome alignments of
294 the seven chromosome-level genome assemblies to the reference genome assembly of the tomato
295 infecting *F. oxysporum* strain 4287 (*Fol*4287)(Ma *et al.*, 2010). We observed eleven homologous
296 chromosomes between the TR4 reference strain II5 and *Fol*4287, that are considered the core
297 genome (chromosome 1-11; Fig. **2a**). In addition to the core chromosomes, we observed two large
298 ARs specific for II5. The first spans 1.8 Mb occupying 27% of chromosome 1 in proximity to one
299 telomere and the second region is 1.1 Mb in size and constitutes the entire contig 12 (Fig. **2a**).
300 Contig 12 only contains one telomere (Fig. **2b**), and therefore it remains unknown whether this
301 accessory region represents an extra chromosome or is attached to one of the core chromosomes.
302 Importantly, we assemble a similar contig in TR4 strain M1, and a similar contig is present in the
303 independently assembled TR4 strain Eden (Warmington *et al.*, 2019), which strongly suggests that
304 this contig represents an separate accessory region.

305 Compared to the eleven core chromosomes, the two ARs have an increased repeat
306 content (28.30% vs 6.41%) and a decreased gene density (32.15% vs 59.71%) and GC content
307 (46.4% vs 47.5%) (Fig. **2b**). Importantly, the ARs in II5 encode 463 genes, nine are homologous to
308 *SIX* genes (Fig. **2b**) and 22 are predicted effectors (out of 629 predicted effectors). Although this

309 suggests that the ARs are important for pathogenicity on banana, they are not collinear with other
310 *Fusarium* strains infecting banana (Supplementary Fig. **1**).

311 To further analyze the diversity of ARs, we identified ARs in our global collection using a
312 pan-genomic approach based on all-vs-all whole genome alignments. ARs were defined as regions
313 longer than 5 kb that are absent in more than 55 of 69 strains (80%). We identified a varying amount
314 of ARs in *Fusarium* strains infecting banana (Fig. **1c**), ranging from 6.7 Mb (15% of the genome
315 size) in strain Indo110 (R1) to 15.5 Mb (30%) in strain C135 (R2), in line with the 19 Mb (29%) of
316 ARs in the reference strain *Fol*4287 (Ma *et al.*, 2010). Next to the two ARs in chromosomes 1 and
317 contig 12, our pan-genomic approach identified additional 8 Mb of ARs in II5, typically localized at
318 sub-telomeric regions (defined as the first and last 10% of each chromosome) (Fig. **2b**). These sub-
319 telomeric regions contain 1,287 genes, of which 80 are predicted effector genes (6.6%). We also
320 identified that chromosomes 9 (42.6% ARs, 87 SNPs per kb), 10 (27% ARs, 80 SNPs per kb), and
321 11 (36% ARs, 92 SNPs per kb) are less conserved than the other core chromosomes (on average
322 20% ARs, 59 SNPs per kb) **(**Supplementary Fig. **2)**, which resembles the situation of the previously
323 identified 'fast-core' chromosomes in *Fol*4287 (Fokkens *et al.*, 2018).

324 We show that ARs can be defined based on sequence conservation and differ from the
325 core genome in gene, repeat and GC-content. Consequently, principal component analyses
326 distinguished ARs from the core genome based on these four features (Fig. **2c**). Unanticipatedly,
327 this analysis also separated the large ARs on chromosome 1 and contig 12 from the ARs localized
328 at the sub-telomeres (Fig. **2c**), this separation is driven by the slightly lower GC content in sub-
329 telomeric regions (44% vs 47%). A reduced GC-content can be caused by repeat induced
330 polymorphisms (RIP), a fungal defense mechanisms that introduces C to T mutations in repetitive
331 regions (Galagan & Selker, 2004). In the sub-telomeric region, 25.4% of the repetitive elements are
332 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.

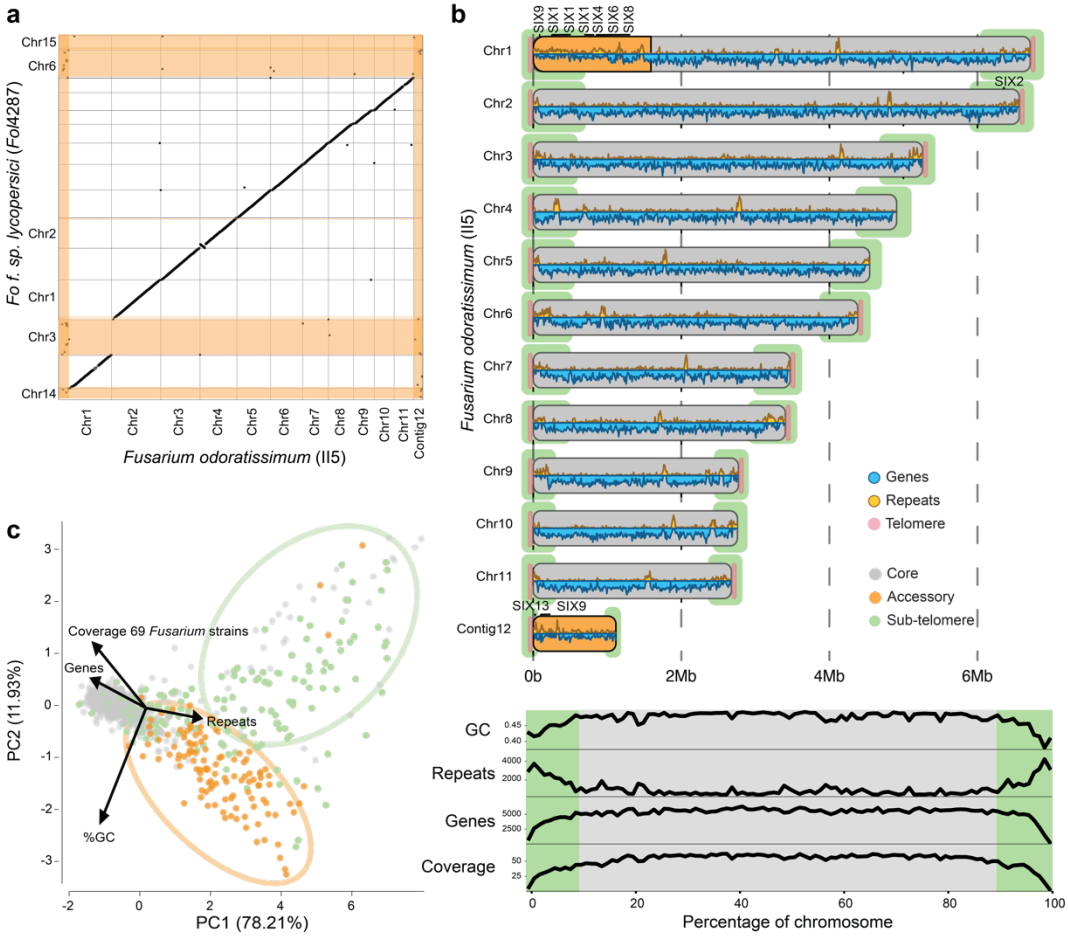


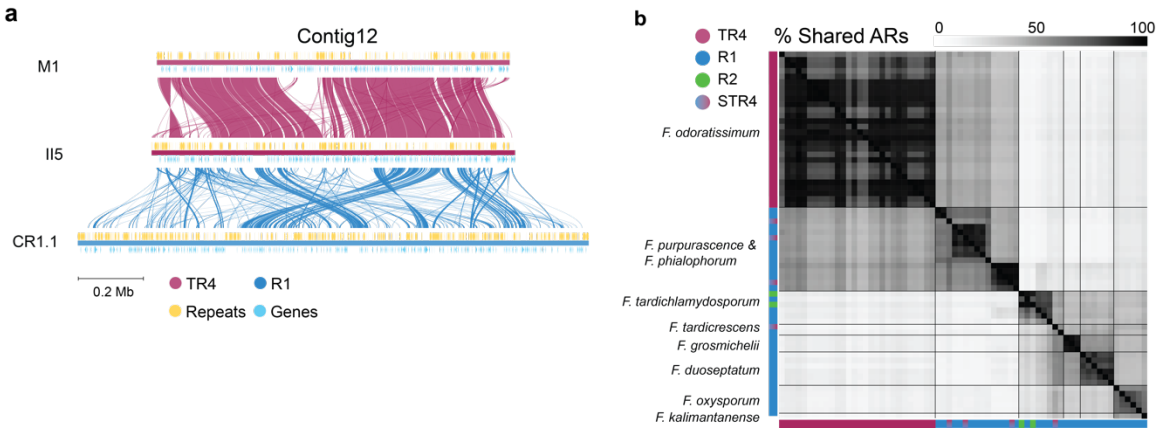
Figure 2. *Fusarium* strains infecting banana carry accessory regions (ARs). a) The TR4 reference strain II5 contains eleven core chromosomes (chromosome 1-11) as well as two ARs (1,8 Mb on chromosome 1 and entire contig 12). Core chromosomes align to *Fol*4287 in contrast to the ARs that are highlighted 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, first and last 10% of the chromosome). **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 contig 12 (orange).

348 ***Highly variable accessory regions in* Fusarium *strains infecting banana***

349 To characterize the similarity of ARs in *Fusarium* strains infecting banana, we compared the ARs
350 from the seven chromosome-level genome assemblies. Notably, the ARs are highly diverse with
351 little sequence and structural similarity (Fig. **3,** Supplementary Fig. **1**). In TR4, the ARs from strain
352 II5 on chromosomes 1 and contig 12 show extensive similarity to the corresponding chromosomes
353 in strain M1 (Fig. **3a**, Supplementary Fig. **1**). On the other hand, isolate 36102, causing milder
354 symptoms in Cavendish than TR4 II5, carries only the AGR on chromosome 1, and does not
355 encode a region similar to contig 12 (Supplementary Figure 1). None of the AGRs from strain II5
356 are present in the strains representing Race 1 and Race 2. Similar to the TR4 strains, the accessory
357 contigs of two Race 2 strains correspond to each other (Supplementary Fig. **1**), but no shared ARs
358 are found among R1 strains. Moreover, we observed that no ARs are conserved among different
359 races, for example only limited genetic material is shared between ARs of TR4 and R2
360 (Supplementary Fig. **1**). The diversity of ARs was further corroborated by pairwise analyses of all
361 69-strains, which revealed that, on average, TR4 shared 91% of the ARs, whereas R1 shared only
362 24% (Fig. **3b**). In line with the high number of shared ARs in TR4 (*F. odoratissimum*, Fig. **3b**), the
363 amount of shared ARs was highest among R1 strains belonging to the same *Fusarium* species
364 (60% shared ARs**,** Fig. **3b**, Supplementary Fig. **3**), however R1 strains from different species share
365 very few ARs (Fig. **3b**).

366 To quantify to which extent *Fusarium* strains infecting banana share ARs with *Fusarium*
367 strains infecting different plant hosts, we identified and compared ARs in a set of 55 additional
368 *Fusarium* strains that are pathogenic to different plant hosts (van Dam *et al.*, 2016). We observed
369 that genetically related strains share most ARs (Supplementary Fig. **4**), however we also noticed
370 that *Fusarium* strains infecting banana encode more diverse ARs than for instance the tomato-
371 infecting strains; on average, *Fol* strains share 62% ARs, with a minimal 30% of the ARs shared
372 between any two *Fol* strains (Supplementary Fig. **4**). In contrast, banana-infecting strains on
373 average share 44% of the ARs and a pair of strains can share as little as 12% (Supplementary Fig.

374 **4**). Thus, we demonstrate that ARs are highly variable among *Fusarium* strains infecting banana
375 and, importantly, strains do not share one host-associated chromosome in contrast to what was
376 previously reported for *Fusarium* infecting tomato (Fokkens *et al.*, 2018).



377
378 **Figure 3. Accessory regions in *Fusarium* strains infecting banana differ significantly. a**)
379 Accessory contig 12 of the TR4 reference strain II5 is shared with TR4 strain M1, however this
380 contig shows little similarity to R1 strain CR1.1. Lines (red and blue) indicate aligned regions
381 between strains. Small alignments are present between R1 and TR4 (blue lines), however these
382 are rearranged and often involve repetitive sequences (yellow blocks). **b**) R1 strains (39 strains,
383 highlighted in blue) have a diverse accessory content sharing on average only 20% of the ARs,
384 whereas TR4 strains (28, highlighted in purple) share most of their ARs (91%). Phylogenetically
385 related strains, here indicated by their species name (Maryani *et al.*, 2019), share more ARs than
386 distantly related strains.

387

388 **Evolutionary *dynamics of the accessory genome***

389 *Fusarium* strains infecting banana do not share large ARs (Fig. **3b**), but we reasoned that ARs
390 nevertheless might share genes that contribute to the pathogenicity towards banana. We therefore
391 grouped 1.1 million protein-coding genes predicted in 69 strains into 22,612 orthologous groups,
392 and subsequently analyzed the gene-content diversity. The pan-genome consisted of 12,101 core
393 groups (53%) that are present in all 69 strains, 2,395 softcore groups present in at least 80% of the
394 strains, 5,595 (25%) accessory groups present in less than 80% of the strains, and 2,521 unique
395 genes (Fig. **4a**). Importantly, the pan-genome based on our collection captures the diversity of
396 protein-coding genes in *Fusarium* strains infecting banana as we did not observe an increase in
397 the pan-genome size after including more than 40 strains (Fig. **4a**). The conserved core genes are,
398 as expected, enriched with genes with housekeeping functions, whilst accessory genes are
399 enriched with genes encoding secondary metabolites (Supplementary Fig. **5**). No enrichment of
400 effector genes is found in any of the three categories (core, accessory, or softcore); the pan-
401 genome consists of 739 gene families encoding predicted effectors that are evenly distributed over
402 the different gene categories (3.7% of accessory genes and 3.5% of the core genes). The role of
403 accessory genes in host infection is suggested by gene expression, as accessory genes are
404 upregulated eight days after banana infection, with a median $\log_2$fold change of 1.59 (compared
405 with the *in vitro* control) (Fig. **4b**), comparable to the upregulation of effector genes (median $\log_2$fold
406 change of 2.4). The core genes, however, show significantly less increase of gene expression upon
407 plant infection (median $\log_2$fold change = 0.25, $P < 0.05$, two-sided Mann-Whitney U test) (Fig. **4b)**.
408 Although this suggests a role of accessory genes in host-pathogenicity, we observed a highly
409 diverse gene content in R1 strains with only 246 out of 7,832 accessory genes (3%) being shared
410 among all R1 strains (Fig. **4f**). TR4 strains, in contrast, have a more similar gene profile and share
411 2,493 of the 4,406 accessory genes (57%).

412 In TR4 strain II5, ARs have a different gene composition than the core regions, with ARs
413 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. **4g,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 (Wolf *et al.*, 2009). 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 based on the defined orthogroups (Fig. **4e**), suggesting that gene duplications play a role in the evolution of ARs in *Fusarium* strains infecting banana.

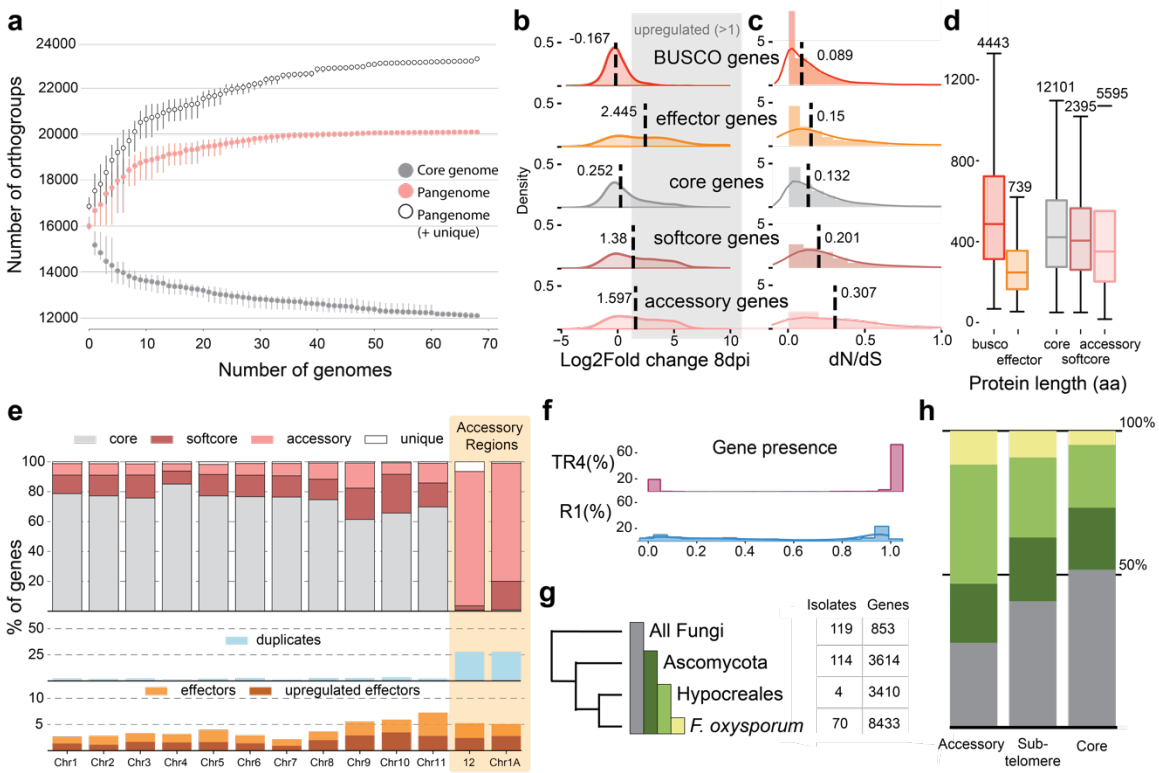


Figure 4. Evolutionary dynamics of genes in the different genomic compartments in *Fusarium* strains infecting banana. a) Global collection of *Fusarium* infecting banana has a pan-genome size of 20,091 orthologous groups (OG, 22,612 when unique genes are included), of which

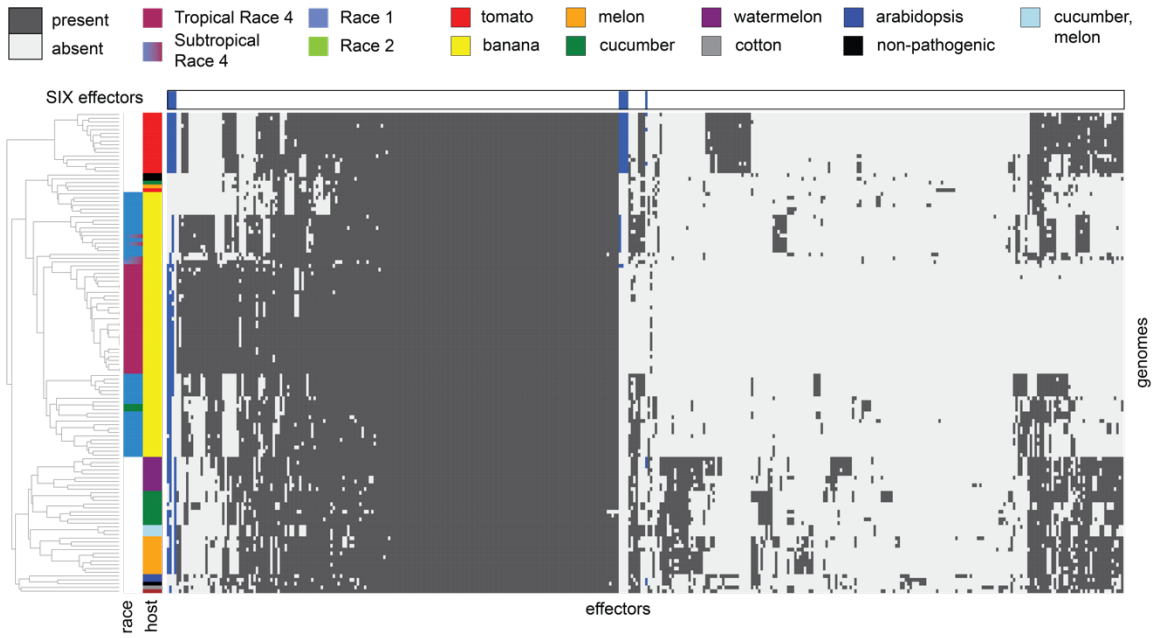
430 12,101 are present in all strains (grey). The pan-genome is saturated after adding 40 genomes.
431 The order of genomes was randomly sampled ten times, dots show the average pangenome size
432 and the vertical grey lines indicate the standard deviation. **b**) Accessory genes are upregulated *in*
433 *planta* at 8 days post inoculation of banana plants. Average log2fold change of accessory genes
434 (2,416 in II5) and effector genes (629 in II5) shows these genes are typically upregulated upon
435 plant infection, which contrasts the core (11,740 genes, log_2-fold change = 0.252) and BUSCO
436 genes (4,364 genes, log_2-fold change = -0.167). The median log2fold change of all gene categories
437 is significantly different ($p<0.05$, two sided Mann-Whitney U test). **c**) Sequence conservation
438 estimated by dN/dS values. BUSCO gene families (4,443 OGs) and core gene families are most
439 conserved, whereas the accessory genes evolve under more relaxed selective pressure. All gene
440 categories show significant variation in dN/dS value ($p<0.05$, Mann-Whitney U test). **d)** Distribution
441 of protein length separated by different gene categories shows that accessory gene families (5,595
442 OGs, pink) are shorter (median length of 315 aa per OG) than core gene families (median length
443 469 aa per OG, 12,101 OGs, grey). All gene categories differed significantly from each other
444 ($p<0.05$, Mann-Whitney U test). **e**) ARs in II5 (orange, contig 12 and chromosome 1A) consist
445 mostly of accessory genes (pink, 405 out of 463, 87%) and 28% of these genes are part of
446 expanded orthogroups (129 out of 463). Accessory genes can also be found in core chromosomes.
447 **f**) Gene content is variable between R1 strains, and only 20% of the genes are present in 80-90%
448 of the R1 strains. Most genes (60%) in TR4 are present in all TR4 strains (90-100%). **g)** Gene
449 families in TR4 strain II5 differ to which extend they have homologs in 308 fungal phyla; most genes
450 are present in related *Fusarium* strains, but only a few gene families (853) are conserved among
451 all considered fungi. **h**) Accessory regions contain most recent genes, present in *Fusarium* (51)
452 and Hypocreales (160). Genes in core regions are evolutionarily older with more genes that have
453 homologs in 'all fungi' (7,764 out of 11,740 genes).

454

455 ***Variable effector repertoire in* Fusarium oxysporum *genomes***

456 Effector gene repertoires define host range (van Dam *et al.*, 2016; Brenes Guallar *et al.*, 2022), and
457 we anticipated that effector presence-absence profiles would distinguish *Fusarium* strains infecting
458 banana from *Fusarium* strains infecting other hosts. We predicted candidate effector genes in all
459 124 *Fusarium* isolates based on their proximity to Miniature IMPala elements (MIMP) that are
460 known to co-localize with some effector genes in *F. oxysporum* (van Dam *et al.*, 2016; Brenes
461 Guallar *et al.*, 2022), which yielded 398 MIMP-associated effector gene families (160 of these were
462 predicted in the complete set of 739 candidate effectors). The predicted MIMP-associated effector
463 repertoires clustered *F. oxysporum* isolates on host range (van Dam *et al.*, 2016; Brenes Guallar
464 *et al.*, 2022) (Fig. **5)**. While banana-infecting strains form a clear cluster, only four MIMP-associated
465 effectors were identified that occur in most (i.e., >80%) of banana-infecting strains but are
466 consistently absent (i.e., present in less than 20%) in other *F. oxysporum* strains. Moreover, races
467 within *Fusarium* strains infecting banana are not clearly separated; TR4 strains have very similar

468 effector profiles, yet R1 strains differ considerably and do not encode shared effectors.
469 Interestingly, two STR4 strains carry an effector profile remarkably similar to the effector profile of
470 TR4 strains (Figure 5). This suggests that this effector profile contributes to the pathogenicity of
471 these STR4 strains to Cavendish banana. However, this effector profile is not STR4 specific as the
472 other two STR4 strains encode effector profiles similar to R1 strains (Figure 5). This difference can
473 be the result of misclassification of STR4 strains, because STR4 infects Cavendish only under
474 environmental stress conditions and this environmental component makes it difficult to distinguish
475 true STR4 isolates from Race 1 isolates (Ploetz, 2006). To further investigate effectors shared
476 between TR4 and STR4, an accurate screening strategy to distinguish true STR4 needs to be
477 conducted. We here reveal a diverse effector repertoire in R1 strains, together with the variable
478 ARs and gene content, this suggests that the species encompassing R1 utilize varying molecular
479 mechanisms to support banana infection.



480
481 **Figure 5. Predicted effector profile clusters *Fusarium* strains infecting the same host.**
482 *Fusarium* strains infecting banana (yellow) are separated from *Fusaria* infecting other hosts. In the
483 set of 124 *Fusarium* isolates, 398 MIMP-associated effector families are predicted. The R1 strains
484 (blue) have diverse effector profiles and do not cluster together, whereas the effector profiles of
485 TR4 strains (purple) are highly similar. Nine SIX effectors (blue) are identified in the 398 MIMP-
486 associated effector families and are present in Fusarium strains infecting tomato, from which they
487 are originally identified, as well as in other Fusarium genomes.

488

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

490 We observed diverse ARs in *Fusarium* strains infecting banana with a high abundance of
491 evolutionary young genes and genes that evolved via gene duplications (Fig. **4e**). Duplications
492 have been previously reported in *F. oxysporum*, including *Fusarium* strains infecting banana (Kistler
493 *et al.*, 1995; Ma *et al.*, 2010; Vlaardingerbroek *et al.*, 2016; Armitage *et al.*, 2018; Li *et al.*, 2020).
494 To better understand the role of gene duplications in the origin and diversification of ARs, we
495 characterized gene duplications in *F. oxysporum* using MCscanX (Wang *et al.*, 2012) and classified
496 them into four categories (dispersed, proximal, tandem, and segmental) (Fig. **6a**). Remarkably,
497 most genes in strain II5 (8,831 genes) have been affected by gene duplication during their evolution
498 (Fig. **6b,** Supplementary Fig. **6**). This high number of duplications include old gene duplications,
499 recognized as distant homologs that acquired considerable number of mutations over time. When
500 we apply a more stringent percent identity filtering, the number of duplicated genes is reduced, yet
501 the number of segmental duplicated genes is least affected (supplementary table **4**), indicating that
502 the segmental duplicated genes acquired less mutations and thus evolved more recently**.** We
503 observed similar results for other *F. oxysporum* strains as well as for other *Fusarium* species; *F.*
504 *verticillioides*, *F. graminearum,* and *F. solani* (syn. *Neocosmospora solani*), suggesting that
505 duplications played a major role in *Fusarium* evolution (**Fig 6b**). Interestingly, we observed that
506 segmental duplicated genes occur specifically in ARs (72 out of 124 in II5) and sub-telomeres (48
507 out of 124 in II5) in all chromosome-level *F. oxysporum* assemblies (Fig**. 6b**, Supplementary Fig.
508 **6**). In *F. solani*, which contains accessory chromosomes analogous to those observed in *F.*
509 *oxysporum* (Coleman *et al.*, 2009), 139 segmental duplicated genes were identified. In contrast, in
510 *F. verticillioides* and *F. graminearum*, for which no accessory chromosomes have been described,
511 no segmental duplicated genes are found (Fig. **6c-e**), underscoring the link between accessory
512 chromosomes and segmental duplications in *Fusarium*. These duplications are possibly driven by
513 TE activity (Wicker *et al.*, 2010; Faino *et al.*, 2016), and we observe that segmental duplications in
514 TR4 strain II5 occur closer to transposable elements (median distance of 1734 kb, $p < 0.05$, two

515 sided Mann-Whitney U test) then other duplication types (dispersed 7721 bp, proximal 7503 bp,
516 tandem 5941 bp), suggesting that TE activity influences the segmental duplications.

517 The highest number of segmental duplicated genes (2,887) is observed in *Fol*4287,
518 carrying the largest accessory genome (19 Mb, 30% of the total genome size) (Fig. **6d**) and 2,366
519 (82%) of the segmental duplicated genes are in ARs. Interestingly, many segmental duplicated
520 genes are shared between accessory chromosomes 3 and 6 (Supplementary Fig. **7**), as previously
521 observed (Ma *et al.*, 2010), indicating that accessory chromosomes evolved by inter-chromosomal
522 segmental duplication that drive the expansion of ARs in *Fusarium*.

523 Duplications of AR genes significantly effect effector repertoires as 336 out of 669 predicted
524 effectors in II5 evolved via duplications, and nine effectors are part of segmental duplications.
525 Among the segmental duplicated effectors is *SIX1*, an effector that is essential for full virulence to
526 banana (Widinugraheni *et al.*, 2018). Three copies of *SIX1* (*SIX1a*,*b*,*c*) are present in the AR on
527 chromosome 1 of the TR4 strains M1, 36102 and II5. *SIX1a* and *SIX1b* are part of a segmentally
528 duplicated block (Fig. **6g**), whereas *SIX1c* is a proximal duplication that shares 82% amino acid
529 sequence identity with *SIX1a* and 73% sequence identity with *SIX1b*. In contrast, the R1 strain
530 CR1.1 contains only one copy of *SIX1*, resembling *SIX1c*, on accessory contig 12. The duplicated
531 blocks with *SIX1* are interspersed with non-duplicated genes (Fig. **6g**), indicating that over time
532 these blocks further diverged by gene gains and losses. Interestingly, the segmentally duplicated
533 *SIX1* blocks also share similarity to a region on contig 12 (Fig. **6g**), yet no copy of *SIX1* is present,
534 and thus *SIX1* is either lost or has been gained on chromosome 1 after the segmental duplication
535 between chromosome 1 and contig 12 occurred. Segmental duplications therefore contribute to the
536 evolution of virulence factors such as *SIX1*, which is essential for successful banana infection
537 (Widinugraheni *et al.*, 2018).

538 To estimate the relative timing of the duplication events, we used the synonymous
539 substitution rate (dS) between duplicated gene pairs as a proxy of time. Dispersed duplicated genes
540 had the highest average dS value (3.08), suggesting that these are the most ancient duplicates,
541 while genes pairs that arose via segmental duplication evolved more recently, with a significantly

542 lower average dS value of 0.61 (Fig. **6f**). To assess if even more recent large-scale duplications
543 are present in our global panel, we determined the read depth of all isolates sequenced with short
544 reads against II5. Based on the genome-wide average read depth (~40x coverage), we determined
545 that contig 12 has been entirely duplicated (read depth of ~80x coverage), while a section (position
546 ~64-116 kb; ~160x coverage) occurs four times (Fig. **6h**). This demonstrates that the II5 carries
547 two (nearly) identical copies of contig 12 (i.e., aneuploidy), in addition to the previously observed
548 segmental duplications (Fig. **6h**). No copy number variation of the accessory region of chromosome
549 1 is observed. The four partial copies of contig 12 possibly originate from an additional duplication,
550 it remains unclear if this duplicated stretch is attached to a chromosome, or if it possibly gave rise
551 to an additional smaller accessory chromosome. To resolve the origin of this duplicated region, we
552 mapped the long reads to the assembly. We observe many clipped reads in this region, which do
553 not entirely correspond to the contig 12 assembly, however the clipped regions do not align to
554 another chromosome. This suggests the region is not attached to another chromosomal arm,
555 leaving the possibility that the duplicated region could represent a duplicated chromosomal arm.
556 We didn't observe a threefold increase on contig 12, suggesting that the duplication happened
557 multiple times. Interestingly, aneuploidy of contig 12 can also be observed in three additional TR4
558 strains from the Middle East, but not in the other 26 TR4 strains (Fig. **6h**). Collectively, we
559 demonstrate that the evolutionary dynamics of ARs in *F. oxysporum* is driven by recent segmental
560 duplications and aneuploidy.

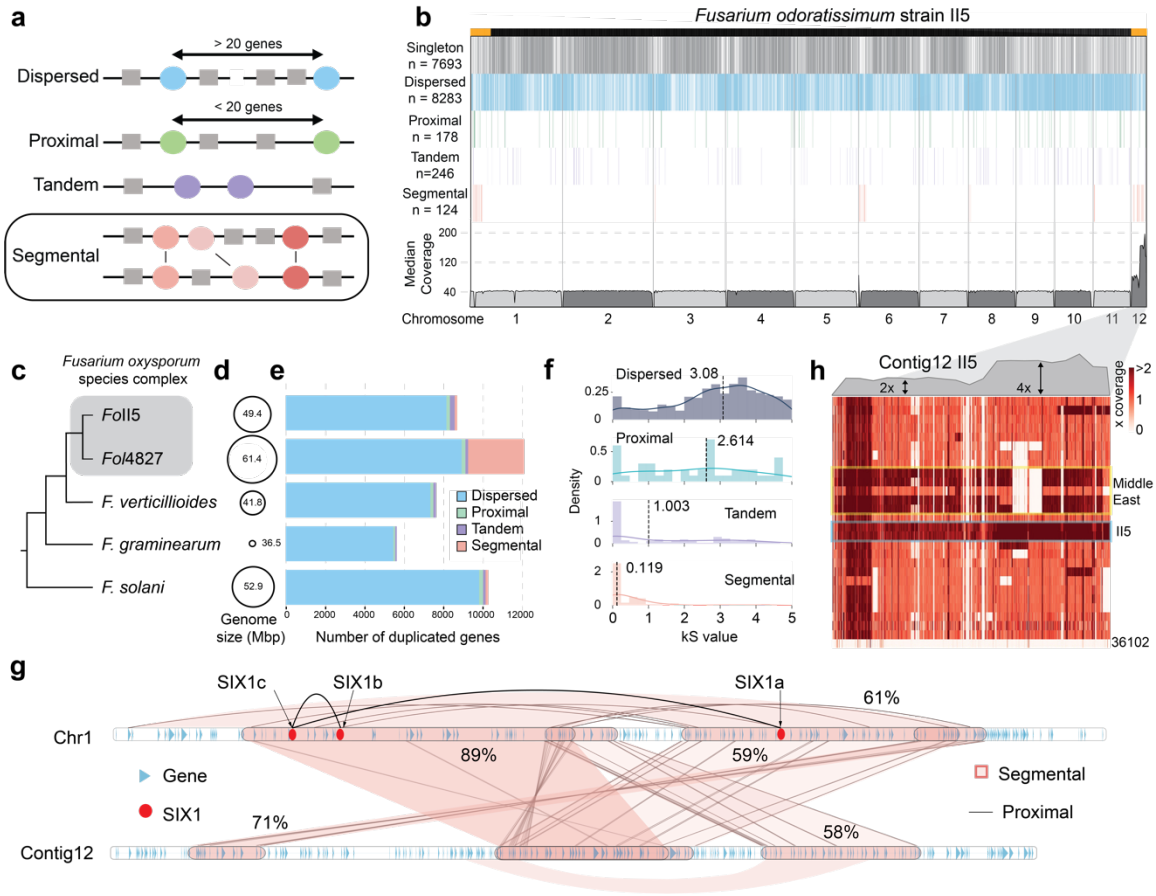


Figure 6. Segmental duplications are involved in the expansion of accessory regions in *F. oxysporum*. a) Four different types of duplications are distinguished. **b**) Segmental duplications (pink) occur mainly in the accessory regions in strain II5 (TR4). Sub-telomeric regions have more dispersed duplicated genes (blue) than the chromosomal arms. On average, the read coverage of II5 mapped onto the II5 reference genome is 40x, whereas contig 12 has a coverage between 80x and 160x, suggesting copy-number variations. **c**) The schematic tree depicts the relationship of *Fusarium* species with **d**) their genome sizes. **e**) Extensive segmental duplications are only found in *F. oxysporum* and *F. solani* and are virtually absent in *F. verticillioides* and *F. graminearum*. **f**) The age of duplications was estimated from the number of synonymous substitutions between duplicated gene pairs. 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). **g**) 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, brown lines between the genes (blue arrows) highlight the genes involved in the segmental duplication. Not all genes in a block are part of the segmental duplication, for example a region on contig 12 is similar to the *SIX1* regions on chromosome 1, yet *SIX1* is absent on contig 12. **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 contig 12.

582 **Discussion**

583 Many fungal plant pathogens evolved compartmentalized genomes with conserved core regions
584 and variable accessory regions (ARs) (Dong *et al.*, 2015; Torres *et al.*, 2020). Some members of
585 the *Fusarium oxysporum* species complex are important plant pathogens and are well-known to
586 have extensive ARs that can encompass entire chromosomes, which have been linked to
587 pathogenicity towards specific hosts (Sánchez-Vallet *et al.*, 2018). Here we use a pan-genomic
588 framework to study the occurrence, composition, and evolution of ARs in a global collection of
589 banana infecting *Fusarium* species. Strains carry up to 15 Mb of ARs that contain *in planta* induced
590 effector genes as well as most *SIX* effectors. ARs are highly divergent among strains and races,
591 and consequently, we cannot identify shared ARs or accessory chromosomes that can be linked
592 to pathogenicity towards banana. We furthermore demonstrate that *Fusarium* evolved by extensive
593 duplications and that ARs specifically are shaped by recent segmental duplications in *F. oxysporum*
594 and *F. solani*. Lastly, we uncover that an accessory contig in II5 underwent recent aneuploidy,
595 suggesting that these processes drive the emergence and evolution of ARs in *F. oxysporum*.

596 Fusarium wilt of banana is a major threat to food security as banana is a major staple crop in
597 tropical and subtropical regions (Drenth & Kema, 2021; van Westerhoven *et al.*, 2022b). Compared
598 to R1 strains, which caused epidemics in Gros Michel bananas in the 1900s, TR4 strains emerged
599 more recently (1967) and cause the ongoing FWB pandemic in Cavendish plantations (van
600 Westerhoven *et al.*, 2022b). TR4 strains are genetically highly similar, corroborating that a single
601 clone that circumvents Cavendish resistance drives the ongoing pandemic (Ordonez *et al.*, 2015;
602 van Westerhoven *et al.*, 2022b). By contrast, R1 strains are genetically diverse, which is possibly
603 due to prolonged co-evolution of R1 strains with a plethora of genetically diverse banana varieties
604 in the center of origin in Southeast Asia (Perrier *et al.*, 2011). Extensive genetic diversity among
605 *Fusarium* strains that infect the same host is not uncommon (Baayen *et al.*, 2000; van Dam *et al.*,
606 2016; Henry *et al.*, 2021; Fayyaz *et al.*, 2023), yet their ARs and effector profiles are typically
607 remarkably similar (van Dam *et al.*, 2016; Fokkens *et al.*, 2018; Brenes Guallar *et al.*, 2022). We,

608 however, observe that the ARs and effector profile of R1 strains show extensive variation,
609 underscoring their diversity and possibly suggesting that different mechanisms contribute to
610 disease in banana (Henry *et al.*, 2021; Brenes Guallar *et al.*, 2022). *Fusarium* strains that infect
611 strawberry have two distinct ARs, one that contributes to yellowing and the other to wilting (Henry
612 *et al.*, 2021). Quantitative differences in banana corm-discoloration have been noted between R1
613 strains (Garcia-Bastidas, 2019), and our results suggest that this might relate to differences in
614 effector profiles, yet further functional assays are needed to elucidate the link between specific
615 effectors and phenotypic variation.

616 We identify abundant and recent segmental duplications in *F. oxysporum* ARs, as well as
617 aneuploidy of contig 12 in TR4 strain II5. Previous studies have observed large-scale duplications
618 and deletions in the ARs of *F. oxysporum* during *in vitro* chromosome transfer (Vlaardingerbroek
619 *et al.*, 2016; Li *et al.*, 2020), yet our results show that duplication of ARs also occur in *Fusarium*
620 strains sampled directly from infected banana plants. Although variations in chromosome number
621 typically comes with a fitness cost (Todd *et al.*, 2017), it can also provide genetic variation
622 necessary for adaptation, for example increasing virulence or fungicide resistance (Sionov *et al.*,
623 2010; Ropars *et al.*, 2018). We speculate that the observed large-scale duplications play a major
624 role in the origin and evolution of ARs in *Fusarium*. In the wheat pathogen *Zymoseptoria tritici*,
625 chromosomal duplication, together with breakage and fusion, shape the evolutionary dynamics of
626 accessory regions (Croll *et al.*, 2013; Möller *et al.*, 2018). Similar dynamics of chromosome
627 duplication, breakage, and fusion may underly the recent segmental duplications observed in
628 *Fusarium*. For example, the fusion of a (partially) duplicated contig 12 with the arm of chromosome
629 1 in strain II5 could explain the observed homology between the ARs. The process underlying these
630 chromosome dynamics remains unclear. *F. oxysporum* has long been considered to evolve strictly
631 asexually (Gordon & Martyn, 1997), however, the presence of an active sexual cycle in *F.*
632 *oxysporum* populations has recently been proposed (McTaggart *et al.*, 2022; Fayyaz *et al.*, 2023)
633 and we similarly observe reticulation between strains that caused FWB **(**Supplementary Fig. **8**),
634 which supports an evolutionary history in which ancient or infrequent sexual recombination is

635 followed by clonal expansion of selected lineages. In this scenario, incomplete chromosome pairing
636 and non-disjunction during meiosis might drive aneuploidy and ARs diversification (Fayyaz *et al.*,
637 2023). We show that ARs are similar within individual *Fusarium* lineages but are genetically distinct
638 when comparing strains from genetically distant lineages. Individual lineages accumulate genomic
639 changes over time, and this process might be sufficient to explain ARs diversity. If new *Fusarium*
640 lineages indeed arise from a meiotic cycle followed by clonal expansion, recombination, gain, or
641 loss of ARs during meiosis might further explain the diversity of ARs between lineages. Similarly,
642 the absence of meiotic recombination during subsequent clonal expansion would explain the
643 similarity of ARs within clonal lineages. However, we observe aneuploidy within individual clonal
644 lineages, for instance the duplication of contig 12 in some *Fusarium* strains, which demonstrates
645 that aneuploidy can occur independent of a meiotic cycle. Thus, chromosome dynamics in *F.*
646 *oxysporum* require further elucidation and, so far, the mechanisms leading to chromosomal
647 duplications are unclear; they might arise from non-disjunction during mitosis or meiosis (Fayyaz
648 *et al.*, 2023), from incomplete chromosome loss following heterokaryon formation (Harrison *et al.*,
649 2014; Shahi *et al.*, 2016), or through consecutive horizontal chromosome transfers when strains
650 acquire the identical chromosome several times.

651 In contrast to the structural variation in the ARs, the core chromosomes in *Fusarium* are
652 remarkably stable. Although ARs are likely more tolerant to structural variation as they encode
653 fewer essential genes, additional processes can influence genome stability. First of all, the
654 accessory genome is enriched in transposable elements (Figure 2b) and the activity of these
655 elements can contribute to structural variation including gene duplications (Faino *et al.*, 2016;
656 Torres *et al.*, 2021; Wang *et al.*, 2022; Stalder *et al.*, 2022). Possibly the activity of transposable
657 elements in the accessory genomic regions gave rise to the observed segmental duplications, for
658 example, in plant genomes large-scale duplications can arise through double stranded break repair
659 upon TE-induced double stranded breaks (Wicker *et al.*, 2010; Wang *et al.*, 2022). For example,
660 the presence of histone modifications such as trimethylation of lysine 27 on histone 3 (H3K27me3),
661 a histone modification often enriched in ARs (Fokkens *et al.*, 2018; Möller *et al.*, 2019; Torres *et*

662 *al.*, 2020) in fungal plant pathogens, has been linked to genome instability (Seidl *et al.*, 2016; Möller
663 *et al.*, 2019). H3K27me3 also occurs in the ARs of *F. oxysporum* infecting tomato (*Fol*4287), but
664 can also be found in smaller core chromosomes ('fast-core' chromosomes, 9 – 11 (Fokkens *et al.*,
665 2018)) that do not seem to undergo segmental duplications or aneuploidy, and thus additional
666 factors likely influence chromosome stability in *F. oxysporum.*

667 Extensive duplications similarly occur in other plant pathogens (Dutheil *et al.*, 2016; Müller *et*
668 *al.*, 2019; Wyka *et al.*, 2021; Wacker *et al.*, 2023), and are thought to be important drivers in co-
669 evolutionary arms races with their hosts. Understanding the evolution of ARs in *Fusarium* can
670 facilitate the discovery of new effector genes and provide insights into effector diversification.
671 Knowledge of effector profiles is crucial for designing effective disease control strategies and
672 supports the identification of durable resistances in crops (Vleeshouwers *et al.*, 2008).

673

674 **Acknowledgments**

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

681

682 **References**

683 **Armitage AD, Taylor A, Sobczyk MK, Baxter L, Greenfield BPJ, Bates HJ, Wilson F, Jackson AC,**
684 **Ott S, Harrison RJ, *et al.* 2018**. Characterisation of pathogen-specific regions and novel effector
685 candidates in *Fusarium oxysporum* f. sp. *cepae*. *Scientific Reports* **8**: 13530.

686 **Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, Fang Q, Xie D, Feng S,**
687 **Stiller J, *et al.* 2020**. Progressive cactus is a multiple-genome aligner for the thousand-genome
688 era. *Nature* **587**: 246–251.

689 **Ayukawa Y, Asai S, Gan P, Tsushima A, Ichihashi Y, Shibata A, Komatsu K, Houterman PM, Rep**
690 **M, Shirasu K, *et al.* 2021**. A pair of effectors encoded on a conditionally dispensable
691 chromosome of *Fusarium oxysporum* suppress host-specific immunity. *Communications Biology*
692 **4**: 1–12.

693 **Baayen RP, O'Donnell K, Bonants PJM, Cigelnik E, Kroon LPNM, Roebroeck EJA, Waalwijk C**.
694 **2000**. Gene genealogies and AFLP analyses in the *Fusarium oxysporum* complex identify
695 monophyletic and nonmonophyletic formae speciales causing wilt and rot disease.
696 *Phytopathology®* **90**: 891–900.

697 **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI,**
698 **Pham S, Prjibelski AD, *et al.* 2012**. SPAdes: a new genome assembly algorithm and its
699 applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455–477.

700 **Bray NL, Pimentel H, Melsted P, Pachter L**. **2016**. Near-optimal probabilistic RNA-seq
701 quantification. *Nature Biotechnology* **34**: 525–527.

702 **Brenes Guallar MA, Fokkens L, Rep M, Berke L, van Dam P**. **2022**. *Fusarium oxysporum* effector
703 clustering version 2: an updated pipeline to infer host range. *Frontiers in Plant Science* **13**.

704 **Chernomor O, von Haeseler A, Minh BQ**. **2016**. Terrace aware data structure for phylogenomic
705 inference from supermatrices. *Systematic Biology* **65**: 997–1008.

706 **Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, Grimwood J, Schmutz J,**
707 **Taga M, White GJ, Zhou S, *et al.* 2009**. The genome of *Nectria haematococca*: contribution of
708 supernumerary chromosomes to gene expansion. *PLOS Genetics* **5**: e1000618.

709 **Cook DE, Mesarich CH, Thomma BPHJ**. **2015**. Understanding plant immunity as a surveillance
710 system to detect invasion. *Annual Review of Phytopathology* **53**: 541–563.

711 **Coombe L, Li JX, Lo T, Wong J, Nikolic V, Warren RL, Birol I**. **2021**. Longstitch: high-quality
712 genome assembly correction and scaffolding using long reads. *BMC Bioinformatics* **22**: 534.

713 **Croll D, Zala M, McDonald BA**. **2013**. Breakage-fusion-bridge cycles and large insertions
714 contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. *PLOS*
715 *Genetics* **9**: e1003567.

716 **Cuomo CA, Güldener U, Xu J-R, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma L-J, Baker SE,**
717 **Rep M, *et al.* 2007**. The *Fusarium graminearum* genome reveals a link between localized
718 polymorphism and pathogen specialization. *Science* **317**: 1400–1402.

719 **van Dam P, Fokkens L, Ayukawa Y, van der Gragt M, ter Horst A, Brankovics B, Houterman PM,**
720 **Arie T, Rep M**. **2017**. A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection
721 of multiple cucurbit species. *Scientific Reports* **7**: 9042.

722 **van Dam P, Fokkens L, Schmidt SM, Linmans JHJ, Kistler HC, Ma L-J, Rep M**. **2016**. Effector
723 profiles distinguish *formae speciales* of *Fusarium oxysporum*. *Environmental Microbiology* **18**:
724 4087–4102.

725 **Derelle R, Philippe H, Colbourne JK**. **2020**. Broccoli: Combining phylogenetic and network
726 analyses for orthology assignment. *Molecular Biology and Evolution* **37**: 3389–3396.

727 **Dong S, Raffaele S, Kamoun S**. **2015**. The two-speed genomes of filamentous pathogens: waltz
728 with plants. *Current Opinion in Genetics & Development* **35**: 57–65.

729 **Drenth A, Kema G**. **2021**. The vulnerability of bananas to globally emerging disease threats.
730 *Phytopathology®* **111**: 2146–2161.

731 **Dutheil JY, Mannhaupt G, Schweizer G, M K Sieber C, Münsterkötter M, Güldener U,**
732 **Schirawski J, Kahmann R**. **2016**. A tale of genome compartmentalization: the evolution of
733 virulence clusters in smut fungi. *Genome Biology and Evolution* **8**: 681–704.

734 **Edel-Hermann V, Lecomte C**. **2019**. Current status of *Fusarium oxysporum* formae speciales and
735 races. *Phytopathology* **109**: 512–530.

736 **Emms DM, Kelly S**. **2019**. OrthoFinder: phylogenetic orthology inference for comparative
737 genomics. *Genome Biology* **20**: 1–14.

738 **Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GCM, Wittenberg AHJ, Thomma BPHJ**.
739 **2016**. Transposons passively and actively contribute to evolution of the two-speed genome of a
740 fungal pathogen. *Genome Research* **26**: 1091–1100.

741 **FAO**. **2020**. *Fruit and vegetables – your dietary essentials: the international year of fruits and*
742 *vegetables, 2021, background paper*. Rome, Italy: FAO.

743 **Fayyaz A, Robinson G, Chang PL, Bekele D, Yimer S, Carrasquilla-Garcia N, Negash K,**
744 **Surendrarao A, von Wettberg EJB, Kemal S-A, *et al.* 2023**. Hiding in plain sight: genome-wide
745 recombination and a dynamic accessory genome drive diversity in *Fusarium oxysporum* f.sp.
746 *ciceris*. *Proceedings of the National Academy of Sciences* **120**: e2220570120.

747 **Fiddes IT, Armstrong J, Diekhans M, Nachtweide S, Kronenberg ZN, Underwood JG, Gordon D,**
748 **Earl D, Keane T, Eichler EE, *et al.* 2018**. Comparative Annotation Toolkit (CAT)—simultaneous
749 clade and personal genome annotation. *Genome Research* **28**: 1029–1038.

750 **Fokkens L, Shahi S, Connolly LR, Stam R, Schmidt SM, Smith KM, Freitag M, Rep M**. **2018**. The
751 multi-speed genome of *Fusarium oxysporum* reveals association of histone modifications with
752 sequence divergence and footprints of past horizontal chromosome transfer events. : 465070.

753 **Galagan JE, Selker EU**. **2004**. RIP: the evolutionary cost of genome defense. *Trends in Genetics*
754 **20**: 417–423.

755 **Garcia-Bastidas F**. **2019**. Panama Disease in Banana: Spread, Screens and Genes.

756 **García-Bastidas FA, Ordóñez N, Konkol J, Al-Qasim M, Naser Z, Abdelwali M, Salem N,**
757 **Waalwijk C, Ploetz RC, Kema GHJ**. **2014**. First report of *Fusarium oxysporum* f. sp. *cubense*
758 tropical race 4 associated with panama disease of banana outside Southeast Asia. *Plant Disease*
759 **98**: 694–694.

760 **Garcia-Bastidas F, Quintero-Vargas JC, Ayala-Vasquez M, Schermer T, Seidl MF, Santos-Paiva**
761 **M, Noguera AM, Aguilera-Galvez C, Wittenberg A, Hofstede R, *et al.* 2020**. First report of
762 Fusarium wilt tropical race 4 in Cavendish bananas caused by *Fusarium odoratissimum* in
763 Colombia. *Plant Disease* **104**: 994–994.

764 **Gordon TR, Martyn RD**. **1997**. The evolutionary biology of *Fusarium oxysporum*. *Annual Review*
765 *of Phytopathology* **35**: 111–128.

766 **Gu Z, Gu L, Eils R, Schlesner M, Brors B**. **2014**. Circlize implements and enhances circular
767 visualization in R. *Bioinformatics (Oxford, England)* **30**: 2811–2812.

768 **Guo L, Han L, Yang L, Zeng H, Fan D, Zhu Y, Feng Y, Wang G, Peng C, Jiang X, *et al.* 2014**.
769 Genome and transcriptome analysis of the fungal pathogen *Fusarium oxysporum* f. sp. *cubense*
770 causing banana vascular wilt disease. *PLOS ONE* **9**: e95543.

771 **Gurevich A, Saveliev V, Vyahhi N, Tesler G**. **2013**. QUAST: quality assessment tool for genome
772 assemblies. *Bioinformatics (Oxford, England)* **29**: 1072–1075.

773 **Harrison BD, Hashemi J, Bibi M, Pulver R, Bavli D, Nahmias Y, Wellington M, Sapiro G, Berman**
774 **J**. **2014**. A tetraploid intermediate precedes aneuploid formation in yeasts exposed to
775 fluconazole. *PLOS Biology* **12**: e1001815.

776 **Henry PM, Pincot DDA, Jenner BN, Borrero C, Aviles M, Nam M-H, Epstein L, Knapp SJ, Gordon**
777 **TR**. **2021**. Horizontal chromosome transfer and independent evolution drive diversification in
778 *Fusarium oxysporum* f. sp. *fragariae*. *New Phytologist* **230**: 327–340.

779 **Hickey G, Monlong J, Ebler J, Novak AM, Eizenga JM, Gao Y, Marschall T, Li H, Paten B**. **2023**.
780 Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nature*
781 *Biotechnology*: 1–11.

782 **Katoh K, Misawa K, Kuma K, Miyata T**. **2002**. Mafft: a novel method for rapid multiple sequence
783 alignment based on Fast Fourier transform. *Nucleic Acids Research* **30**: 3059–3066.

784 **Kistler HC, Benny U, Boehm EW, Katan T**. **1995**. Genetic duplication in *Fusarium oxysporum*.
785 *Current Genetics* **28**: 173–176.

786 **Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM**. **2017**. Canu: Scalable and
787 accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome*
788 *Research* **27**: 722–736.

789 **Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL**. **2004**.
790 Versatile and open software for comparing large genomes. *Genome Biology* **5**: 1–9.

791 **Latorre SM, Were VM, Foster AJ, Langner T, Malmgren A, Harant A, Asuke S, Reyes-Avila S,**
792 **Gupta DR, Jensen C, *et al.* 2023**. Genomic surveillance uncovers a pandemic clonal lineage of
793 the wheat blast fungus. *PLOS Biology* **21**: e3002052.

794 **Li H**. **2013**. Aligning sequence reads, clone sequences and assembly contigs with BWA-mem.

795 **Li J, Fokkens L, Conneely LJ, Rep M**. **2020**. Partial pathogenicity chromosomes in *Fusarium*
796 *oxysporum* are sufficient to cause disease and can be horizontally transferred. *Environmental*
797 *Microbiology* **22**: 4985–5004.

798 **Lindenbaum P**. **2015**. JVarkit: Java-based utilities for bioinformatics.

799 **Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, Dufresne M,**
800 **Freitag M, Grabherr M, Henrissat B, *et al.* 2010**. Comparative genomics reveals mobile
801 pathogenicity chromosomes in Fusarium. *Nature* **464**: 367–373.

802 **Maryani N, Lombard L, Poerba YS, Subandiyah S, Crous PW, Kema GHJ**. **2019**. Phylogeny and
803 genetic diversity of the banana fusarium wilt pathogen *Fusarium oxysporum* f. sp. *cubense* in the
804 Indonesian centre of origin. *Studies in Mycology* **92**: 155–194.

805 **Maymon M, Sela N, Shpatz U, Galpaz N, Freeman S**. **2020**. The origin and current situation of
806 *Fusarium oxysporum* f. sp.*cubense* tropical race 4 in Israel and the Middle East. *Scientific Reports*
807 **10**: 1590.

808 **McTaggart AR, James TY, Idnurm A, Park RF, Shuey LS, Demers MNK, Aime MC**. **2022**. Sexual
809 reproduction is the null hypothesis for life cycles of rust fungi. *PLOS Pathogens* **18**: e1010439.

810 **Mesny F, Miyauchi S, Thiergart T, Pickel B, Atanasova L, Karlsson M, Hüttel B, Barry KW,**
811 **Haridas S, Chen C, *et al.* 2021**. Genetic determinants of endophytism in the Arabidopsis root
812 mycobiome. *Nature Communications* **12**: 7227.

813 **Möller M, Habig M, Freitag M, Stukenbrock EH**. **2018**. Extraordinary genome instability and
814 widespread chromosome rearrangements during vegetative growth. *Genetics* **210**: 517–529.

815 **Möller M, Schotanus K, Soyer JL, Haueisen J, Happ K, Stralucke M, Happel P, Smith KM,**
816 **Connolly LR, Freitag M, *et al.* 2019**. Destabilization of chromosome structure by histone H3
817 Lysine 27 methylation. *PLOS Genetics* **15**: e1008093.

818 **Möller M, Stukenbrock EH**. **2017**. Evolution and genome architecture in fungal plant pathogens.
819 *Nature Reviews Microbiology* **15**: 756–771.

820 **Müller MC, Praz CR, Sotiropoulos AG, Menardo F, Kunz L, Schudel S, Oberhänsli S, Poretti M,**
821 **Wehrli A, Bourras S, *et al.* 2019**. A chromosome-scale genome assembly reveals a highly
822 dynamic effector repertoire of wheat powdery mildew. *New Phytologist* **221**: 2176–2189.

823 **O'Donnell K, Kistler HC, Cigelnik E, Ploetz RC**. **1998**. Multiple evolutionary origins of the fungus
824 causing panama disease of banana: concordant evidence from nuclear and mitochondrial gene
825 genealogies. *Proceedings of the National Academy of Sciences* **95**: 2044–2049.

826 **Ordonez N, Seidl MF, Waalwijk C, Drenth A, Kilian A, Thomma BPHJ, Ploetz RC, Kema GHJ**.
827 **2015**. Worse comes to worst: bananas and panama disease—when plant and pathogen clones
828 meet. *PLOS Pathogens* **11**: e1005197.

829 **Palmer J, Stajich J**. **2019**. nextgenusfs/funannotate: funannotate v1.5.3.

830 **Perrier X, De Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F, Carreel F, Hippolyte I,**
831 **Horry J-P, Jenny C, *et al.* 2011**. Multidisciplinary perspectives on banana (Musa spp.)
832 domestication. *Proceedings of the National Academy of Sciences* **108**: 11311–11318.

833 **Ploetz RC**. **2006**. Fusarium Wilt of Banana Is Caused by Several Pathogens Referred to as
834 Fusarium oxysporum f. sp. cubense. *Phytopathology®* **96**: 653–656.

835 **Ploetz RC**. **2015**. Management of Fusarium wilt of banana: a review with special reference to
836 tropical race 4. *Crop Protection* **73**: 7–15.

837 **Raffaele S, Kamoun S**. **2012**. Genome evolution in filamentous plant pathogens: why bigger can
838 be better. *Nature Reviews Microbiology* **10**: 417–430.

839 **Rep M, Dekker HL, Vossen JH, de Boer AD, Houterman PM, Speijer D, Back JW, de Koster CG,**
840 **Cornelissen BJC**. **2002**. Mass spectrometric identification of isoforms of PR proteins in xylem sap
841 of fungus-infected tomato. *Plant Physiology* **130**: 904–917.

842 **Ropars J, Maufrais C, Diogo D, Marcet-Houben M, Perin A, Sertour N, Mosca K, Permal E, Laval**
843 **G, Bouchier C, *et al.* 2018**. Gene flow contributes to diversification of the major fungal pathogen
844 *Candida albicans*. *Nature Communications* **9**: 2253.

845 **Rovenich H, Boshoven JC, Thomma BPHJ**. **2014**. Filamentous pathogen effector functions: of
846 pathogens, hosts and microbiomes. *Current Opinion in Plant Biology* **20**: 96–103.

847 **Sánchez-Vallet A, Fouché S, Fudal I, Hartmann FE, Soyer JL, Tellier A, Croll D**. **2018**. The genome
848 biology of effector gene evolution in filamentous plant pathogens. *Annual Review of*
849 *Phytopathology* **56**: 21–40.

850 **Seidl MF, Cook DE, Thomma BPHJ**. **2016**. Chromatin biology impacts adaptive evolution of
851 filamentous plant pathogens. *PLOS Pathogens* **12**: e1005920.

852 **Seidl MF, Thomma BPHJ**. **2017**. Transposable elements direct the coevolution between plants
853 and microbes. *Trends in genetics: TIG* **33**: 842–851.

854 **Shahi S, Beerens B, Bosch M, Linmans J, Rep M**. **2016**. Nuclear dynamics and genetic
855 rearrangement in heterokaryotic colonies of *Fusarium oxysporum*. *Fungal genetics and biology:*
856 *FG & B* **91**: 20–31.

857 **Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM**. **2015**. BUSCO: Assessing
858 genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:
859 3210–3212.

860 **Sionov E, Lee H, Chang YC, Kwon-Chung KJ**. **2010**. *Cryptococcus neoformans* overcomes stress of
861 azole drugs by formation of disomy in specific multiple chromosomes. *PLOS Pathogens* **6**:
862 e1000848.

863 **Stalder L, Oggenfuss U, Mohd-Assaad N, Croll D**. **2022**. The population genetics of adaptation
864 through copy number variation in a fungal plant pathogen. *Molecular Ecology* **n/a**.

865 **Stanke M, Morgenstern B**. **2005**. AUGUSTUS: A web server for gene prediction in eukaryotes
866 that allows user-defined constraints. *Nucleic Acids Research* **33**: W465–W467.

867 **Suyama M, Torrents D, Bork P**. **2006**. Pal2nal: robust conversion of protein sequence
868 alignments into the corresponding codon alignments. *Nucleic Acids Research* **34**: W609-612.

869 **Todd RT, Forche A, Selmecki A**. **2017**. Ploidy variation in fungi: polyploidy, aneuploidy, and
870 genome evolution. *Microbiology Spectrum* **5**: 5.4.09.

871 **Torres DE, Oggenfuss U, Croll D, Seidl MF**. **2020**. Genome evolution in fungal plant pathogens:
872 looking beyond the two-speed genome model. *Fungal Biology Reviews* **34**: 136–143.

873 **Torres DE, Thomma BPHJ, Seidl MF**. **2021**. Transposable Elements Contribute to Genome
874 Dynamics and Gene Expression Variation in the Fungal Plant Pathogen Verticillium dahliae.
875 *Genome Biology and Evolution* **13**: evab135.

876 **Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T,**
877 **Shakir K, Roazen D, Thibault J, *et al.* 2013**. From fastQ data to high-confidence variant calls: the
878 genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics* **43**: 11.10.1-
879 11.10.33.

880 **Vaser R, Sović I, Nagarajan N, Šikić M**. **2017**. Fast and accurate de novo genome assembly from
881 long uncorrected reads. *Genome Research* **27**: 737–746.

882 **Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E,**
883 **Peterson P, Weckesser W, Bright J, *et al.* 2020**. Scipy 1.0: fundamental algorithms for scientific
884 computing in Python. *Nature Methods* **17**: 261–272.

885 **Vlaardingerbroek I, Beerens B, Schmidt SM, Cornelissen BJC, Rep M**. **2016**. Dispensable
886 chromosomes in *Fusarium oxysporum* f. sp. *lycopersici*. *Molecular Plant Pathology* **17**: 1455–
887 1466.

888 **Vleeshouwers VGAA, Rietman H, Krenek P, Champouret N, Young C, Oh S-K, Wang M,**
889 **Bouwmeester K, Vosman B, Visser RGF, *et al.* 2008**. Effector genomics accelerates discovery
890 and functional profiling of potato disease resistance and *Phytophthora infestans* avirulence
891 genes. *PLOS ONE* **3**: e2875.

892 **Wacker T, Helmstetter N, Wilson D, Fisher MC, Studholme DJ, Farrer RA**. **2023**. Two-speed
893 genome evolution drives pathogenicity in fungal pathogens of animals. *Proceedings of the*
894 *National Academy of Sciences* **120**: e2212633120.

895 **Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q,**
896 **Wortman J, Young SK, *et al.* 2014**. Pilon: an integrated tool for comprehensive microbial variant
897 detection and genome assembly improvement. *PLOS ONE* **9**: e112963.

898 **Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, *et al.* 2012**.
899 MCscanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity.
900 *Nucleic Acids Research* **40**: e49.

901 **Wang X, Yan X, Hu Y, Qin L, Wang D, Jia J, Jiao Y**. **2022**. A recent burst of gene duplications in
902 Triticeae. *Plant Communications* **3**: 100268.

903 **Warmington RJ, Kay W, Jeffries A, O'Neill P, Farbos A, Moore K, Bebber DP, Studholme DJ**.
904 **2019**. High-quality draft genome sequence of the causal agent of the current panama disease
905 epidemic. *Microbiology Resource Announcements* **8**: e00904-19.

906 **van Westerhoven AC, Meijer H, Houdijk J, Martínez de la Parte E, Matabuana EL, Seidl M,**
907 **Kema GHJ**. **2022a**. Dissemination of Fusarium wilt of banana in Mozambique caused by
908 *Fusarium odoratissimum* tropical race 4. *Plant Disease*.

909 **van Westerhoven AC van, Meijer HJG, Seidl MF, Kema GHJ**. **2022b**. Uncontained spread of
910 Fusarium wilt of banana threatens African food security. *PLOS Pathogens* **18**: e1010769.

911 **Wicker T, Buchmann JP, Keller B**. **2010**. Patching gaps in plant genomes results in gene
912 movement and erosion of colinearity. *Genome Research* **20**: 1229–1237.

913 **Widinugraheni S, Niño-Sánchez J, Does HC van der, Dam P van, García-Bastidas FA,**
914 **Subandiyah S, Meijer HJG, Kistler HC, Kema GHJ, Rep M**. **2018**. A SIX1 homolog in *Fusarium*

915 *oxysporum* f.sp. *cubense* tropical race 4 contributes to virulence towards Cavendish banana.
916 *PLOS ONE* **13**: e0205896.

917 **Wolf YI, Novichkov PS, Karev GP, Koonin EV, Lipman DJ**. **2009**. The universal distribution of
918 evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent
919 ages. *Proceedings of the National Academy of Sciences* **106**: 7273–7280.

920 **Wyka SA, Mondo SJ, Liu M, Dettman J, Nalam V, Broders KD**. **2021**. Whole-genome
921 comparisons of ergot fungi reveals the divergence and evolution of species within the genus
922 claviceps are the result of varying mechanisms driving genome evolution and host range
923 expansion. *Genome Biology and Evolution* **13**: evaa267.

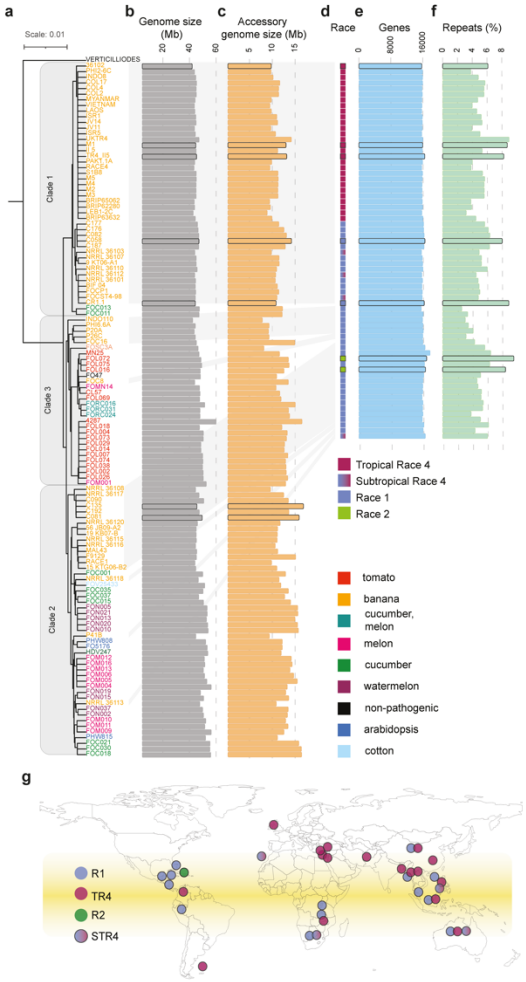
924 **Yang Z**. **2007**. PAML4: phylogenetic analysis by maximum likelihood. *Molecular Biology and*
925 *Evolution* **24**: 1586–1591.

926 **Zhang Y, Yang H, Turra D, Zhou S, Ayhan DH, DeIulio GA, Guo L, Broz K, Wiederhold N,**
927 **Coleman JJ, *et al.* 2020**. The genome of opportunistic fungal pathogen *Fusarium oxysporum*
928 carries a unique set of lineage-specific chromosomes. *Communications Biology* **3**: 1–12.

929 **Zheng S-J, García-Bastidas FA, Li X, Zeng L, Bai T, Xu S, Yin K, Li H, Fu G, Yu Y, *et al.* 2018**. New
930 geographical insights of the latest expansion of *Fusarium oxysporum* f.sp. *cubense* tropical race 4
931 into the greater Mekong subregion. *Frontiers in Plant Science* **9**.

932

933 **Figures and Tables**



934

935 **Figure 1. *Fusarium* strains infecting banana are genetically diverse. a**) *Fusarium* strains that
936 infect banana are polyphyletic (yellow labels). The relationship between *Fusarium* infecting banana
937 and *Fusarium* strains that infect other plant hosts (see label color) is estimated using maximum-
938 likelihood phylogenetic analysis based on 3,811 concatenated conserved single-copy BUSCO
939 genes. Black outline indicates high quality genome assemblies. **b)** Genome sizes of *Fusarium*
940 strains range between 43 to 61 Mb. **c**) The sizes of accessory regions in *Fusarium* strains range
941 from 6.7 to 16.5 Mb. **d**) TR4 strains (purple squares) occur in clade 1 and are genetically similar,
942 as well as the two R2 strains (green). By contrast, the R1 strains (blue squares) are polyphyletic.
943 **e**) *Fusarium* strains that infect banana encode between 15,664 and 17,865 genes and **f**) between
944 2.9 and 8.9% repeats. **g)** The collection of 69 *Fusarium* strains originates from different banana
945 growing regions worldwide; the approximate origin of isolation is indicated by dots, colored
946 according to the race designation (R1, R2, STR4, and TR4). Strains sequenced by long-read
947 sequencing technology are highlighted in black.

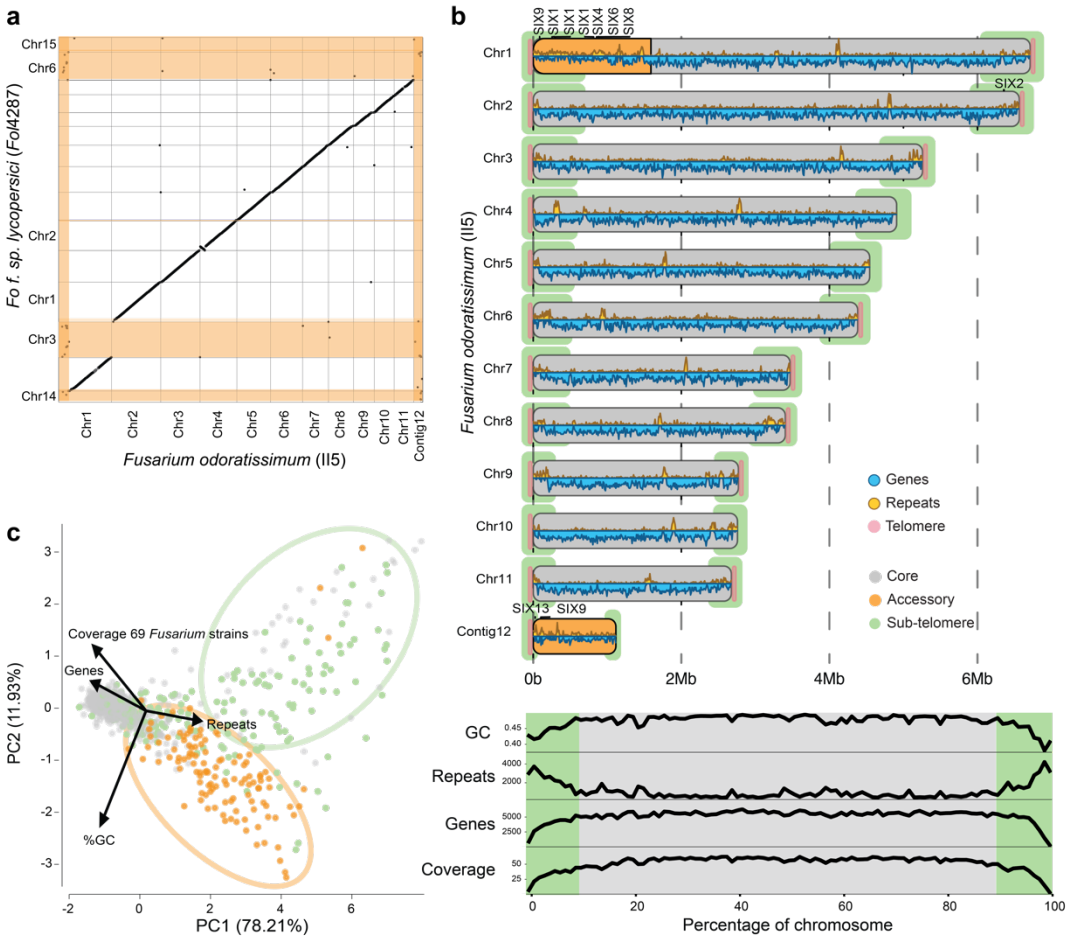


Figure 2. *Fusarium* strains infecting banana carry accessory regions (ARs). a) The TR4 reference strain II5 contains eleven core chromosomes (chromosome 1-11) as well as two ARs (1,8 Mb on chromosome 1 and entire contig 12). Core chromosomes align to *Fol*4287 in contrast to the ARs that are highlighted in orange. b) The ARs (orange) in II5 carry most *S*ecreted *I*n *X*ylem (*SIX*) genes and show an increased repeat and lower gene content. A similar pattern is found in the sub-telomeric regions (green, first and last 10% of the chromosome). 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 contig 12 (orange).

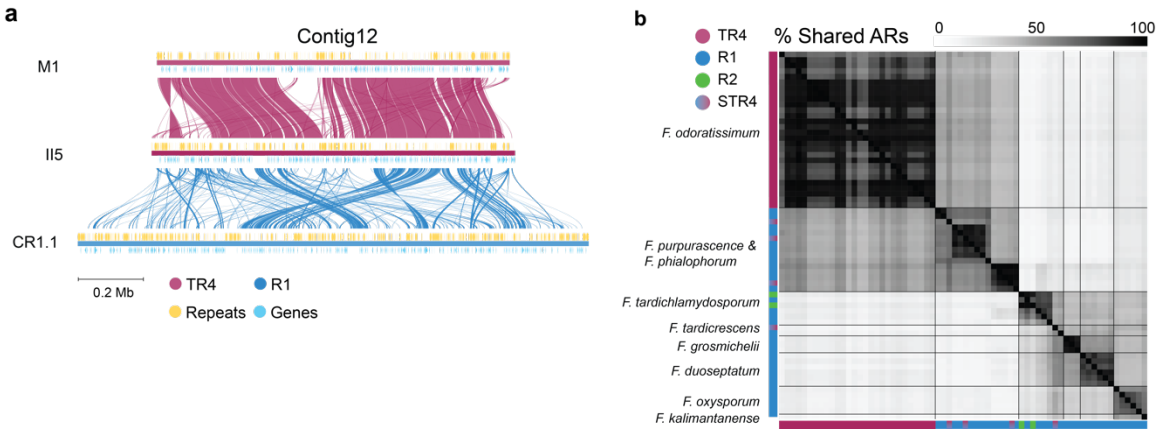


Figure 3. Accessory regions in *Fusarium* strains infecting banana differ significantly. a) Accessory contig 12 of the TR4 reference strain II5 is shared with TR4 strain M1, however this contig shows little similarity to R1 strain CR1.1. Lines (red and blue) indicate aligned regions between strains. 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 (Maryani *et al.*, 2019), share more ARs than distantly related strains.

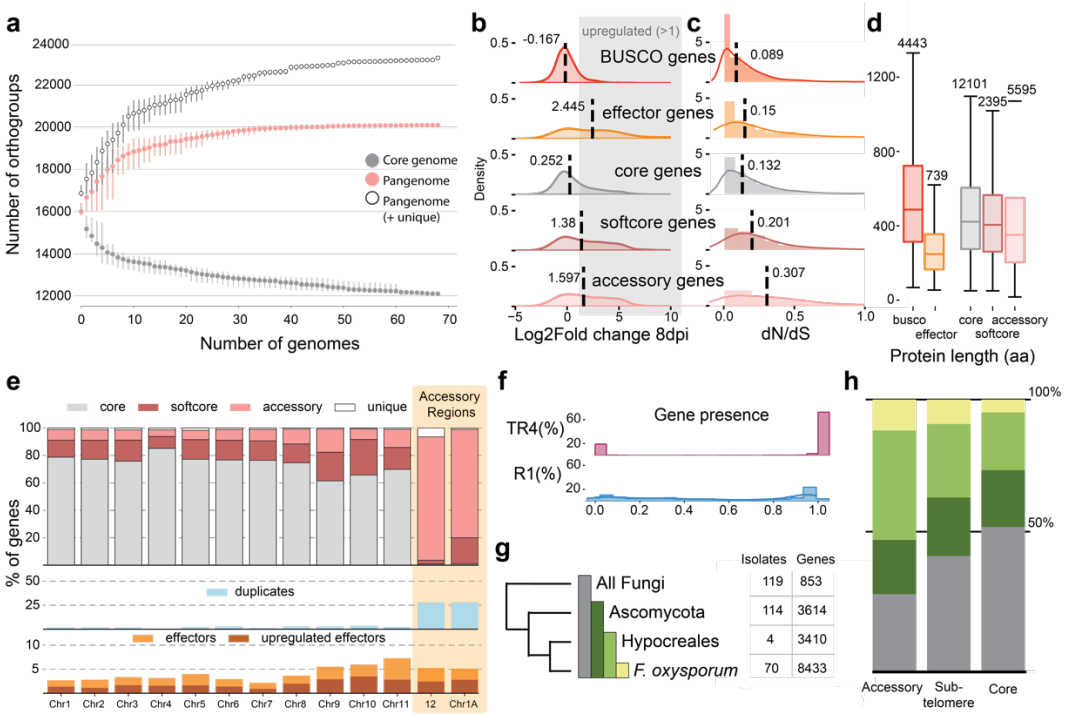


Figure 4. Evolutionary dynamics of genes in the different genomic compartments in *Fusarium* strains infecting banana. a) Global collection of *Fusarium* infecting banana has a pan-genome size of 20,091 orthologous groups (OG, 22,612 when unique genes are included), of which 12,101 are present in all strains (grey). The pan-genome is saturated after adding 40 genomes. The order of genomes was randomly sampled ten times, dots show the average pangenome size and the vertical grey lines indicate the standard deviation. **b**) Accessory genes are upregulated *in planta* at 8 days post inoculation of banana plants. 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). **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)** 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). **e**) ARs in II5 (orange, contig 12 and chromosome 1A) consist mostly of accessory genes (pink, 405 out of 463, 87%) and 28% of these genes are 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 are 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).

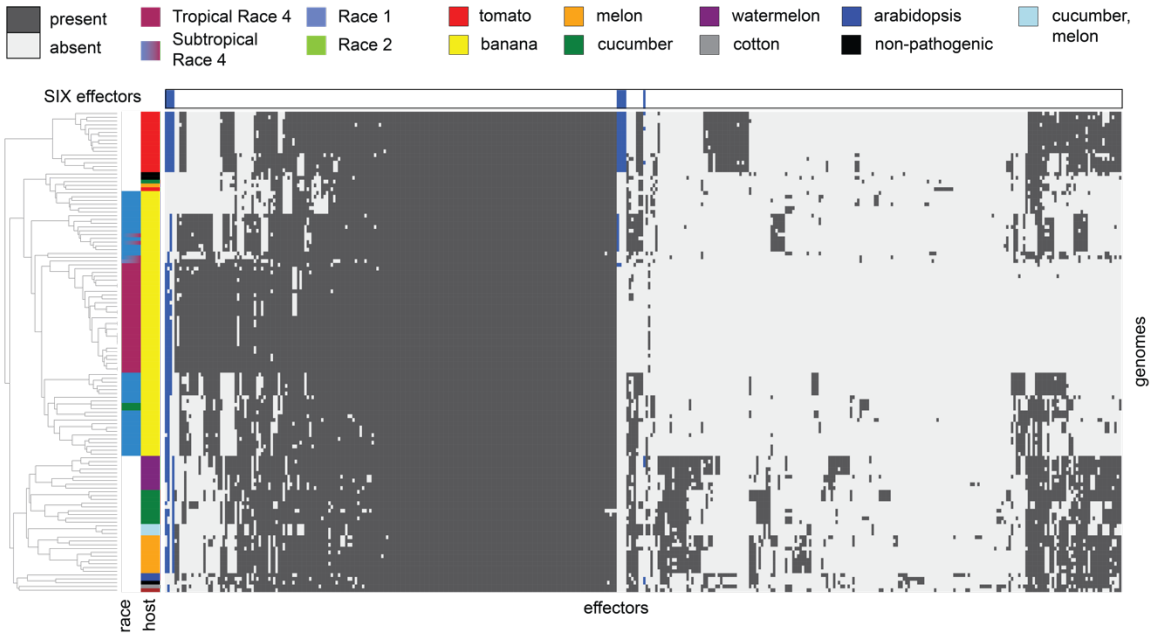


Figure 5. Predicted effector profile clusters *Fusarium* strains infecting the same host. *Fusarium* strains infecting banana (yellow) are separated from *Fusaria* infecting other hosts. In the set of 124 *Fusarium* isolates, 398 MIMP-associated effector families are predicted. The R1 strains (blue) 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.

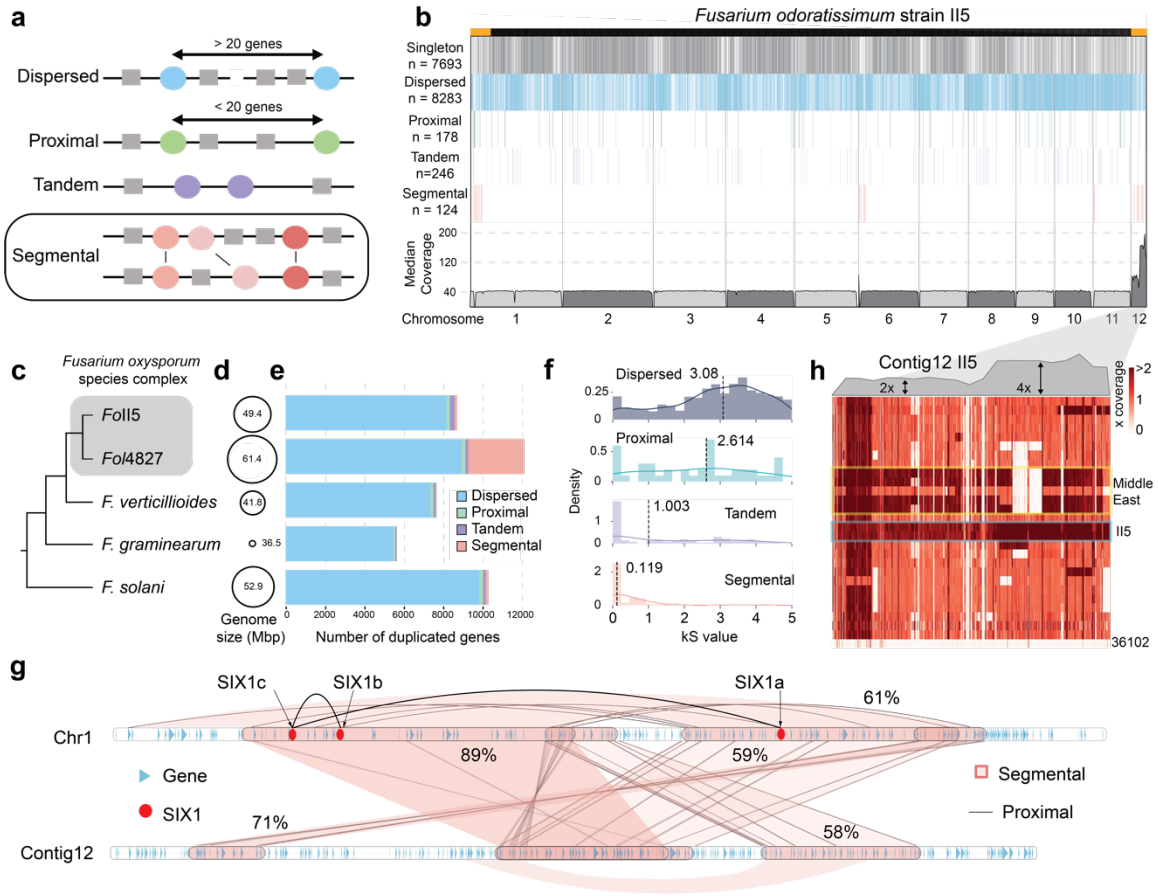


Figure 6. Segmental duplications are involved in the expansion of accessory regions in *F. oxysporum*. a) Four different types of duplications are distinguished. **b**) Segmental duplications (pink) occur mainly in the accessory regions in strain II5 (TR4). Sub-telomeric regions have more dispersed duplicated genes (blue) than the chromosomal arms. On average, the read coverage of II5 mapped onto the II5 reference genome is 40x, whereas contig 12 has a coverage between 80x and 160x, suggesting copy-number variations. **c**) The schematic tree depicts the relationship of *Fusarium* species with **d**) their genome sizes. **e**) Extensive segmental duplications are only found in *F. oxysporum* and *F. solani* and are virtually absent in *F. verticillioides* and *F. graminearum*. **f**) The age of duplications was estimated from the number of synonymous substitutions between duplicated gene pairs. 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). **g**) 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, brown lines between the genes (blue arrows) highlight the genes involved in the segmental duplication. Not all genes in a block are part of the segmental duplication, for example a region on contig 12 is similar to the *SIX1* regions on chromosome 1, yet *SIX1* is absent on contig 12. **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 contig 12.