

# **Complementary Theory of Evolutionary Genetics**

Xiaoqiu Huang  
Department of Computer Science  
Iowa State University  
Ames, Iowa 50011, USA

Corresponding Author:

Xiaoqiu Huang

Email address: xqhuang@iastate.edu

Keywords: Species definition, Speciation, Species maintenance, Supernumerary chromosomes, Horizontal chromosome transfer

## Abstract

This theory seeks to define species and to explore evolutionary forces and genetic elements in speciation and species maintenance. The theory explains how speciation and species maintenance are caused by natural selection acting on non-Mendelian and Mendelian variation, respectively. The emergence and maintenance of species as groups of populations are balanced by evolutionary forces including complementary mechanisms of gene flow within and between populations at population-specific rates: sexual and asexual reproduction, recombining and nonrecombining genome regions, vertical and horizontal DNA transfer, and transposon proliferation and control. While recombining genome regions carry conserved genes and are subjected to meiotic recombination, nonrecombining genome regions carry accessory genes and are not subjected to such structural restraint. Sexual reproduction, vertical DNA transfer, recombining genome regions and transposon control keep species in existence by maintaining recombining chromosome number and structure, while asexual reproduction, horizontal DNA transfer, nonrecombining genome regions and transposon proliferation help species emerge by promoting reproductive isolation and changes in chromosome number and structure. The theory is based on the analysis of the genome sequences of isolates in the *Fusarium oxysporum* complex. The rate of horizontal supernumerary chromosome transfer in this complex was estimated to be 0.1 per genome per year.

## 30 Introduction

31 Darwin (1859) explained the role of natural selection in the origin of species, and  
32 Mendel (1901) discovered the mechanism of inheritance of traits. Fisher, Haldane  
33 and Wright in the 1920s and 1930s developed mathematical models of evolution  
34 as a change in the frequency of gene variants over time. Muller (1932) predicted  
35 that asexual populations accumulate irreversible deleterious mutations. Dobzhan-  
36 sky (1970) proposed that chromosome translocation contributes to the birth of new  
37 species, and Franchini et al. (2020) suggested that chromosomal rearrangements play  
38 an important role in speciation. Coluzzi (1982) proposed a chromosomal speciation  
39 model of suppressed recombination. Growing evidence supports a significant role of  
40 selfish genetic elements in eukaryotic speciation (Werren 2011). Mayr (1942) empha-  
41 sized reproduction isolation in the concept of species; reproductive isolation is the  
42 critical factor behind the emergence of new eukaryotic populations of organisms that  
43 reproduce sexually. Reproduction isolation was linked with variation in *Drosophila*  
44 Y chromosome, a nonrecombining genome region (Hafezi et al. 2020). Asexual re-  
45 production is associated with polyploidy in both plants and animals (Herben et al.  
46 2017). Asexual reproduction can be viewed as a genetic form of reproductive iso-  
47 lation that is common in some groups of eukaryotes such as fungi. An intriguing  
48 question is whether asexual reproduction is the critical factor behind the emergence  
49 of new eukaryotic populations of organisms that reproduce both sexually and asexu-  
50 ally. Is there a concise genetic definition of eukaryotic populations of organisms that  
51 reproduce asexually most of the time? What are the nature and rate of gene flow  
52 within such populations? What kinds of changes in genome structure occur in such  
53 populations? In this study, we attempted to collect evidence for these questions in  
54 a complex group of fungal populations called the *Fusarium oxysporum* speices com-  
55 plex, and propose a theory based on the evidence to improve the current theory of

56 evolutionary genetics.

57 The availability of genome sequence data presents a historic opportunity to ad-  
58 dress classical questions in evolutionary genetics: the evolutionary consequences of  
59 changes in chromosome number and structure (Peichel 2017). Extensive amounts of  
60 genome sequence data are available for isolates in the *F. oxysporum* species complex,  
61 which holds the potential to understand the role of horizontal chromosome trans-  
62 fer in evolution (Kistler et al. 2013). The *F. oxysporum* species complex contained  
63 pathogenic fungal populations for a large number and a wide range of hosts from  
64 plants to animals including humans, where each population had a narrow host range  
65 (van Dam et al. 2016). This implies that a large number of populations were present  
66 in the complex. Although no sexual cycle was observed in the complex, mating-type  
67 loci cloned from this complex was functional in a close sexual relative (Arie et al.,  
68 2000).

69 The genome of the individual or isolate in the complex was composed of the core  
70 genome containing single-copy genes (except ribosomal DNA (rDNA) genes) that  
71 were conserved among the isolates in the complex and the supernumerary genome  
72 containing accessory genes that were present only in certain isolates and that could  
73 have multiple copies in the genome (Covert 1998; Ma et al. 2010; Rep and Kistler,  
74 2010). The supernumerary genome constituted nearly 40% of the genome, based on  
75 the difference in genome size between *F. oxysporum* and *F. graminearum*. Rates of  
76 single nucleotide polymorphisms (SNPs) in conserved genes between isolates were  
77 mostly 0-3%. The core genome was distributed among core chromosomes, while the  
78 supernumerary genome was distributed among supernumerary chromosomes and spe-  
79 cific regions (e.g. the ends) of core chromosomes. The supernumerary genome evolved  
80 more rapidly than the core genome (Raffaele and Kamoun 2012; Croll and McDonald  
81 2012; Dong et al. 2015; Huang et al. 2016). Supernumerary chromosomes have been

shown in *in vitro* experiments to transfer between vegetatively incompatible isolates or to transfer from a pathogenic isolate to a non-pathogenic isolate in asexual filamentous fungi (He et al. 1998; Akagi et al. 2009; Ma et al. 2010; Vlaardingerbroek et al. 2016a; van Dam et al. 2017). In *F. oxysporum*, phylogenetic studies suggest horizontal transfer of supernumerary chromosomes and supernumerary effector genes (van Dam et al. 2016; Fokkens et al. 2018), and supernumerary chromosomes are likely acquired by horizontal transfer through vegetative fusion of hyphae (Eschenbrenner et al. 2020). Several high-quality genome assemblies contained 11 core chromosomes, one or more supernumerary chromosomes, and sometimes keeper chromosomes, which contained large supernumerary regions fused with large core regions of the two smallest core chromosomes that could split into two segments through fission. Chromosome rearrangements generated supernumerary chromosomes in the wheat blast fungus (Langner et al. 2021), and supernumerary regions of core chromosomes in the *Verticillium* wilt fungus were thought to be acquired horizontally (Huang 2014). In this paper, we present a general theory, a particular model of evolution for the *F. oxysporum* complex, and evidence for the model from the analysis of sequencing and genomic data in the complex.

## Methods

We obtained the genome assemblies of the following isolates (by their GenBank assembly accessions) from GenBank at National Center for Biotechnology Information (NCBI): *F. oxysporum* f.sp. *conglutinans* (Foc) race 1 isolate Fo5176 (GCA\_014154955.1), *F. oxysporum* f.sp. *conglutinans* (Foc) race 1 isolate IVC-1 (GCA\_014839635.1), *F. oxysporum* f.sp. *conglutinans* (Foc) race 1 isolate Cong1-1 (GCA\_018894095.1), *F. oxysporum* f.sp. *conglutinans* (Foc) race 2 isolate 54008 (GCA\_000260215.2), *F. oxys-*

106 *porum* f.sp. *conglutinans* (Foc) race 2 isolate 58385 (GCA\_002711385.1), *F. oxyspo-*  
 107 *rum* f.sp. *conglutinans* (Foc) isolate FGL03-6 (GCA\_002711405.2), *F. oxysporum* f.sp.  
 108 *cubense* (Focb) race 1 isolate 160527 (GCA\_005930515.1), *F. oxysporum* f.sp. *cubense*  
 109 (Focb) tropical race 4 (TR4) isolate UK0001 (GCA\_007994515.1), *F. oxysporum* f.sp.  
 110 *lycopersici* (Fol) race 3 isolate D11 (GCA\_003977725.1), *F. oxysporum* f.sp. *lycopersici*  
 111 (Fol) race 2 isolate 4287 (GCA\_001703175.2), *F. oxysporum* f.sp. *melongenae* (Fom)  
 112 isolate 14004 (GCA\_001888865.1), *F. oxysporum* f.sp. *melonis* (Fom) isolate 26406  
 113 Fom001 (GCA\_002318975.1), *F. oxysporum* f.sp. *radicis-cucumerinum* (Forc) iso-  
 114 late Forc016 (GCA\_001702695.2), *F. oxysporum* isolate FISS-F4 (GCA\_004292535.1),  
 115 *F. oxysporum* isolate Fo47 (GCA\_013085055.1), and 99 *F. oxysporum* isolates whose  
 116 genome assemblies produced by Achari et al. (2020).

117 We also obtained the datasets of short reads for the following isolates (by their  
 118 SRA accessions) from Sequence Read Archive (SRA) at NCBI:

119 2 *F. oxysporum* f.sp. *conglutinans* (Foc) race 1 isolates IVC-1 (SRR11823424),  
 120 Cong1-1 HS1 (SRR12709665);

121 1 *F. oxysporum* f.sp. *conglutinans* (Foc) race 2 isolate 58385 (SRR8640621);

122 3 *F. oxysporum* f.sp. *cubense* (Focb) race 1 isolates N2 (SRR550150), VCG0124  
 123 (SRR13311630), VCG0125 (SRR13311629);

124 18 *F. oxysporum* f.sp. *cubense* (Focb) TR4 isolates Col4 (SRR10125423), Col17  
 125 (SRR10747097), Col2 (SRR10103605), FOC.TR4-1 (SRR10054450), FOC.TR4-5 (SRR10054449),  
 126 Hainan.B2 (SRR550152), II-5 (SRR10054446), JV11 (SRR7226881), JV14 (SRR10054448),  
 127 La-2 (SRR7226878), Leb1.2C (SRR7226880), My-1 (SRR7226877), Pak1.1A (SRR7226883),  
 128 Phi2.6C (SRR7226882), S1B8 (SRR10054447), VCG01213/16 (SRR13311628), Vn-2  
 129 (SRR7226879), UK0001 (SRR9733598);

130 13 *F. oxysporum* f.sp. *lycopersici* (Fol) isolates with a total of 155 runs (11-12 runs  
 131 per isolate) CA92/95 (12 runs: SRR307095, SRR307102, SRR307126, SRR307129,

SRR307254, SRR307274, SRR307299, SRR307331, SRR307235, SRR307240, SRR307246,  
 SRR307276), LSU-3 (12 runs: SRR307087, SRR307089, SRR307093, SRR307118,  
 SRR307233, SRR307234, SRR307252, SRR307267, SRR307256, SRR307271, SRR307307,  
 SRR307328), LSU-7 (12 runs: SRR307111, SRR307237, SRR307239, SRR307249,  
 SRR307268, SRR307327, SRR307341, SRR307345, SRR307261, SRR307284, SRR307323,  
 SRR307347), IPO1530/B1 (12 runs: SRR307080, SRR307094, SRR307098, SRR307103,  
 SRR307104, SRR307288, SRR307292, SRR307298, SRR307291, SRR307296, SRR307301,  
 SRR307312), DF0-41 (12 runs: SRR307244, SRR307265, SRR307302, SRR307334,  
 SRR307108, SRR307112, SRR307121, SRR307124, SRR307242, SRR307253, SRR307311,  
 SRR307325), WCS852/E241 (12 runs: SRR307084, SRR307241, SRR307264, SRR307273,  
 SRR307282, SRR307286, SRR307303, SRR307348, SRR307272, SRR307280); SRR307315  
 SRR307324), 14844(M1943) (12 runs: SRR307081, SRR307116, SRR307117, SRR307119,  
 SRR307269, SRR307332, SRR307342, SRR307346, SRR307236, SRR307255, SRR307293,  
 SRR307316), 5397 (12 runs: SRR307083, SRR307085, SRR307109, SRR307110, SRR307120,  
 SRR307260, SRR307314, SRR307319, SRR307247, SRR307277, SRR307310, SRR307333),  
 DF0-40 (12 runs: SRR307088, SRR307099, SRR307105, SRR307125, SRR307127,  
 SRR307245, SRR307309, SRR307339, SRR307279, SRR307313, SRR307321, SRR307344),  
 DF0-23 (12 runs: SRR307106, SRR307107, SRR307113, SRR307115, SRR307123,  
 SRR307238, SRR307257, SRR307266, SRR307295, SRR307297, SRR307322, SRR307336),  
 DF0-38 (12 runs: SRR307086, SRR307090, SRR307091, SRR307092, SRR307122,  
 SRR307250, SRR307262, SRR307278, SRR307281, SRR307306, SRR307320, SRR307326),  
 DF0-62 (12 runs: SRR307100, SRR307101, SRR307128, SRR307130, SRR307248,  
 SRR307258, SRR307259, SRR307270, SRR307275, SRR307300, SRR307317, SRR307343),  
 MN-14 (11 runs: SRR307082, SRR307096, SRR307097, SRR307243, SRR307263,  
 SRR307285, SRR307290, SRR307318, SRR307335, SRR307338, SRR307340).

Each paired-end dataset of short reads was represented by a pair of Fastq files in

158 compressed format. For example, the names of the two files for a dataset with SRA  
159 accession SRR3139043 were SRR3139043\_1.fastq.gz and SRR3139043\_2.fastq.gz. To  
160 process many datasets in a batch mode, their pairs of files were placed in a data  
161 directory. A Linux shell script was written to go through each pair of files and to  
162 call a Perl script to map the files of short reads onto a reference genome assembly.  
163 The path of the data directory was included in the Perl script. The Perl script  
164 takes as input an SRA accession number and the name of a fasta file containing the  
165 reference genome assembly in the working directory. Then it calls Bowtie2 (Langmead  
166 and Salzberg 2012) to map the two files of short reads onto the reference genome  
167 assembly, generating an alignment output file in BAM format. Next the Perl script  
168 calls Picard to transform the BAM alignment file and calls GATK with command  
169 option HaplotypeCaller to produce a file (in VCF format) of SNPs and indels between  
170 the short reads and the reference. After that, it calls Bedtools (Quinlan and Hall 2010)  
171 with command option genomecov to report a file (whose name ends in '.cov') of reads  
172 depths at each reference genome position and to report a file of reference regions  
173 with zero coverage in BedGraph format. Finally, the Perl script calls a custom AWK  
174 (named 'z.cov.awk') script to calculate the average read depth for the reference from  
175 the file of reads depths and calls another custom AWK script to calculate a total size  
176 of reference regions of at least 5 kb with zero coverage for each chromosome or contig  
177 in the reference genome assembly from the BedGraph file.

178 The z.cov.awk AWK script for calculating the average read depth for the reference  
179 from the .cov file was based on the following algorithm. The reference can be a  
180 whole genome assembly or a chromosome. If the reference is a chromosome with a  
181 subtelomere element at an end, then the subtelomere element would be covered at  
182 high depths by short reads from the multiple copies of the subtelomere element in  
183 the genome. These high depths of coverage would inflate the average depth of the



184 chromosome. To address this problem, we used the standard formula for calculating  
185 the original mean and standard deviation of the read depths for the reference. Then we  
186 calculated a revised mean by using only those reference positions whose read depths  
187 were not more than the original mean plus three times the standard deviation. The  
188 revised mean was reported in a file whose name ends in ‘.average2’ by the z.cov.awk  
189 AWK script from the the .cov file.

190 To determine if an isolate with a dataset of short reads belong to the asexual  
191 population of a reference isolate with a genome assembly in a fasta sequence file, the  
192 fasta sequence file was processed by a custom AWK script to produce an output file  
193 (whose name ends in ‘.ATareas’) of AT rich regions of sizes at least 2 kb, where each  
194 region is composed of multiple consecutive lines of the fasta sequence file such that  
195 the AT content of each line is at least 65%. Then another custom AWK script was  
196 written to calculate the percentage of the AT rich regions that was covered by the  
197 dataset of short reads at read depths of at least 5, where the AWK script takes as  
198 input the .ATareas file and the .cov file. The isolate belongs to the asexual population  
199 of the reference isolate if the percentage of the reference AT rich regions covered by  
200 short reads from the isolate was above a cutoff, say, 10%. In addition, the SNP rate  
201 between the short read isolate and the reference isolate was calculated as the number  
202 of SNPs with a read depth of at least 10 (given in the DP field in the VCF file) and a  
203 quality value of at least 80 (given in column 6 of the VCF file) divided by the number  
204 of reference positions with a read depth of at least 10 (obtained by using the .cov  
205 file).

206 Candidate genes in a genome assembly were found by AUGUSTUS (Stanke and  
207 Waack 2003). Functional annotation of predicted protein sequences wer performed  
208 by HMMER (Finn et al. 2011). Gap-free matches within a genome assembly were  
209 computed by the DDS2 program (Huang et al. 2004). The output from the program

210 was filtered by an AWK script to select gap-free matches of at least 5 kb in length  
211 and at least 99% in identity. This procedure was applied to the genome assemblies  
212 of isolates Fo47 and Foc Fo5176.

213 A genome assembly may still contain subtelomeres that were inactivated by RIP  
214 during the last sexual cycle. To find such subtelomeres in the genome assembly,  
215 we used the AAT package (Huang et al. 1997) to search the genome assembly for  
216 matches to an ATP-dependent DNA helicase *hus2/rqh1* of 1,428 residues (accession:  
217 KAG7001869.1) encoded in isolate Fo5176 supernumerary chromosome 18. Genome  
218 sequences that were similar to the DNA helicase were potential subtelomeres as they  
219 encoded a helicase, a signature of subtelomeres. The DNA-protein alignment pro-  
220 duced by AAT was used to count the number of stop codons in the reading frame of  
221 the DNA sequence, where each stop codon in the reading frame was marked with three  
222 stars. Note that some of the matches might be intact subtelomeres that contained  
223 no stop codons in their reading frames. For example, the sequences of the 3' 30-kb  
224 ends of isolate Forc Forc016 chromosomes 4 and 10 were similar and contained intact  
225 subtelomeres adjacent to inactivated subtelomeres, as indicated by an alignment of  
226 these sequences produced by the SIM program (Huang and Miller 1991), where all  
227 434 base mismatches in the alignment were located in the inactivated subtelomeres.  
228 This also happened to those of Forc016 chromosomes 4 and contig 7.

229 All isolates with sufficiently high percent coverage of the AT rich regions in a  
230 reference isolate had low SNP rates with the reference isolate. Conversely, all isolates  
231 with high SNP rates with a reference isolate had low high percent coverage of the AT  
232 rich regions in the reference isolate. This association means that a measure of SNP  
233 rate is useful in predicting the mode of reproduction in isolates. The distribution  
234 of SNP rates between isolates can be used to predict how long asexual reproduction  
235 lasts. Assume that 10 SNPs occurred per genome per year. Then two isolates with

236 thousands of SNPs were estimated to diverge hundreds of years ago. If a sample of  
237 isolates from a population were estimated to have hundreds of SNPs among them,  
238 then the population was estimated to have existed for decades.

## 239 **Results**

### 240 **Theory**

241 This theory explains how species maintenance is caused by natural selection acting  
242 on Mendelian variation in gene structure, and how speciation is caused by natural se-  
243 lection acting on non-Mendelian variation in chromosome number and structure. The  
244 emergence and maintenance of species as groups of populations are balanced by evolu-  
245 tionary forces including complementary mechanisms of gene flow within and between  
246 populations at population-specific rates: sexual and asexual reproduction, recom-  
247 bining and nonrecombining genome regions, vertical and horizontal DNA transfer,  
248 and transposon proliferation and control. Nonrecombining genome regions include  
249 B and sex chromosomes in plants and animals, and supernumerary chromosomes  
250 (also called dispensable, lineage-specific, or accessory chromosomes) in fungi. While  
251 core chromosomes carry conserved genes and are subjected to meiotic recombination,  
252 nonrecombining genome regions carry accessory genes and are not subjected to such  
253 structural restraint. Sexual reproduction, vertical DNA transfer, recombining genome  
254 regions and transposon control play major roles in maintaining chromosome number  
255 and structure, while asexual reproduction, horizontal DNA transfer nonrecombining  
256 genome regions and transposon proliferation are main genetic factors behind repro-  
257 ductive isolation and changes in chromosome number and structure. Nonrecombining  
258 genome regions are enriched in genes involved in genome dynamics, adaptation to en-  
259 vironments and reproductive isolation, where some of those genes arise by horizontal

gene transfer, which is an ongoing evolutionary force during asexual reproduction. Thus, eukaryotic populations, especially populations of organisms that reproduce both sexually and asexually, emerge and adapt by undergoing changes more frequently in nonrecombining genome structure than in recombining chromosome structure. The theory is based on a model of evolution for populations in the *F. oxysporum* complex. The model was formulated by studying changes in chromosome number and structure within and between populations in the complex. In this complex, for example, the generation and transfer of supernumerary chromosome structural variants alongside the formation of population-specific subtelomeric palindromes at the ends of chromosomes allow pathogenic fungal populations to emerge and evolve during asexual reproduction. Some of those populations contained fusions between core and supernumerary chromosomes, as well as translocations between core chromosomes, which could be potential barriers to meiotic recombination. The rate of horizontal supernumerary chromosome transfer was  $1/y$  per genome per year, where  $y$  was the number of years for the isolate to acquire population-specific subtelomeric palindromes at the ends of its chromosomes for the first time. The parameter  $y$  was estimated to be less than 10 years.

## Model of evolution for populations in the *F. oxysporum* species complex

Variation in chromosome structure is a major driver of divergence and speciation, and stability in chromosome structure is a major keeper of species in existence. In this sense, speciation and species maintenance are in conflict, and the evolutionary and molecular mechanisms of speciation and species maintenance are also in conflict or complementary. For example, selfish genetic elements are in conflict with other genes in the eukaryotic genome (Werren 2011). Possible complementary modes and mecha-

nisms include sexual and asexual reproduction, recombination and nonrecombination,  
 vertical and horizontal DNA transfer, core and supernumerary chromosome, strong  
 and weak selection, rapid and slow gene flow, rapid and slow mutation, transposon  
 proliferation and control, change in the frequency of gene variants and chromosome  
 structural variants, variation and stability in chromosome number, genes in single  
 and multiple copies, chromosomes in single and multiple copies, genes with benefits  
 in general and particular environments, large and small genomes, large and small  
 numbers of individuals, intact and mutated subtelomeres, and and haploidy, diploidy  
 and polyploidy. Most importantly, speciation and species maintenance are strongly  
 affected by the mode of reproduction. Sexual reproduction is effective at maintaining  
 species by preserving their chromosome structure through recombination, but less  
 efficient for variation in chromosome structure to occur. Because of lack of recom-  
 bination, asexual reproduction is efficient for variation in chromosome structure to  
 occur, but less effective at maintaining species by preserving their chromosome struc-  
 ture. The weaknesses in sexual and asexual reproduction are compensated by using  
 mixed modes of reproduction and/or other complementary mechanisms. In general,  
 speciation and species maintenance are balanced by those complementary modes and  
 mechanisms at appropriate frequencies and rates.

We explain how speciation and species maintenance in the *F. oxysporum* species  
 complex were balanced by complementary modes and mechanisms at appropriate  
 frequencies and rates. The complex was estimated to have existed for at least 10  
 thousand years. In this complex, two types of genome were formed to maintain the  
 complex and to allow different populations to adapt to diverse environments (Croll  
 and McDonald, 2012; Raffaele and Kamoun, 2012). The core genome was intended  
 to carry genes with benefits in all environments, and the supernumerary genome to  
 carry genes with benefits in particular environments. The core genome was conserved

among isolates, while the supernumerary genome was highly variable among isolates. The stability in the core genome was maintained by strong purifying selection as well as infrequent sexual reproduction, during which the proliferation of transposable and repetitive elements was controlled by Repeat Induced Point Mutation (RIP) (Cambareri et al. 1989; Gladyshev 2017). The infrequent sexual reproduction offered little time for a change in the frequency of conserved gene variants to occur in the complex. The variation in the supernumerary genome was promoted by the frequent horizontal transfer of supernumerary chromosomes carrying transposons within and between populations, where the rate of horizontal transfer was more frequent than sexual reproduction. Haploidy was the predominant state in this complex, with single-copy conserved genes (except for rDNA genes) in core chromosomes, where deleterious mutations in conserved genes were removed more effectively than in two-copy genes, since their effects were not shielded (Orr and Otto 1994). On the other hand, genes in supernumerary chromosomes were under weaker purifying selection and could be present in multiple copies, and differences in supernumerary chromosome number and structure were prevalent among isolates. Selection plays a major role in removing supernumerary chromosomes with deleterious mutations from the complex and in increasing the frequency of beneficial supernumerary chromosomes in the complex. The construction of accurate phylogenetic trees from core chromosomes of isolates across the complex (Achari et al. 2020; Fokkens et al. 2020) revealed insignificant flow of core chromosomes in the complex either vertically or horizontally.

The genomes of isolates in the complex contained the hallmarks of RIP in the form of C-T and G-A mutations in and around repetitive elements (except rDNA repeats), which is regarded as the signature of past sexual reproduction, since RIP occurs during sexual cycles. Because sexual reproduction in the complex was infrequent and short, it may be effective in maintaining chromosome number and structure through

337 recombination and in controlling transposon proliferation through RIP, but it may  
338 not be able to contribute significantly to generating recombinations of gene variants.  
339 Therefore, most of the population adaptation and divergence in the complex occurred  
340 during long periods of asexual reproduction.

341 All core and supernumerary and keeper chromosomes were flanked on both sides  
342 by inverted copies of a population-specific subtelomeric element, called a subtelomeric  
343 palindrome. All isolates with subtelomeric palindromes of the same kind belonged  
344 to the same population, which persisted on specific hosts or in specific environments.  
345 One reason why all periods of sexual reproduction were short is that isolates un-  
346 dergoing sexual reproduction could not persist on any hosts or in any environments  
347 for a long time because their host- or environment-specific supernumerary genomes  
348 were all inactivated by RIP during sexual cycles. The complex contained a diverse  
349 set of supernumerary chromosomes distributed in a large number of populations.  
350 The supernumerary genome was enriched in transposable and repetitive elements,  
351 pathogenicity genes and HET (heterokaryon incompatibility) domain genes (Paoletti  
352 and Clavé 2007; Vlaardingerbroek et al. 2016b).

353 The lineage for each isolate was comprised of long periods of asexual reproduction  
354 and short periods of sexual reproduction. During a period of sexual reproduction,  
355 the duplicated regions of the supernumerary genome were inactivated by RIP into  
356 AT-rich regions that contained no functional genes; such duplicated regions included  
357 transposons and subtelomeric palindromes. At the beginning of a period of asexual  
358 reproduction, in order to survive on a host or in an environment, the lineage had  
359 to reconstruct its supernumerary genome quickly by horizontally acquiring an in-  
360 tact supernumerary chromosome with a population-specific subtelomeric palindrome  
361 and duplicating the palindrome at the end of each core chromosome. During the  
362 period of asexual reproduction, within the nucleus, gene exchanges between supernu-

merary chromosomes and core chromosomes occurred through the homology of the subtelomeric palindromes at the ends of these chromosomes, and transposons along with pathogenicity genes moved from supernumerary chromosomes to AT-rich regions of core chromosomes.

Gene flow manifested in the form of the horizontal transfer of supernumerary chromosomes within and between populations during asexual reproduction. The horizontal transfer of a supernumerary chromosome within a population led to isolates with two or more structurally different copies of the supernumerary chromosome. These copies underwent chromosome rearrangements so that deleterious variants were lost and beneficial variants became more prevalent within the population. When an isolate from one population came into contact with an isolate from another population, only supernumerary chromosomes might move from one isolate to the other; the core chromosomes in one isolate could be separated from those in the other isolate based on the differences in the sequences of their subtelomeric palindromes, which was one of the reasons why all core chromosomes in an isolate were flanked on both sides by subtelomeric palindromes of its population-specific type. Note that phylogenetic trees of isolates in the complex were accurately constructed on sequences from core chromosomes, indicating that the extent to which core chromosomes from different populations were mixed through horizontal transfer during a long period of asexual reproduction was minimal, and also indicating the absence of extensive meiotic recombination in this complex. This absence of extensive meiotic recombination means that the evolution of this species complex was reflected to a lesser extent through change in conserved gene frequency. Instead, the horizontal transfer of supernumerary chromosomes is proposed as a major driver of the evolution of this species complex through change in genome structure. Gene flow within and between populations during asexual reproduction was controlled by over 100 HET domain genes.



A new population of pathogenic isolates could emerge for a host through the horizontal transfer of a supernumerary chromosome as follows. First, the host (called *a*) developed resistance to a population *A* of pathogenic isolates. Then a new supernumerary chromosome emerged from another population *B* (for a different host *b*) that contained multiple supernumerary and keeper chromosomes. Next the new supernumerary chromosome with the *B*-specific subtelomeric palindrome arrived in isolates of population *A* through horizontal transfer, and subsequently underwent subtelomeric palindrome changes from type *B* to type *A*. After that, the old and new supernumerary chromosomes in population *A* underwent gene exchanges so that a resulting supernumerary chromosome had a new subtelomeric palindrome different from that of population *A* and could cause disease to host *a*. Finally, the new population was founded when the resulting supernumerary chromosome arrived in isolates of a population and caused the core chromosomes of those isolates to obtain its subtelomeric palindrome. With asexual reproduction, the new population underwent growth and expanded by spreading its supernumerary chromosome to isolates of other populations through horizontal transfer. The above description was based on the similarity between supernumerary chromosomes from several real populations, and the observation that the subtelomeric palindrome of one population were found in a supernumerary chromosome in another population.

During a period of asexual reproduction, certain deleterious mutations such as nucleotide substitutions in core chromosomes from some nuclei could be removed through mitotic recombination with matching core chromosomes from other nuclei within the individual (Nieuwenhuis and James 2016). However, when the AT-rich regions of core chromosomes were all taken up by active transposons, deleterious mutations to core chromosomes caused by the proliferation of transposons in all nuclei within the individual could no longer be fixed through mitotic recombination. This

415 caused the current period of asexual reproduction to be terminated. A new short  
 416 period of sexual cycles was needed to control the proliferation of transposons by  
 417 RIP, which is known to induce massive point mutations in duplicated regions rapidly.  
 418 These sexual cycles also maintained the number and structure of core chromosomes  
 419 through meiotic recombination. Because transposons arrived on supernumerary chro-  
 420 mosomes during a period of asexual reproduction, the composition and transfer rate  
 421 of supernumerary chromosomes affected the rate at which structural variation was  
 422 generated and the length of this asexual period. Note that the length of existence  
 423 of a pathogenic population depended more on the availability and susceptibility of  
 424 its host than on the inability to purge deleterious mutations during asexual repro-  
 425 duction. An explanation for this is that the horizontal transfer of supernumerary  
 426 chromosomes would allow the population to expand in young asexual populations if  
 427 the current asexual populations carrying the population had lasted for a long period  
 428 of time and accumulated deleterious mutations.

429 The rate of horizontal transfer was estimated based on the above model. At least  
 430 one horizontal transfer was needed for any asexual lineage to acquire a population-  
 431 specific supernumerary chromosome in order to persist in the environment for the  
 432 population. Let  $y$  be the longest length in years of asexual lineages in the *F. oxys-*  
 433 *porum* complex to acquire their first population-specific supernumerary chromosomes  
 434 after sexual reproduction. Then the rate of horizontal transfer for supernumerary  
 435 chromosomes between populations was bounded from below by  $1/y$  per genome per  
 436 year. If  $y$  was less than 10 years, then the rate of transfer was greater than 0.1 per  
 437 genome per year. An argument in favor of a small value for the parameter  $y$  could be  
 438 supported by the inability to find natural isolates without intact subtelomeric palin-  
 439 dromes. Let  $\lambda$  be the longest length in years of asexual lineages in the *F. oxysporum*  
 440 complex. The rate of transposition for transposons was bounded from below by  $\tau/\lambda$ ,

441 where  $\tau$  is the minimum number of intact transposons in core chromosomes in any  
442 asexual lineage. If  $\lambda_i$  was the length in years of the asexual lineage for isolate  $i$  with  
443 a population-specific subtelomeric palindrome, then the rate of horizontal transfer  
444 within the population for this lineage was  $c_i/\lambda_i$  per genome per year, where  $c_i$  was  
445 the number copies of a supernumerary chromosome in isolate  $i$ .

446 In the complex, dynamic genetic populations were carried by static asexual popula-  
447 tions. As some asexual populations became old, they passed their genetic populations  
448 to young asexual populations through the horizontal transfer of the supernumerary  
449 chromosomes in the genetic populations. Then the old asexual populations under-  
450 went sexual cycles to become young asexual populations. A major role of sexual  
451 reproduction was to produce young asexual populations with the correct number and  
452 structure of core chromosomes and with fewest functional transposons so that those  
453 young asexual populations would be able to carry dynamic genetic populations. Note  
454 that gene duplication, which is an important mechanism by which evolution occurs,  
455 was restricted to the supernumerary genome. An explanation for this restriction is  
456 that the core genome went through sexual cycles, which would inactivate all dupli-  
457 cated genes, so that only the unique core genome could stay intact during sexual  
458 cycles. Supernumerary chromosomes with duplicated genes were transferred from old  
459 asexual populations to young ones, bypassing sexual cycles so that they were not  
460 subjected to RIP. Note that supernumerary chromosomes that remained in isolates  
461 undergoing sexual cycles would be inactivated by RIP and then were lost. Trans-  
462 posons in supernumerary chromosomes in the wheat blast fungus lacked signature of  
463 RIP (Peng et al. 2019).

## Evidence: Population-specific subtelomeres and AT rich regions

We previously found that in isolates Fol race 3 D11, Forc Forc016, Fom Fom001 and Focb race 1 160527, core and supernumerary chromosomes were flanked on both sides by inverted copies of a host- or population-specific subtelomeric element (Huang 2019). That is, the 5' copy in forward orientation was highly similar to the 3' copy in reverse orientation, with both copies located within 10- to 15-kb ends of the chromosome, respectively. This pair of inverted copies was called a subtelomeric palindrome here. The finding also held for isolates Fo47, Foc Fo5176 and Focb TR4 UK0001. The subtelomeric element contained a gene encoding a helicase, where the reading frame of the gene was free of stop codons (Huang 2019).

We examined AT rich regions in the genomes of the 15 *F. oxysporum* isolates listed in Methods. Of the 15 isolates, 5 isolates (Fo47, Foc Fo5176, Fol D11, Focb 160527 and Focb TR4 UK0001) had a high-quality genome assembly; in each of these 5 isolates, the AT content of each core and supernumerary chromosome was around 52%. On the other hand, the genome assemblies of all the 15 isolates contained AT rich regions of total sizes ranging from 511.3 to 2,270.1 kb with a mean of 1,028.9 kb, where an AT rich region was a block of consecutive lines of sequences with the AT content of each line at 65% or higher and with the size of the block at 2 kb or larger.

Some AT rich regions were inactivated subtelomeric palindromes containing helicase genes with many stop codons in their reading frames. For example, isolate Fo47 chromosome 2 contained a pair of inactivated subtelomeric elements that were 15 to 16 kb away from the respective chromosome ends, each of which contained an intact subtelomere with no stop codons in its reading frame. The 5' subtelomere had an AT content of 73% over a length of 3.7 kb, and was 45% identical over 1313 residues to an ATP-dependent DNA helicase *hus2/rqh1* of 1,428 residues (ac-

cession: KAG7001869.1) encoded in isolate Fo5176 supernumerary chromosome 18. This DNA-protein alignment contained 100 stop codons in the reading frame of the 5' subtelomere. Similarly, the 3' subtelomere had an AT content of 69% over a length of 1.4 kb, and was 37% identical over 593 residues to the same helicase protein, with 37 stop codons in its reading frame. This example suggested that an ancestor of Fo47 chromosome 2 contained inactivated subtelomeres at both ends and later acquired an intact subtelomere attached to each of its ends, where the intact subtelomere was located within the new end of length 15 kb.

In another example, isolate Forc Forc016 supernumerary chromosome RC contained a pair of intact subtelomeric elements, each of whose reading frames was 90% identical over 984 residues to the above helicase protein with no stop codons in the reading frame. This also happened with the intact subtelomeric elements in the core chromosomes of this isolate. Those similarities could explain the origin of the current set of subtelomeric palindromes in this isolate. Moreover, the isolate contained several inactivated subtelomeric palindromes that were similar to the helicase protein with many stop codons in their reading frames. For instance, we found two 3' inactivated subtelomeres that were 91.7% identical over 4.0 kb, adjacent to two intact subtelomeres with a nearly perfect match over 13 kb. A similar case also happened to another two 3' inactivated subtelomeres next to two intact ones. These examples suggested that inact subtelomeres were attached to ancetor chromosome ends containing inactivated subtelomeres. Lastly, isolate Focb race 1 160527 contig 7 of 4.2 Mb contained a 3' inactivated partial 4.0-kb subtelomere with an AT content of 68% that was 54% identical over 1,276 residues to the helicase protein with 78 stop codons in the reading frame, and isolate TR4 UK0001 contig 4 of 5.2 Mb contained an 5' inactivated partial 3.7-kb subtelomere that was 42% identical over 1,078 residues to the helicase protein with 93 stop codons in the reading frame. Note that these two

subtelomeres were located at 14 kb away from the 3' end and at 21 kb away from the 5' end, respectively. Those AT regions were evidence for past RIP activities during sexual cycles, which mutated repetitive elements including subtelomeric palindromes.

Extremely high and variable rates of accessory chromosome loss were reported in the plant pathogenic fungus *Zymoseptoria tritici*, which reproduces both sexually and asexually, with sexual cycles observed. In *Z. tritici* isolate ST00Arg\_1D1a1, a 3' subtelomere of 11.4 kb of chromosome 12 was 94-95% homologous to the subtelomeres of 14 other chromosomes with 99.4% of 22,679 substitutions being G-A and C-T mutations, the signature of the RIP process during a sexual cycle. A longest region with an AT content of  $\geq 65\%$  in the *Z. tritici* isolate ST00Arg\_1D1a1 was 11 kb, and that in the *F. oxysporum* isolate UK0001 was 149 kb.

## **Evidence: Gene duplication and structural variation**

We considered gene duplication and structural variation in isolate Fo47, Foc race 1 isolates Fo5176 and IVC-1, each of which had a high-quality genome assembly. A hallmark of asexual reproduction in *F. oxysporum* pathogenic isolates was extensive duplication of accessory genes within the genome (within and between supernumerary chromosomes and the ends of core chromosomes). For example, we found more than 46,000 gap-free matches of at least 5 kb in length and at least 99% identity in the genome assembly of isolate Fo5176, and as a comparison, 18 such matches in the non-pathogenic isolate Fo47. All of these 18 matches in isolate Fo47 were between subtelomeres and other supernumerary regions. This showed that gene duplication was limited to accessory genes in isolate Fo47, with all conserved genes (except for rDNA genes) as single copy genes. We also checked on matches between the genome assemblies of isolates Fo47 and Fo5176, and found that all non-unique matches were between supernumerary regions of the two isolates. Those comparisons confirmed

that all the gene duplications in isolate Fo5176 were within its supernumerary genome region.

Next we quantified the amount of large-scale structural variations between Foc race 1 isolates Fo5176 and IVC-1 with respect to the number of SNPs between them. We used minimap2 (Li 2018) with a stringent assembly-to-reference mapping option (the preset -x asm5 option) to compute unique alignments of lengths at least 200 kb between the Fo5176 and IVC-1 genome assemblies, where the minimum alignment length of 200 kb was selected because all duplications within the Fo5176 assembly were of lengths less than 200 kb. These unique alignments contained 88 deletion gaps of lengths from 1,146 to 8,662 bp with a total deletion gap length of 303,479 bp, and 85 insertion gaps of lengths from 1,323 to 12,268 bp with a total insertion gap length of 335,954 bp. Here, a deletion (an insertion) gap was composed only of a sequence region from the Fo5176 (IVC-1) assembly, and the length of the gap was the length of the sequence region. These large deletion and insertion gaps were likely to be some of the structural variations (SVs) between the two race 1 isolates with a total of 1,072 SNPs, resulting in a ratio of the SV total length to the SNP number at the level of  $639,433/1,072 = 596.5$ . Many of those sequence regions in the gaps were comprised of transposons present in multiple copies in the genome. So this approach allowed us to quantify SVs in repetitive regions of the genome.

To quantify SVs between races 1 and 2 in unique regions of the genome, we mapped short reads from race 1 isolate IVC-1 and race 2 isolate 58385 onto the genome assembly of race 1 isolate Fo5176 as a reference. An SV between race 1 isolate IVC-1 and race 2 isolate 58385 was defined as a reference region of length at least 1 kb such that the read depths for the two isolates at every position of the reference region consistently indicated the presence of coverage by one isolate (called P) and the absence of coverage by the other isolate (called A). Precisely, the following

567 conditions hold at every position  $x$  of the region:  $dep(P, x) \geq 0.7 * average\_dep(P)$   
 568 and  $dep(A, x) \leq 0.05 * average\_dep(A)$ , where  $dep(I, x)$  is the read depth of isolate  $I$   
 569 at position  $x$ , and  $average\_dep(I)$  is the genome-wide average read depth of isolate  $I$ .  
 570 We found 154 SVs of a total length of 315,107 bp with the presence of coverage only  
 571 by race 1 isolate IVC-1 and 108 SVs of a total length of 318,650 bp with the presence  
 572 of coverage only by race 2 isolate 58385. The number of SNPs between race 1 isolate  
 573 IVC-1 and race 2 isolate 58385 was 3,225. Thus, these numbers resulted in a ratio of  
 574 the SV total length to the SNP number at the level of  $633,757/3225 = 196.5$ .

## 575 **Evidence: Fusions between large core and supernumerary re-** 576 **gions**

577 We studied chromosome-level variation in several isolates by comparing their genome  
 578 assemblies. Isolates in the species complex are known to carry 11 core chromosomes  
 579 and one or more supernumerary chromosomes (Ma et al. 2010). The genome assembly  
 580 of Fo47 was composed of 12 chromosomes named 1 through 12 in decreasing order  
 581 of chromosome sizes, with 11 core and 1 supernumerary (chromosome 7). The two  
 582 smallest core chromosomes were less conserved than the other core chromosomes. By  
 583 comparing the genome assembly of Fo47 to those of the other isolates, we found in  
 584 some of the other isolates that one of the two smallest core chromosomes (homologous  
 585 to chromosomes 11 and 12 of Fo47) or a large region ( $\geq 900$  kb) of it was fused with  
 586 a supernumerary chromosome or a large region of it.

587 In the genome assembly of Fol race 3 isolate D11, contig 1 was a result of a fusion  
 588 between a 1,120-kb core region (as a 5' portion) and a 4,680-kb supernumerary region  
 589 (as a 3' portion), where the 1,112-kb core region was homologous to a 3' portion of  
 590 core chromosome 12 in isolate Fo47. Similarly, contig 6 of isolate D11 was composed  
 591 of a 1,206-kb supernumerary region (as a 5' portion) and a 929-kb core region (as a 3'



portion), with the core region being homologous to a 3' portion of core chromosome 11 in isolate Fo47. No such fusion events were detected in the genome assembly of Fol race 2 isolate 4287.

The genome assembly of Focb race 1 isolate 160527 was composed of 12 contigs named 1 through 12. Contig 2 of 5,885.8 kb was a result of a fusion between a core chromosome (homologous to core chromosome 11 in isolate Fo47) and a 3,211-kb supernumerary region (as a 3' portion). This fusion was confirmed by long reads. Contig 12 was a standalone supernumerary chromosome of 1,261 kb. All other contigs were core chromosomes based on comparison with the Fo47 chromosomes; their ends may contain short supernumerary regions ( $\leq 200$  kb).

The genome assembly of Focb TR4 isolate UK0001 was made up of 15 contigs, three of which were of lengths less than 120 kb. Among the remaining 12 contigs, contig 13 of 1.24 Mb was a standalone supernumerary chromosome, and contig 14 of 3.74 Mb was fused between a core chromosome (homologous to core chromosome 12 in isolate Fo47) and a supernumerary region of 1.30 Mb (as a 3' portion). All other contigs were core chromosomes based on comparison with the Fo47 chromosomes.

We examined a genome assembly of Foc race 1 isolate Fo5176, which was composed of 19 chromosomes named 1 through 19. To find chromosomal fusions and translocations in this genome assembly, we compared it with that of isolate Fo47. A first chromosomal difference between the two genome assemblies involved Fo5176 core chromosome 5 of 5.04 Mb and chromosome 6 of 5.01 Mb as well as Fo47 core chromosome 2 of 5.61 Mb and chromosome 4 of 4.73 Mb. To describe this difference, a minor portion was used to refer to a smaller 5' or 3' region of a chromosome and a major portion to the larger remaining part. Specifically, a 5' minor portion of Fo5176 chromosome 5 (chromosome 6) was syntenic to a 3' minor portion of Fo47 chromosome 2 (chromosome 4); the major portion of Fo5176 chromosomes 5 (chro-

618 mosome 6) was syntenic to the major portion of Fo47 chromosome 4 (chromosome  
619 2). Each of Fo47 core chromosomes 2 and 4 was syntenic to a single contig or scaffold  
620 in each genome assembly of isolates 160527 and UK0001. This observation suggests  
621 an event of reciprocal translocation in the Fo5176 lineage in which a 5' minor portion  
622 (totalling 1.79 Mb) of an ancestor core chromosome was exchanged with a 5' minor  
623 portion (totalling 1.45 Mb) of another ancestor core chromosome. This exchange was  
624 not present in isolates Fo47, 160527 or UK0001.

625 A second chromosomal difference associated Fo5176 chromosome 4 of 5.26 Mb  
626 and Fo5176 chromosome 13 of 2.80 Mb with Fo47 core chromosome 11 of 2.85 Mb. A  
627 core portion (at positions 0.54 to 1.80 Mb) of Fo5176 chromosome 4 was syntenic to a  
628 portion (at positions 0.57 to 1.66 Mb) of Fo47 core chromosome 11, and a core portion  
629 (at positions 1.68 to 2.69 Mb) of Fo5176 chromosome 11 was syntenic to a portion  
630 (at positions 1.66 to 2.50 Mb) of Fo47 core chromosome 11. The remaining portions  
631 of Fo5176 chromosomes 4 and 11 were mostly supernumerary. This difference was  
632 a result of chromosomal fusions between core and supernumerary chromosomes or  
633 chromosomal regions in the Fo5176 lineage (Fokkens et al. 2020).

634 A third difference connected Fo5176 chromosome 10 of 3.19 Mb to Fo47 core chro-  
635 mosome 10 of 2.89 Mb. A 5' minor portion (totalling 0.80 Mb) of Fo5176 chromosome  
636 10 was syntenic to a 5' minor portion of Fo47 core chromosome 10, and the major  
637 portion of Fo5176 chromosome 10 was supernumerary; the major portion of Fo47 core  
638 chromosome 10 was syntenic to the majority of Fo5176 chromosome 15 of 2.40 Mb.  
639 This finding indicated that Fo5176 chromosome 10 resulted from a fusion between a  
640 core chromosomal region and a supernumerary chromosome (Fokkens et al. 2020).

641 A last difference concerned Fo5176 chromosome 11 of 3.09 Mb and chromosome 12  
642 of 3.02 Mb as well as Fo47 core chromosome 5 of 4.52 Mb. A 5' minor portion (totalling  
643 1.23 Mb) of Fo5176 chromosome 11 was syntenic to a 3' minor portion (1.13 Mb)

644 of Fo47 core chromosome 5, and a major 5' portion of Fo47 core chromosome 5 was  
645 syntenic to Fo5176 chromosome 12. The major portion of Fo5176 chromosome 11 was  
646 supernumerary. Fo47 core chromosomes 5 was syntenic to a single contig or scaffold  
647 in each of the genome assemblies of isolates 160527 and UK0001. This observation  
648 indicated that Fo5176 chromosome 11 was composed of a core chromosomal region  
649 and a supernumerary one, and that Fo5176 chromosome 12 was a core one (Fokkens  
650 et al. 2020).

651 Put together, the genome assembly of isolate Fo5176 consisted of 11 core chromo-  
652 somes, 4 supernumerary chromosomes and 4 keeper chromosomes. A keeper chromo-  
653 some was composed of a core chromosomal region of at least 0.80 Mb and a larger  
654 supernumerary region. Unlike supernumerary chromosomes, which contained no es-  
655 sential genes, keeper chromosomes contained a core chromosomal region with con-  
656 served genes, and their losses may be deleterious. Keeper chromosomes play roles  
657 in the evolution of supernumerary chromosomes (see below). The four keeper chro-  
658 mosomes resulted from translocations involving core regions of at 0.8 Mb. Two of  
659 the core chromosomes underwent a reciprocal translocation involving core regions of  
660 at least 1.4 Mb. Regions around each translocation breakpoint were confirmed by  
661 long reads. Several Foc race 1 and 2 isolates belonged to the population contain-  
662 ing Foc isolate Fo5176 (see below), and these isolates all contained the same core  
663 chromosome translocation as isolate Fo5176. Because this translocation was a ge-  
664 netic barrier to meiotic recombination with other populations in the *F. oxysporum*  
665 complex, this population is an example of potential new species that emerged during  
666 asexual reproduction over the last hundreds of years.

## Evidence: Two or more copies of a supernumerary chromosome in isolates

Below we focus on two or more structurally different copies of a supernumerary chromosome in isolates. For two or more types of supernumerary chromosomes in isolates, see the above Fusion subsection on isolate Fo5176.

We estimated the copy numbers of Fol isolate D11 supernumerary chromosome 14 in 155 datasets of short reads, with multiple datasets generated, one per sequencing run, from each Fol isolate. To obtain a copy number estimate for supernumerary chromosome 14 in a dataset of short reads, we mapped the short reads onto D11 supernumerary chromosome 14 as a reference and separately onto the core chromosome carrying the *EF1- $\alpha$*  gene as a reference. Then we computed a copy number estimate by dividing the average read depth for supernumerary chromosome 14 by that for the core chromosome. In 143 of the 155 datasets, the copy number estimates for D11 supernumerary chromosome 14 were in the range 0.92–1.38, and in the remaining 12, the values were 1.76–1.78. These 12 datasets were generated in 12 sequencing runs from the same Fol isolate named DF0-62. A manual examination of the read depths of D11 supernumerary chromosome 14 in these 12 datasets revealed higher read depths over a 5' portion at 13.1 to 254.8 kb and over a 3' portion at 1,128.1 to 2,066.4 kb of D11 supernumerary chromosome 14 than over the middle portion between them. When only the 5' and 3' portions were used to estimate the copy numbers in these 12 datasets, the estimates were in the range 2.09–2.12; when only the middle portion was used to compute those, the estimates were in 1.16–1.19. These observations suggested that Fol isolate DF0-62 contained two structurally different copies of D11 supernumerary chromosome 14. In addition, we found 509 common SNPs in the 12 datasets of short reads from isolate DF0-62; 231 of the 509 SNPs had both reference and alternate allele frequencies above 30% in one of the 12 datasets, implying that

the two copies contained SNP differences as well. The presence of two structurally different copies of a supernumerary chromosome in an isolate suggested that these copies were acquired horizontally.

After mapping short reads from an isolate onto a reference chromosome or contig, we calculated the total uncovered size of the isolate for the reference by collecting all uncovered reference regions of sufficient lengths and totalling their sizes. For this analysis, the minimum length of each uncovered reference region was set to 5 kb. Of the 155 Fol isolates, 11 isolates had total uncovered sizes of 1,660 to 1,754 kb for supernumerary chromosome 14 of 2,139 kb, and the other 144 isolates had their values less than 900 kb. For D11 supernumerary contig 38 of 1,574 kb, these 11 isolates had total uncovered sizes of 962 to 1,046 kb, and the rest each had their values less than 600 kb. For D11 keeper contig 1 of 5,802 kb, 47 isolates had total uncovered sizes of 3,765 to 4,205 kb, and the remaining 108 isolates each had their values less than 2,500 kb. These 47 isolates include all of the 11 isolates mentioned above. For D11 keeper contig 6 of 2,135 kb, 83 isolates had total uncovered sizes of 913 to 1,184 kb, and the remaining 72 isolates each had their values less than 150 kb. These 83 isolates include all of the 47 isolates. Of the 155 isolates, 47 isolates were more distant to the Fol D11 isolate than the rest; for example, the SNP rates between these 47 isolate and the D11 core chromosome carrying the EF1-alpha gene were estimated to be at least 84 times more than those for the rest. This group of 47 distant isolates is exactly the same group of 47 isolates identified above to have the largest total uncovered sizes for keeper contig 1. Put together, these observations revealed that the 47 isolates contained little supernumerary portions of D11 keeper contigs 1 or 6, but of these 47 isolates, 36 isolates contained significant portions of D11 supernumerary chromosome 14 and supernumerary contig 38. In fact, the 36 isolates had total uncovered sizes of 280 to 411 kb for D11 supernumerary chromosome 14, and of 72 to

598 kb for D11 supernumerary contig 38. In other words, standalone supernumerary chromosomes, but not supernumerary portions of keeper chromosomes, were found in distant isolates, suggesting that supernumerary chromosomes moved from one isolate to another, but not keeper chromosomes. Note that keeper chromosomes contained core chromosomal regions and their horizontal transfer would result in the duplication of these core regions in isolates.

We also estimated the copy numbers of Focb TR4 UK0001 supernumerary chromosome 13 in 18 Focb TR4 isolates. We obtained a highest copy number value of 2.80 for isolate II-5, and values in a range of 1.96 to 2.12 for 4 isolates JV11, Leb1.2C, JV14 and FOC.TR4-1, and values in a range of 0.99 to 1.19 for the remaining 13 TR4 isolates. These numbers suggested three copies of UK0001 supernumerary chromosome 13 in isolate II-5, and two copies of this supernumerary chromosome in the 4 TR4 isolates. A manual examination of the read depths revealed that the copies in each of these isolates were structurally different.

Race 1 isolate N2 was distant to the TR4 isolates in core chromosome, and its population-specific subtelomere was different from that of the TR4 isolates. But it was closer to the TR4 isolates in Focb TR4 UK0001 supernumerary chromosome 13. A section of this chromosome from 278.7 to 518.4 kb was present in two copies in isolate N2, which contained 651 SNPs with both reference and alternate allele frequencies above 30%. In the remaining portion (totalling 1004.9 kb) of the chromosome, the number of SNPs between isolate N2 and TR4 isolate UK0001 was 230. Moreover, the SNP rate between isolates N2 and UK0001 in this portion of the chromosome 13 was 26 times lower than their genome-wide SNP rate, which is inconsistent with the expectation that a supernumerary chromosome is not more variable than the core chromosomes. An explanation to this inconsistency is that a recent horizontal transfer event involving a version of supernumerary chromosome 13 occurred to an

745 ancestor of isolates N2 or UK0001. This transfer event was preceded or followed by  
746 a change in the subtelomere of the supernumerary chromosome. For isolate N2, an  
747 average read depth of 454.6 for a region of 661 bp at 3.55 to 4.21 kb of a reference  
748 subtelomere for the TR4 isolates was obtained, indicating a partial similarity between  
749 their subtelomeres. Also, a retrotransposon of length 3.3 kb, which was located next  
750 to the 5' subtelomere of isolate TR4 UK0001 supernumerary chromosome 13, was  
751 present in 47 copies in isolate N2, in 12 copies in isolate TR4 II-5, but was present in  
752 0 or 1 copy in the other 17 TR4 isolates.

753 Some of the 18 TR4 isolates underwent changes in the subtelomeres of their chro-  
754 mosomes. Those changes were located in region 1 at 0.5-3.7 kb and region 2 at 5.2-8.6  
755 kb of a reference subtelomere, the 5' subtelomere of Focb TR4 isolate UK0001 su-  
756 pernumerary chromosome 13. For each of the 18 isolates, the ratio of the average  
757 read depth of region 1 to that of region 2 was calculated. For 9 of the 18 isolates,  
758 their ratios were between 0.88 and 1.47, and for 8 of them, their ratios were between  
759 0.11 and 0.49, and for the last one (isolate Pak1.1A), its ratio was 21.22. These large  
760 ratio differences suggested that some of these isolates underwent changes in many  
761 of their chromosome subtelomeres in the same region, during asexual reproduction.  
762 Note that all of these 18 isolates belonged to the same asexual population.

## 763 **Evidence: Origin of supernumerary chromosomes**

764 A supernumerary chromosome in one isolate may be a composition of portions of sev-  
765 eral supernumerary chromosomes in another isolate, so a global measure of similarity  
766 is less informative than a local measure of similarity. Below we first identified similar  
767 supernumerary regions between Focb TR4 isolate UK0001 and race 1 isolate 160527.  
768 Then we examined those regions between isolate 160527 and Foc race 1 isolate Fo5176,  
769 and within isolate Fo5176. Next we confirmed the presence of supernumerary regions

in 99 *F. oxysporum* isolates assembled by Achari et al. (2020), with 16 of these isolates collected from natural ecosystem soil, by estimating the total length of similar regions between the genome assembly of each isolate and the supernumerary genome of isolate Fo5176. Finally we compared Fol isolates 4287 and D11. Results from these comparisons indicated that supernumerary chromosomes in a new population evolved from ones in old populations.

Supernumerary contig 13 of 1,245 kb in isolate UK0001 contained 7 regions (totalling about 80% of the contig) that were syntenic to regions of the 3,211-kb supernumerary section of keeper contig 2 in isolate 160527. Let the 7 regions in isolate 160527 be denoted by a, b, c, d, e, f and g in forward orientation in the 5'-to-3' order. Then the order and orientation of the 7 syntenic regions in isolate UK0001 was A, F-, D, E-, C-, B- and G-, where a letter in upper case denotes a region in isolate UK0001 that was syntenic to the region denoted by the letter in lower case in isolate 160527, and region F- denotes the reverse complement of region F. The best gap-free matches in these 7 regions in the order in isolate UK0001 were 1/24,561, 2/52,381, 3/55,553, 0/33,736, 0/60,915, 1/20763 and 1/41,883, where a gap-free alignment of  $n$  paired nucleotides with  $d$  nucleotide differences was given in the form of  $d/n$ . For comparison, a best gap-free match between the core chromosomes of isolates 160527 and UK0001 was 61/10,946.

Isolate 160527 contained a keeper contig (contig 2) and a supernumerary contig (contig 12) as well as 10 core contigs. Four of the seven regions mentioned above in contig 2 of isolate 160527, regions c, d, e and f, overlapped with the first three of the following four regions of keeper contig 2 that were similar to supernumerary regions in isolate Fo5176. Region 1 of 133 kb at 3,468 to 3,601 Mb was 91% identical to a supernumerary region of Fo5176 keeper chromosome 11 at 2,227 to 2,344 Mb. Region 2 of 100 kb at 3,620 to 3,719 Mb was 93% identical to a supernumerary region



796 of Fo5176 chromosome 11 at 2,127 to 2,225 Mb. Region 3 of 290 kb at 3,986 to  
 797 4,275 Mb was 89% identical to a region of Fo5176 supernumerary chromosome 16  
 798 at 0.893 to 0.580 Mb (in reverse order denoting reverse orientation). Region 4 of  
 799 145 kb at 5,237 to 5,381 Mb was 86% identical to a supernumerary region of Fo5176  
 800 keeper chromosome 10 at 1,997 to 2,153 Mb. This example presented a case where  
 801 a supernumerary region in one isolate was composed of regions that were similar to  
 802 supernumerary regions of separate chromosomes in another isolate. In supernumerary  
 803 contig 12 of 1.261 Mb of isolate 160527, 4 of its regions (totalling 487 kb) were 87-90%  
 804 identical to supernumerary regions of Fo5176 keeper chromosome 4.

805 We examined homologous regions between supernumerary and keeper chromo-  
 806 somes in isolate 5176 to shed light on the evolution of these chromosomes. We found  
 807 three sets of highly similar regions between Fo5176 supernumerary chromosome 2 and  
 808 keeper chromosome 13. A first set contained a 5' region of 55 kb from keeper chromo-  
 809 some 13 that was 99% identical to two regions of Fo5176 supernumerary chromosome  
 810 2 at 750 to 801 kb and at 922 to 976 kb. A second set was a list of 6 close regions to-  
 811 talling 83 kb conserved at 98% identity between Fo5176 supernumerary chromosome  
 812 2 at 5.680 to 5.886 Mb and keeper chromosome 13 at 4.3 to 127.7 kb. The last one  
 813 was a set of 11 close regions totalling 60 kb conserved at 98% identity between chro-  
 814 mosome 2 at 128.8 to 246.0 kb and chromosome 13 at 885 to 1,027 kb. These matches  
 815 indicated recent segmental duplications between the two types of chromosomes in the  
 816 Fo5176 lineage. We also found 12 more sets of regions conserved at 87-97% identity  
 817 between supernumerary and keeper chromosomes in isolate 5176.

818 We checked on the presence of supernumerary regions in the genome assemblies of  
 819 99 isolates, with 16 of them collected from natural ecosystem soil. A dataset of super-  
 820 numerary sequence regions was prepared by taking all supernumerary chromosomes  
 821 and supernumerary portions of all keeper chromosomes in the genome assembly of

isolate 5176. The total lengths of similar regions of at least 5 kb between the supernumerary sequence regions of isolate 5176 and each of the genome assemblies of 99 *F. oxysporum* isolates ranged from 26.7 to 546.8 kb with a mean value of 186.5 kb. This observation supported the hypothesis that all *F. oxysporum* isolates contained supernumerary chromosomes or regions.

The supernumerary portion of contig 1 in Fol race 3 isolate D11 contained regions that were syntenic to supernumerary contigs 4, 18, 47, 65 in the genome assembly of Fol race 2 isolate 4287, but the rest of it, a 1,251-kb supernumerary region at 1,124-2,375 kb, had no long syntenic matches to any part of the 4287 genome assembly. This supernumerary region contained long syntenic matches to a genome assembly of *F. oxysporum* isolate ISS-F4, with a best gap-free match of 12.7 kb at a percent difference of 0.01%. As a comparison, a best gap-free core match of 13.5 kb between isolates D11 and ISS-F4 had a percent difference of 0.18%, which was 18 times that of the best supernumerary match. Note that isolates D11 and 4287 were highly similar in core chromosome; their best gap-free core match was of length 107 kb with no differences. This example revealed a case where a supernumerary chromosome in isolate D11 contained two kinds of regions that originated from supernumerary chromosomes in different lineages whose core chromosomes were more distant. Transfer, not loss, is best at explaining this observation. However, the comparison of the genome assemblies of isolates D11 and 4287 revealed many inter-chromosomal rearrangements between their supernumerary chromosomes; the two isolates were in the same population (see below).

## **Evidence: Examples of populations**

The asexual subpopulation containing Focb TR4 isolate UK0001 as an individual was defined as a group of *F. oxysporum* isolates whose chromosome ends were highly

847 similar to the subtelomere of 9.2 kb at the ends of chromosomes in Focb TR4 isolate  
 848 UK0001, and whose lineages to the most recent common ancestor (MRCA) were all  
 849 asexual. Besides isolate UK0001, we found 17 Focb TR4 isolates (see Methods).  
 850 Short reads from all 17 Focb TR4 isolates covered internal core AT rich regions of the  
 851 UK0001 genome assembly, suggesting that the lineages for the 17 isolates and UK0001  
 852 to the MRCA were all asexual. For 16 of the 17 isolates, their short reads covered  
 853 the UK001 subtelomere at average read depths from 680 to 4,371, with up to 2 SNPs.  
 854 This indicated the presence of multiple highly similar copies of this subtelomere in  
 855 these 16 TR4 isolates. The other TR4 isolate, Pak1.1A, had the lowest average read  
 856 depth of 440. A manual check on this coverage file found read depths of 0 to 2  
 857 for a 2.0-kb section of the subtelomere, and read depths of up to 90 for a 2.4-kb  
 858 section of the subtelomere, and an average depth of 675 for the rest (totalling 4.8 kb)  
 859 of the subtelomere. This showed that the subtelomeres at the ends of the Pak1.1A  
 860 chromosomes underwent significant changes. Thus, the 16 Focb TR4 isolates belonged  
 861 to the UK0001 asexual subpopulation, but isolate Pak1.1A did not. Note that all 17  
 862 isolates was extremely close to isolate UK0001 with a SNP rate less than 0.00001.  
 863 For example, Focb TR4 isolate Pak1.1A had an average read depth of 51.6 over the  
 864 whole reference genome with a SNP rate of  $6.7 \times 10^{-6}$  to the reference isolate (UK0001).  
 865 The low SNP rates indicated that the UK001 asexual subpopulation was young, and  
 866 that the changes in the subtelomere of isolate Pak1.1A happened recently, perhaps  
 867 within decades.

868 Another asexual subpopulation was defined based on Focb race 1 isolate 160527  
 869 with a subtelomere of 8.5 kb identified from its high-quality genome assembly. We  
 870 found 3 additional Focb race 1 isolates: N2, VCG0124 and VCG0125. Isolate N2  
 871 had a SNP rate of 0.00634 with isolate 160527. Because of a lack of read coverage  
 872 by short reads from isolate N2 over internal AT rich regions of the core chromosomes

873 for isolate 160527, either or both lineages for isolates 160527 and N2 had undergone  
874 meiotic recombination since their split. But their subtelomeres were still homologous;  
875 an initial 5.6-kb section of the isolate 160527 subtelomere and the rest (totalling 2.9  
876 kb) were covered at average depths of 918 and 135 by short reads from isolate N2.

877 Isolates VCG0124 and VCG0125 were close with a SNP rate of 0.00011, but were  
878 less close to isolate 160527, with SNP rates of 0.00148 and 0.00156, respectively.  
879 Still, the genome assemblies of isolates VCG0125 and 160527 contained 35 gap-free  
880 matches (with at least 99% percent identity) of lengths 3.1 to 66.8 kb totalling 403  
881 kb over regions with an AT content of at least 60%, suggesting that the lineages  
882 for these two isolates had remained asexual since their divergence. However, isolates  
883 VCG0124 and VCG0125 contained different types of subtelomeres. Short reads from  
884 isolate VCG0125 revealed that the subtelomere for isolate VCG0125 was globally  
885 similar to that of isolate 160527 with an average read depth of 1,448, where sharp  
886 drops in read depths occurred in only two locations (at locally lowest depths of 67  
887 and 349, respectively). But mapping short reads from isolate VCG0124 and from  
888 all TR4 isolates to the isolate 160527 subtelomere showed that the subtelomeres for  
889 these isolates including VCG0124 were only similar to a short section (totalling 1.5  
890 kb) of the isolate 160527 subtelomere; for example, only this section was covered at  
891 an average read depth of 1,093 by short reads from isolate VCG0124 and at 879 by  
892 those from Focb TR4 isolate JV11.

893 On the other hand, two adjacent sections (at sizes of 2.8 and 1.8 kb) of the  
894 isolate UK0001 subtelomere were covered at average read depths of 1,159 and 207  
895 by short reads from isolate VCG0124 and at 911 and 344 by those from Focb TR4  
896 isolate Pak1.1A. Note that this TR4 isolate was the only one not belonging to the  
897 UK0001 population (see above). Put together, the subtelomere for isolate VCG0124  
898 was more similar to that for isolate UK0001 than to that for isolate 160527. This

899 observation indicated that at least one of isolates VCG0124 and VCG0125 switched  
 900 to a different type of subtelomere since their divergence. Focb race 1 isolate VCG0124  
 901 had a SNP rate of 0.00814 with isolate UK0001. Short reads from isolate VCG0124  
 902 did not cover internal AT rich regions of the core UK0001 chromosomes, suggesting  
 903 that the lineage for isolate UK0001 underwent sexual reproduction after they split.  
 904 Thus, isolate VCG0124 did not belong to the UK0001 asexual subpopulation. Note  
 905 that short reads from isolates VCG0124 and VCG0125 covered, at read depths of at  
 906 least 10, 0.61.8% and 0.55.8% of supernumerary contig 2 of isolate 160527, and 45.9%  
 907 and 17.1% of supernumerary contig 13 of isolate UK0001. The two reference isolates,  
 908 UK0001 and 160527, were more distant, with a SNP rate of 0.00959. Although isolate  
 909 VCG0124 did not belong to the population that included the asexual subpopulation  
 910 with Focb TR4 isolate UK001, because of a lack of global subtelomere similarity,  
 911 the local subtelomere similarity indicated a common ancestor for the portions of the  
 912 subtelomeres for isolates VCG0124 and UK0001.

913 A third asexual subpopulation was defined based on Foc race 1 isolate Fo5176  
 914 with a subtelomere of 9.8 kb identified from its high-quality genome assembly. We  
 915 also selected Foc race 1 isolates Cong1-1 and IVC-1, and Foc race 2 isolates 54008  
 916 and 58385, based on the availability of genome assemblies or datasets of short reads.  
 917 The SNP rate between each pair of race 1 isolates was less than 0.00003, that between  
 918 the two race 2 isolates was 0.00009, and those between the race 1 and race 2 isolates  
 919 were between 0.00006 to 0.00010. The average read depth for the dataset of short  
 920 reads from isolate IVC-1 on the Fo5176 genome assembly as a reference was 49,  
 921 that from isolate 58385 was 64, and that from isolate Cong1-1 HS1 was 42. (The  
 922 same values for the read depths for the three datasets were obtained on the Cong1-1  
 923 genome assembly as a reference.) The percent coverage for the AT rich region of the  
 924 race 1 Fo5176 genome assembly was 80% for the dataset of short reads from race

1 isolate IVC-1, 94% for that from race 1 isolate Cong1-1 HS1, and 91% for that from race 2 isolate 58385. The values for those of the race 2 54008 genome assembly from the three datasets were 73%, 80% and 96%. These numbers reflected the mode of asexual reproduction in the lineages of these races 1 and 2 isolates up to their MRCA. The average read depths on the 9.8-kb Fo5176 subtelomere by the short reads from the three isolates were 3,250, 1,627 and 2,903, indicating multiple copies of the subtelomere the genomes of these isolates. Thus, these races 1 and 2 isolates belonged to the Fo5176 population.

A fourth population was centered on Fol isolate D11 with a high-quality genome assembly. After mapping short reads from isolate Fol 4287 onto the genome assembly of Fol isolate D11, we found that 57.6% of 974 kb of AT rich regions in the D11 genome assembly was covered by short reads from isolate Fol 4287, suggesting that Fol isolates 4287 and D11 were in the same asexual population. Both isolates carried nearly identical subtelomeres of 10.7 kp (with only a few short indel differences) at the ends of their chromosomes. Moreover, of 13 more Fol isolates, 9 were also in the D11 asexual population based on their percent coverage of the D11 AT rich regions in a range of 32.8% to 76.3%, but the other 4 were not in the population because of their low coverage in a range of 0.5% to 4%. The names of these 9 Fol isolates were 14844 (M1943), 5397, CA92.95, DF0-40, DF0-41, IPO1530/B1, LSU-3, LSU-7, WCS852/E241, and those of the other 4 were DF0-23, DF0-38, DF0-62, MN-14. Except isolate MN-14, the isolates also had highly similarly subtelomere sequences in high copy numbers. Thus, except isolate MN-14, the isolates belonged to the D11 population. A detailed description of the analysis of the 155 runs from these 13 isolates is given below.

The Fol D11 population was a well-studied population, with more than 100 datasets of short reads at NCBI. To check if isolates in the population carried mul-

951 tiple copies of the subtelomere at chromosome ends of the Fol D11 reference genome  
952 assembly, we mapped datasets of short reads from 155 runs of 13 Fol isolates (11-12  
953 runs per isolate) onto supernumerary chromosome 14 and contig 38 (as well as keeper  
954 contigs 1 and 6 each with both supernumerary and core regions) of Fol isolate D11 as  
955 a reference. Of the 155 runs, 144 runs each carried a significant portion (more than  
956 30%) of each of supernumerary chromosome 14 and contig 38, and carried multiple  
957 copies of the subtelomere based on read coverage. None of the remaining 11 runs  
958 carried a significant portion of either of supernumerary chromosome 14 and contig  
959 38, or carried any copies of the subtelomere. Moreover, none of these 11 runs carried  
960 a significant portion of the supernumerary region of either keeper contig 1 or contig  
961 6. Of the 144 runs carrying multiple copies of the subtelomere, 108 runs carried a  
962 significant portion of the supernumerary region of keeper contig 1, and 72 runs car-  
963 ried that of keeper contig 6. By defining the Fol population as a group of isolates  
964 with the Fol subtelomere at the ends of their chromosomes, we classified the 144 runs  
965 as belonging to the Fol population. Thus, all isolates in the population carried a  
966 supernumerary region of each of supernumerary chromosome 14 and contig 38, but  
967 some did not carry any of the supernumerary regions of keeper contigs 1 and 6.

## 968 Discussion

969 Complementary mechanisms of gene flow play important roles in speciation and  
970 species maintenance. Nonrecombining genome regions, horizontal DNA transfer and  
971 transposon proliferation are critical factors in speciation, while recombining genome  
972 regions, vertical DNA transfer and transposon control are critical factors in main-  
973 taining species. In populations of organisms that reproduce asexually and sexually,  
974 asexual reproduction, which is different from clonal reproduction, is an efficient mode

by which speciation occurs, and sexual reproduction is an effective mode by which species are maintained. In the *F. oxysporum* complex, a large number of populations arose for diverse plant and animal hosts during asexual reproduction, where extensive structural variation among isolates in the same asexual population overwhelmed nucleotide substitution variation among those isolates, and some populations contained core chromosome rearrangements. In homothallic fungi, the recombination between chromosomes from different nuclei in the same individual plays an important role in controlling transposons and maintaining chromosome core number and structure. In plants, polyploid speciation occurs during asexual reproduction in one or two generations and it may take many thousands of generations to new species to occur during sexual reproduction (Rieseberg and Willis 2007). Asexual reproduction is likely to play an important role in polyploid speciation (Herben et al. 2017). The rareness of ancient asexuals supports the claim that asexual reproduction is not an effective mode by which species are maintained for a long time. Bacteria is an exception because they contain single circular chromosomes, without separate chromosomes. This suggests that recombination is the most effective mechanism for maintaining chromosome number and structure, which plays a key role in maintaining species.

The *F. oxysporum* complex illustrates difficulty in defining species. On the other hand, the presence of gene flow between asexual populations in this complex offers support for the definition of species as groups of populations with gene flow within and between populations at population-specific rates, where complementary mechanisms of gene flow could operate in populations. Lack of extensive variation in core chromosome number and structure among isolates in the complex over 10,000 years suggests that it was important to maintain core chromosome number and structure among isolates in the complex through sexual reproduction. Thus, the maintenance of core chromosome number and structure as well as the presence of gene flow may



1001 be necessary properties of any eukaryotic species, whether they are sexual or not.

1002     The emergence of new species and the maintenance of existing species may be in  
1003 conflict with each other, because new species may compete with existing species for  
1004 resources. We discuss how this competition is addressed in general and in particular in  
1005 the *F. oxysporum* complex. According to the complementary theory, for populations  
1006 of organisms that reproduce sexually and asexually, new species emerge during asexual  
1007 reproduction. It is known that organisms reproduce asexually when resources are  
1008 abundant and switch to sex when resources are limited. Put together, new species  
1009 emerge when resources are abundant, making the competition with existing species  
1010 less likely to occur. In the *F. oxysporum* complex, populations contained molecular  
1011 signatures in the form of population-specific subtelomeric palindromes at the end  
1012 of each chromosome. Thus, populations can be distinguished from each other as  
1013 far as gene flow is concerned. Also, populations had specific plant hosts. Thus,  
1014 new populations are less likely to compete with existing populations for resources.  
1015 Moreover, the availability of new resources causes new populations to emerge.

1016     Plants and their fungal pathogens use different forms of gene duplication, an im-  
1017 portant evolutionary strategy, to compete against each other. Plants can multiply  
1018 their large genomes, while fungi cannot carry large genomes. The *F. oxysporum* com-  
1019 plex used a novel form of gene duplication. The complex contained a large number of  
1020 populations for different hosts, with each population carrying specific supernumerary  
1021 chromosomes, while the core chromosomes in all populations were similar and free  
1022 of duplicated genes. Novel supernumerary chromosomes emerged in existing popula-  
1023 tions and were duplicated in new populations by horizontal transfer during asexual  
1024 reproduction. This discussion reveals that horizontal transfer plays a more important  
1025 role than vertical transfer in the emergence of new populations in this complex. It  
1026 remains unclear whether the horizontal transfer of B chromosomes is important for

new plant species to emerge.

In population genetics, evolution is defined as a change in the frequency of gene variants. However, our analysis of genome data in the *F. oxysporum* complex showed that evolution also involves a change in the frequency of chromosome structural variants as well as horizontal chromosome transfer. These new types of variation and transfer are important in speciation. Also, asexual reproduction is not only different clonal reproduction, but also is an efficient mode by which speciation occurs. Deleterious mutations during asexual reproduction could take the form of transposon proliferation. The new theory calls for new mathematical models of and experimental approaches to population genetics to address evolutionary forces acting on non-Mendelian genetic variation, which enables new species to emerge.

## Additional Information and Declarations

### Competing Interests

The author is interested in exploring the potential of the genomic insights in industrial applications.

### Author Contributions

Xiaoqiu Huang conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper.

## Data and Code Availability

All sequencing and genomic data were downloaded from NCBI (see Methods for their accession numbers). Scripts used to process the data and sample output from them are available on the Open Science Framework at <https://osf.io/86y5r/>

## Funding

This work was supported by Iowa State University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Acknowledgments

The analysis of genome data was performed on high performance computing clusters managed by the Research IT team in College of Liberal Arts and Sciences at Iowa State University.

## References

- [1] Achari SR, Kaur J, Dinh Q, Mann R, Sawbridge T, Summerell BA, Edwards J. 2020. Phylogenetic relationship between Australian *Fusarium oxysporum* isolates and resolving the species complex using the multispecies coalescent model. *BMC Genomics* **21**: 248. doi: 10.1186/s12864-020-6640-y  
doi: 10.3390/cells8020156

- [2] Akagi Y, Akamatsu H, Otani H, Kodama M. 2009. Horizontal chromosome transfer, a mechanism for the evolution and differentiation of a plant-pathogenic fungus. *Eukaryotic Cell* **8**: 1732-1738. doi: 10.1128/EC.00135-09
- [3] Arie T, Kaneko I, Yoshida T, Noguchi M, Nomura Y, Yamaguchi I. 2000. Mating-type genes from asexual phytopathogenic ascomycetes *Fusarium oxysporum* and *Alternaria alternata*. *Molecular Plant-Microbe Interactions* **13**: 1330-1339 doi: 10.1094/MPMI.2000.13.12.1330
- [4] Cambareri E, Jensen B, Schabtach E, Selker E. 1989. Repeat-induced G-C to A-T mutations in *Neurospora*. *Science* **244**: 1571-1575. doi: 10.1126/science.2544994
- [5] Coluzzi M. 1982. Spatial distribution of chromosomal inversions and speciation in anopheline mosquitoes, pp. 143-153. In Barigozzi C (ed), Mechanisms of speciation. New York: Alan R. Liss, Inc.
- [6] Covert SF. 1998. Supernumerary chromosomes in filamentous fungi. *Current Genetics* **33**: 311-319. doi: 10.1007/s002940050342
- [7] Croll D, McDonald BA. 2012. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathogens* **8**: e1002608. doi: 10.1371/journal.ppat.1002608
- [8] van Dam P, Fokkens L, Schmidt SM, Linmans JHJ, Kistler HC, Ma L-J, Rep M. 2016. Effector profiles distinguish formae speciales of *Fusarium oxysporum*. *Environmental Microbiology* **18**: 4087-4102. doi: 10.1111/1462-2920.13445
- [9] van Dam P, Fokkens L, Ayukawa Y, van der GM, Ter Horst A, Brankovics B, Houterman PM, Arie T, Rep M. 2017. A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of multiple cucurbit species. *Scientific Reports* **7**: 9042. doi: 10.1038/s41598-017-07995-y

- [10] Darwin C. 1859. On The Origin of Species by Means of Natural Selection, or The Preservation of Favoured Races in the Struggle for Life. London: John Murray.
- [11] Dobzhansky TG. 1970. Genetics of the Evolutionary Process. New York: Columbia University Press.
- [12] Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development* **35**: 57-65. doi: 10.1016/j.gde.2015.09.001
- [13] Eschenbrenner CJ, Feurtey A, Stukenbrock EH. 2020. Population genomics of fungal plant pathogens and the analyses of rapidly evolving genome compartments. *Methods in Molecular Biology* **2090**: 337-355. doi: 10.1007/978-1-0716-0199-0\_14
- [14] Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **39**: W29-W37. doi: 10.1093/nar/gkr367
- [15] Fokkens L, Guo L, Dora S, Wang B, Ye K, Sánchez-Rodríguez C, Croll D. 2020. A chromosome-scale genome assembly for the *Fusarium oxysporum* strain Fo5176 to establish a model *Arabidopsis*-fungal pathosystem. *G3 Genes—Genomes—Genetics* **10**: 3549-3555. doi: 10.1534/g3.120.401375
- [16] Fokkens L, Shahi S, Connolly LR, Stam R, Schmidt SM, Smith KM, Freitag M, Rep M. 2018. The multi-speed genome of *Fusarium oxysporum* reveals association of histone modifications with sequence divergence and footprints of past horizontal chromosome transfer events. doi: 10.1101/465070
- [17] Franchini P, Kautt AF, Nater A, Antonini G, Castiglia R, Meyer A, Solano E. 2020. Reconstructing the evolutionary history of chromosomal races on islands:

- 1112 A genome-wide analysis of natural house mouse populations. *Molecular Biology*  
 1113 *and Evolution* **37**: 2825-2837. doi: 10.1093/molbev/msaa118  
 1114 doi: 10.1093/gbe/evy100
- 1115 [18] Gladyshev E. 2017. Repeat-induced point mutation (RIP) and other  
 1116 genome defense mechanisms in fungi. *Microbiology Spectrum* **5**:  
 1117 10.1128/microbiolspec.FUNK-0042-2017. doi: 10.1128/microbiolspec.FUNK-  
 1118 0042-2017
- 1119 [19] Hafezi Y, Sruba SR, Tarrash SR, Wolfner MF, Clark AG. 2020. Dissecting fer-  
 1120 tility functions of *Drosophila* Y chromosome genes with CRISPR. *Genetics* **214**:  
 1121 977-990. doi: 10.1534/genetics.120.302672
- 1122 [20] He C, Rusu AG, Poplawski AM, Irwin JA, Manners JM. 1998. Transfer of a su-  
 1123 pernumerary chromosome between vegetatively incompatible biotypes of the fun-  
 1124 gus *Colletotrichum gloeosporioides*. *Genetics* **150**: 1459-1466. doi: 10.1093/ge-  
 1125 netics/150.4.1459
- 1126 [21] Herben T, Suda J, Klimešová J. 2017. Polyploid species rely on vegetative re-  
 1127 production more than diploids: a re-examination of the old hypothesis. *Annals*  
 1128 *of Botany* bf 120: 341-349. doi:10.1093/aob/mcx009
- 1129 [22] Huang X. 2014. Horizontal transfer generates genetic variation in an asexual  
 1130 pathogen. *PeerJ* **2**:e650. doi: 10.7717/peerj.650
- 1131 [23] Huang X. 2019. Host-specific subtelomere: Genomic architecture of pathogen  
 1132 emergence in asexual filamentous fungi. *bioRxiv* 721753. doi: 10.1101/721753
- 1133 [24] Huang X, Adams MD, Zhou H, Kerlavage AR. 1997. A tool for analyzing and an-  
 1134 notating genomic sequences. *Genomics* **46**: 37-45. doi: 10.1006/geno.1997.4984

- 1135 [25] Huang X, Chao K-M. 2003. A generalized global alignment algorithm. *Bioinform-*  
1136 *atics* **19**: 228-233. doi: 10.1093/bioinformatics/19.2.228
- 1137 [26] Huang X, Das A, Sahu BB, Srivastava SK, Leandro LF, O'Donnell K, Bhat-  
1138 tacharyya MK. 2016. Identification of highly variable supernumerary chro-  
1139 mosome segments in an asexual pathogen. *PLoS ONE* **11**: e0158183. doi:  
1140 10.1371/journal.pone.0158183
- 1141 [27] Huang X, Miller W. 1991. A time-efficient, linear-space local similarity al-  
1142 gorithm. *Advances in Applied Mathematics* **12**: 337-357. doi: 10.1016/0196-  
1143 8858(91)90017-D
- 1144 [28] Huang X, Ye L, Chou H-H, Yang IH, Chao K-M. 2004. Efficient combination  
1145 of multiple word models for improved sequence comparison. *Bioinformatics* **20**:  
1146 2529-2533. doi: 10.1093/bioinformatics/bth279
- 1147 [29] Kistler HC, Rep M, Ma L-J. 2013. Structural dynamics of *Fusarium* genomes,  
1148 pp. 31-41. In Rown DW, Proctor RH (ed), *Fusarium: Genomics, Molecular and*  
1149 *Cellular Biology*. Norwich: Horizon Scientific Press.
- 1150 [30] Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. *Na-*  
1151 *ture Methods* **9**: 357-359. doi.org/10.1038/nmeth.1923
- 1152 [31] Langner T, Harant A, Gomez-Luciano LB, Shrestha RK, Malmgren A, Latorre  
1153 SM, Burbano HA, Win J, Kamoun S. 2021. Genomic rearrangements gener-  
1154 ate hypervariable mini-chromosomes in host-specific isolates of the blast fungus.  
1155 *PLoS Genetics* **17**: e1009386. doi: 10.1371/journal.pgen.1009386
- 1156 [32] Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinform-*  
1157 *atics* **34**: 3094-3100. doi: 10.1093/bioinformatics/bty191

- 1158 [33] Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro  
1159 A, Dufresne M, Freitag M, Grabherr M, Henrissat B, et al. 2010. Comparative  
1160 genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**:  
1161 367-373. doi: 10.1038/nature08850
- 1162 [34] Mayr E. 1942. Systematics and the Origin of Species. New York: Columbia  
1163 University Press.
- 1164 [35] Mendel G. 1901. Experiments in plant hybridisation. *Journal of the Royal Hor-*  
1165 *ticultural Society* **26**: 1-32.
- 1166 [36] Muller HJ. 1932. Some genetic aspects of sex. *American Naturalist* **66**: 118-138.  
1167 doi: 10.1086/280418
- 1168 [37] Nieuwenhuis BP, James TY. 2016. The frequency of sex in fungi. *Philosophical*  
1169 *Transactions of the Royal Society B: Biological Sciences* **371**: 20150540. doi:  
1170 10.1098/rstb.2015.0540
- 1171 [38] Orr HA, Otto SP. 1994. Does diploidy increase the rate of adaptation? *Genetics*  
1172 **136**: 1475-1480. doi: 10.1093/genetics/136.4.1475
- 1173 [39] Peichel CL. 2017. Chromosome evolution: Molecular mechanisms and evolution-  
1174 ary consequences. *Journal of Heredity* **108**: 1-2. doi: 10.1093/jhered/esw079
- 1175 [40] Peng Z, Oliveira-Garcia E, Lin G, Hu Y, Dalby M, Migeon P, Tang H, Farman  
1176 M, Cook D, White FF, Valent B, Liu S. 2019. Effector gene reshuffling involves  
1177 dispensable mini-chromosomes in the wheat blast fungus. *PLoS Genetics* **15**:  
1178 e1008272. doi: 10.1371/journal.pgen.1008272



- 1179 [41] Paoletti M, Clavé C. 2007. The fungus-specific HET domain mediates pro-  
 1180 grammed cell death in *Podospora anserina*. *Eukaryotic Cell* **6**: 2001-2008. doi:  
 1181 10.1128/EC.00129-07
- 1182 [42] Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for compar-  
 1183 ing genomic features. *Bioinformatics* **26**: 841-842. doi: 10.1093/bioinformat-  
 1184 ics/btq033
- 1185 [43] Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens:  
 1186 why bigger can be better. *Nature Reviews Microbiology* **10**: 417-430. doi:  
 1187 10.1038/nrmicro2790
- 1188 [44] Rep M, Kistler HC. 2010. The genomic organization of plant pathogenic-  
 1189 ity in *Fusarium* species. *Current Opinion in Plant Biology* **13**: 420-426. doi:  
 1190 10.1016/j.pbi.2010.04.004
- 1191 [45] Rieseberg LH, Willis JH. 2007. Plant speciation. *Science* **317**: 910-914.  
 1192 doi:10.1126/science.1137729
- 1193 [46] Stanke M, Waack S. 2003. Gene prediction with a hidden-markov model and a  
 1194 new intron submodel. *Bioinformatics* **19**: ii215-ii225. doi: 10.1093/bioinformat-  
 1195 ics/btg1080
- 1196 [47] Vlaardingerbroek I, Beerens B, Rose L, Fokkens L, Cornelissen BJ, Rep M.  
 1197 2016a. Exchange of core chromosomes and horizontal transfer of lineage-specific  
 1198 chromosomes in *Fusarium oxysporum*. *Environmental Microbiology* **18**: 3702-  
 1199 3713. doi: 10.1111/1462-2920.13281
- 1200 [48] Vlaardingerbroek I, Beerens B, Schmidt SM, Cornelissen BJ, Rep M. 2016b.  
 1201 Dispensable chromosomes in *Fusarium oxysporum* f. sp. *lycopersici*. *Molecular*  
 1202 *Plant Pathology* **17**: 1455-1466. doi: 10.1111/mpp.12440

1203 [49] Werren JH. 2011. Selfish genetic elements, genetic conflict, and evolutionary  
1204 innovation. *Proceedings of the National Academy of Sciences* **108**: 10863-10870.  
1205 <https://doi.org/10.1073/pnas.1102343108> doi: 10.1073/pnas.1102343108