1   **How transposons drive evolution of virulence in a fungal pathogen**

2   Luigi Faino[1][#], Michael F Seidl[1][#], Xiaoqian Shi-Kunne[1], Marc Pauper[1], Grardy CM van den

3   Berg[1], Alexander HJ Wittenberg[2], and Bart PHJ Thomma[1]*

4

5   [1]Laboratory of Phytopathology, Wageningen University, Droevendaalsesteeg 1, 6708 PB

6   Wageningen, The Netherlands

7   [2]Keygene N.V., Agro Business Park 90, 6708 PW Wageningen, The Netherlands

8

9   [#]These authors contributed equally to this work

10   *Corresponding author: Prof. dr. Bart PHJ Thomma

11   Chair, Laboratory of Phytopathology

12   Wageningen University

13   Droevendaalsesteeg 1

14   6708 PB Wageningen

15   The Netherlands

16   Email: bart.thomma@wur.nl

17

18

19   Running title: Genome evolution by transposable elements

20   Keywords: Genome evolution; Transposable element; Plant pathogen; *Verticillium dahlia*e;

21   segmental genome duplication

22 **Abstract**

23 Genomic plasticity enables adaptation to changing environments, which is especially relevant

24 for pathogens that engage in arms races with their hosts. In many pathogens, genes

25 mediating aggressiveness cluster in highly variable, transposon-rich, physically distinct

26 genomic compartments. However, understanding of the evolution of these compartments,

27 and the role of transposons therein, remains limited. We now show that transposons are the

28 major driving force for adaptive genome evolution in the fungal plant pathogen *Verticillium*

29 *dahliae*. Highly variable genomic regions evolved by frequent segmental duplications

30 mediated by erroneous homologous recombination, often utilizing transposons, leading to

31 genetic material that is free to diverge. Intriguingly, the duplicated regions are enriched in

32 active transposons that further contribute to local genome plasticity. Thus, we provide

33 evidence for genome shaping by transposons, both in an active and passive manner, which

34 impacts the evolution of pathogen aggressiveness.

Introduction

Genomic plasticity enables organisms to adapt to environmental changes and occupy novel niches. While such adaptation occurs in any organism, it is particularly relevant for pathogens that engage in co-evolutionary arms races with their hosts (Raffaele and Kamoun 2012; Seidl and Thomma 2014; Dong et al. 2015). In these interactions, hosts utilize their surveillance system to detect invaders and mount appropriate defenses, which involves detection of invasion patterns by immune receptors, while pathogens secrete so-called effector molecules that support host colonization to counteract these immune responses (Rovenich et al. 2014; Cook et al. 2015). The tight interaction, in which the microbe tries to establish a symbiosis that the host tries to prevent, exerts strong selection pressure on both partners and incites rapid genomic diversification (McDonald and Linde 2002).

Sexual reproduction is an important mechanism to establish genotypic diversity by combining genetic information of two parental lineages, followed by meiotic recombination leading to novel combinations of existing alleles (McDonald and Linde 2002; de Visser and Elena 2007; Heitman 2010; Speijer et al. 2015). However, not all eukaryotes regularly reproduce sexually, including many fungal phyla that are thought to reproduce mainly asexually (Heitman et al. 2007; Flot et al. 2013). However, also in such asexual organisms adaptive genome evolution occurs, which is mediated by various mechanisms ranging from single-nucleotide polymorphisms to large-scale structural variations that can affect chromosomal shape, organization and gene content (Seidl and Thomma 2014). *Verticillium dahliae* is a soil-borne fungal pathogen that infects susceptible hosts through their roots and colonizes the water-conducting xylem vessels, leading to vascular wilt disease (Fradin and Thomma 2006). Despite its presumed asexual nature, *V. dahliae* is a highly successful pathogen that causes disease on hundreds of plant hosts (Fradin and Thomma 2006; Klosterman et al. 2009; Inderbitzin and Subbarao 2014). Verticillium wilt diseases are difficult to control due to the long viability of fungal resting structures in the soil, the broad host range of the pathogen, and inability of fungicides to eliminate the pathogen from infected xylem tissues. Moreover, little genetic resistance is available (Fradin and Thomma 2006). Thus,

3

63 novel methods for disease control are urgently required, which requires increased

64 understanding of the biology of the fungus in its interaction with host plants (Fradin and

65 Thomma 2006).

66   Using comparative genomics, we recently identified genomic rearrangements in *V.*

67 *dahliae* that lead to extensive chromosomal length polymorphisms between strains (de

68 Jonge et al. 2013). Moreover, the genomes of *V. dahliae* strains were found to contain highly

69 dynamic, repeat-rich regions that are only present in subsets of the *V. dahliae* population,

70 and thus have been referred to as lineage-specific (LS) (Klosterman et al. 2011; de Jonge et

71 al. 2013). Intriguingly, LS regions are enriched for *in planta*-induced effector genes that

72 contribute to fungal virulence (de Jonge et al. 2013). Therefore, it was hypothesized that

73 elevated genomic plasticity in LS regions contributes to diversification of effector repertoires

74 of *V. dahliae* lineages and thus mediates their aggressiveness (de Jonge et al. 2013; Seidl

75 and Thomma 2014). In a similar fashion, many filamentous pathogens evolved so-called

76 'two-speed' genomes with gene-rich, repeat-poor genomic compartments that contain core

77 genes that mediate general physiology and evolve slowly, whereas plastic, gene-poor and

78 repeat-rich compartments are enriched in effector genes that mediate aggressiveness in

79 interactions with host plants and evolve relatively quickly (Raffaele and Kamoun 2012).

80   Plastic, fast-evolving genomic compartments in plant and animal pathogen genomes

81 concern particular regions that are either embedded within the core chromosomes or reside

82 on conditionally dispensable chromosomes (Thon et al. 2006; Fedorova et al. 2008; Haas et

83 al. 2009; Ma et al. 2010; Goodwin et al. 2011; Klosterman et al. 2011; Rouxel et al. 2011; de

84 Jonge et al. 2013). Irrespective of the type of organization of the 'two-speed' genome, the

85 fast-evolving compartment is generally enriched for transposable elements (TEs). These

86 elements have been hypothesized to actively promote genomic changes by causing DNA

87 breaks during excision, or by acting as a substrate for rearrangement (Seidl and Thomma

88 2014). Nevertheless, the exact role of TEs in the evolution of effector genes currently

89 remains unknown, as well as the molecular mechanisms and the evolutionary trajectories

90 that govern genome plasticity (Dong et al. 2015). Thus far, in-depth studies of the precise

4

91   roles of TEs were hampered by fragmented pathogen genome assemblies caused by

92   limitations in sequencing technologies, leading to poor assemblies of repeat-rich areas

93   particularly (Faino et al. 2015; Thomma et al. 2015). We recently obtained gapless *V. dahliae*

94   whole-genome assemblies using a combination of long-read sequencing and optical

95   mapping (Faino et al. 2015). Here, we exploit these novel assemblies for in-depth

96   investigation into the molecular mechanisms that are responsible for genomic variability, and

97   are instrumental for adaptive genome evolution in the plant pathogen *V. dahliae*.

98   **Results**

99   *Genomic rearrangements in* Verticillium dahliae *are associated with extensive sequence*

100   *similarity*

101   To investigate the molecular mechanisms that underlie genomic rearrangements in *V.*

102   *dahliae*, we performed whole-genome alignments between strains JR2 and VdLs17 (Faino et

103   al. 2015), revealing 24 synteny disruptions (Fig.1A; Supplemental Fig. 1). To further increase

104   the resolution of these synteny breakpoints, we subsequently aligned long (average ~9 kb)

105   sequencing reads derived from strain VdLs17 (Faino et al. 2015) to the completed, gapless

106   assembly of *V. dahliae* stain JR2, and reads of which the two sides aligned to two distinct

107   genomic locations were used to determine breakpoints at high-resolution (Supplemental Fig.

108   2). After manual refinement, this procedure yielded 19 large-scale chromosomal

109   rearrangements, 13 of which concern inter-chromosomal translocations and 6 concern intra-

110   chromosomal inversions (Fig. 1; Table 1). Of these 19 rearrangements, seven were confined

111   to regions smaller than 100 bp (Table 1). The remaining 12 genomic rearrangements, in

112   particular intra-chromosomal inversions, concern to larger genomic regions that could not be

113   further refined due to the presence of strain-specific and/or repeat-rich regions (Fig. 1B).

114   We subsequently assessed the occurrence of the 13 inter-chromosomal

115   rearrangements in nine other *V. dahliae* strains by querying paired-end reads derived from

116   these genomes for pairs of discordantly mapped reads (i.e. both reads fail to map at the

117   expected distance and/or location) when mapped onto the assembly of *V. dahliae* strain JR2

118   (Supplemental Fig. 3). This analysis revealed distinct rearrangement patterns among various

119   strains (Supplemental Fig. 3B). For example, while some synteny breakpoints identified in *V.*

120   *dahliae* strain JR2 are either specific to *V. dahliae* strain VdLs7 (Chr1: 2,843,014-2,843,020;

121   Supplemental Fig. 4A) or common to all other *V. dahliae* strains (Chr1: 8,013,865-8,013,868;

122   Supplemental Fig. 4B), some are observed in only a subset of *V. dahliae* strains (Chr3:

123   3,130,776-3,130,781; Supplemental Fig. 4C). Thus, genomic rearrangements commonly

124   occur in different *V. dahliae* strains, suggesting a common mechanism driving their

125   formation.

126   Genomic rearrangements are generally caused by double-strand DNA breakages

127   followed by unfaithful repair by DNA repair mechanisms that utilize homologous sequences

128   to repair such breaks (Hedges and Deininger 2007; Krejci et al. 2012; Seidl and Thomma

129   2014). Repetitive genomic elements, such as transposable elements (TEs), may give rise to

130   genomic rearrangements by providing an ectopic substrate that interferes with faithful repair

131   of the original break (Hedges and Deininger 2007; Seidl and Thomma 2014). Notably, out of

132   19 chromosomal rearrangements identified in *V. dahliae* strain JR2, 15 co-localize with a

133   repetitive element (Table 1; Supplemental Table 1). However, even though genomic

134   rearrangements in *V. dahliae* were previously associated with LTR-type retrotransposons (de

135   Jonge et al. 2013), we were neither able to confirm this association nor systematically

136   associate these structural variations to any other class of repetitive elements when using the

137   completely assembled genomes of *V. dahliae* strains JR2 and VdLs17 (Faino et al. 2015)

138   (Table 1). To understand the mechanisms that contributed to their formation, we studied the

139   13 inter-chromosomal rearrangements between *V. dahliae* strain JR2 and VdLs17 in more

140   detail. Of these 13 rearrangements, 12 could be reconstructed in detail (Fig. 1B; Table 1;

141   Supplemental Fig. 5). Eight of these 12 syntenic breakpoints occur over highly similar TEs in

142   both strains, albeit that they belong to different TE families (Fig. 1C; Table 1; Supplemental

143   Fig. 5). Closer inspection of the four remaining breakpoints revealed a TE at two of the

144   breakpoints in *V. dahliae* strain VdLs17, while this TE was absent at those breakpoints in the

145   JR2 strain. For the final two breakpoints no association to a TE was found in either strain, but

146   extended sequence similarity surrounding the rearrangement site was identified (Fig. 1C).

147   Therefore, we conclude that not necessarily TEs or their activity, but rather stretches of

148   sequence similarity are associated with ectopic chromosomal rearrangements in *V. dahliae*,

149   likely mediating unfaithful homology-based DNA repair. Since TEs are more abundant

150   compared to other (types of) sequences, these are more likely to become substrate for

151   double-strand repair pathways.

152   *Lineage-specific genomic regions in* Verticillium dahliae *evolved by segmental genomic*

153   *duplications*

154   Chromosomal rearrangements can lead to ectopic insertions, duplications or deletions of

155   genomic material, which play relevant roles in pathogen evolution during the arms race with

156   host plants (Seidl and Thomma 2014). Whole-genome alignments between *V. dahliae* strains

157   JR2 and VdLs17 revealed four large (>10 kb) repeat-rich genomic regions that do not display

158   synteny between the strains (Supplemental Fig. 6; Supplemental Table 2), suggesting that at

159   least part of the genomic material is absent in the respective other *V. dahliae* strain and

160   indicating extensive differential gene loss at these specific genomic locations. Moreover,

161   mapping of reads from nine additional strains (de Jonge et al. 2012) onto the JR2 and

162   VdLs17 assemblies revealed absence of read coverage mainly at their respective LS regions

163   (Supplemental Fig. 7). Notably, several inter-chromosomal rearrangements co-localize with

164   these LS regions (Fig. 1A), suggesting that genomic rearrangements contribute to the

165   formation of LS regions.

166   Chromosomal rearrangements not only lead to deletion of genetic material, e.g. in LS

167   regions (Fig. 1A), but also foster genomic duplications (Seidl and Thomma 2014). Even

168   though duplicated genes have previously been observed in LS regions of *V. dahliae*

169   (Klosterman et al. 2011), the extent of such duplications and their role in the evolution of LS

170   regions remains unknown. To determine the extent of segmental duplications in LS regions,

171   we employed two approaches. First, we performed whole-genome nucleotide alignments *of*

172   *V. dahliae* strain JR2 to itself to identify highly similar (>80% identity), large-scale duplication

173   events (Fig. 2A). Unanticipated, the vast majority of highly similar large-scale duplications

174   occurs within LS regions, while only very few of such duplications occur outside these

175   regions (Fig. 2A). Second, we performed homology detection between protein-coding genes

176   in *V. dahliae* strain JR2, establishing a set of ~1,000 paralogous sequences (Fig. 2B).

177   Notably, 40% of the 418 genes located in LS regions have a paralog (Fig. 2B), which is a

178   4.5x enrichment when compared to the core genome where only 7% of the ~11,000 genes

179   has a paralog (hypergeometic test; p = $1.31*10^{-69}$). Therefore, duplications of genomic

180   material are instrumental for the constitution of LS regions in *V. dahliae* (Fig. 2B).

181   The high level of similarity between sequences located at LS regions (Fig. 2A)

182   suggests that the duplications occurred rather recently. To firmly establish when these

183   duplications occurred during the evolution of *V. dahliae*, we used the rate of synonymous

184   substitutions per synonymous site (Ks) between paralogous gene pairs as a proxy for time

185   since these sequences diverged (Fig. 2C). While the Ks distribution of paralogous pairs

186   located in the core genome displays a single peak, indicating a single and distinct period in

187   which the majority of these duplications occurred, the distribution of paralogous pairs where

188   at least one gene is located in the LS regions displays two distinct peaks (Fig. 2C). Notably,

189   the older of the two peaks coincides with the peak observed for the core paralogs (Fig. 2C),

190   indicating that the expansion of core genes and a subset of genes in LS regions occurred in

191   the same period. The additional peak points towards duplications that occurred more recently

192   (Fig. 2C). To place these periods in relation to speciation events, we estimated Ks

193   distributions for orthologous gene pairs between *V. dahliae* strain JR2 and a number of

194   closely related fungi from the taxonomic class of Hypocreomycetidae (Fig. 2D). Within this

195   group of close relatives, the tomato wilt pathogen *Fusarium oxysporum* f.sp. *lycopersici* was

196   the first one to diverge from the last common ancestor, while the most recent split was the

197   divergence of *V. dahliae* strains JR2 and VdLs17 (Fig. 2D). The first duplication period

198   affected both the core genome and LS compartments, and occurred after the divergence of

199   *F. oxysporum* f.sp. *lycopersici*, but before the divergence of *Colletotrichum higginsianum*, the

200   causal agent of anthracnose leaf-spot disease of crucifers (Fig. 2D). Notably, all fungi that

201   diverged after speciation of *F. oxysporum* f.sp. *lycopersici* display a peak in the Ks

202   distribution for paralogous gene pairs at approximately the same period, although *A.*

203   *alcalophilium* shows a less pronounced peak (Fig. 2D). The second duplication event in *V.*

204   *dahliae* strain JR2 that specifically concerned genes located at LS regions occurred much

205   more recently, after the speciation of *Verticillium alfalfae* (Fig. 2D). As it was previously

206   shown that genes in the LS regions are particularly relevant for pathogen aggressiveness in

9

207   *V. dahliae* (de Jonge et al. 2013), this suggests that recent gene duplications are contributing

208   to the evolution of aggressiveness.

209   The recent duplications that affected LS regions have generated raw genetic material

210   that can be subject to subsequent rapid evolutionary diversification, leading to novel or

211   altered gene functionality, but can also be subject to differential loss of one of the duplicated

212   gene copies (Fig. 3A; Supplemental Fig. 8). In general, LS regions in *V. dahliae* strain JR2

213   display considerable gene loss, since for ~100 of the ~400 genes located in LS regions no

214   ortholog could be detected in *V. dahliae* strain VdLs17. For example, a repeat-rich, ~100 kb

215   LS region that contains ~20 genes that duplicated on chromosome 2 displays loss of multiple

216   genes (Fig. 3A; Supplemental Fig. 8), a process that is similarly observed in VdLs17. Thus,

217   differential gene loss significantly contributed to the diversification of LS regions. To

218   determine if LS regions display signs of increased gene diversification and selection pressure

219   acting on protein-coding genes, we used the rate of nonsynonymous substitutions per

220   nonsynonymous site (Ka) as well as the ratio of Ka to Ks values calculated between

221   orthologous gene pairs of *V. dahliae* strain JR2 and the closest related *Verticillium* species,

222   *V. alfalfae*, as a proxy. Genes located at LS regions display moderately increased Ka values,

223   as well as Ka/Ks values, when compared to genes residing in the core genome (median of

224   0.05 compared to 0.02, and 0.37 compared to 0.17, respectively). While these indicate

225   accelerated sequence divergence of genes located within LS regions, the moderate

226   differences also corroborate previous results (de Jonge et al. 2013; Seidl et al. 2015),

227   suggesting that sequence divergence only plays a minor role in *V. dahliae* genome evolution.

228

229   *The* Ave1 *effector gene is located in a highly dynamic genomic region*

230   As shown above, highly dynamic LS regions are characterized by frequent gene duplications

231   (Fig. 2), and by differential gene loss (Fig. 3; Supplemental Fig. 8). Moreover, effector genes

232   located in LS regions play decisive roles in pathogen-host interactions (de Jonge et al. 2013).

233   For example, Ave1 is an important LS effector that determines *V. dahliae* aggressiveness on

234   various host plants (de Jonge et al. 2012). As expected in a co-evolutionary arms race

235 (Thomma et al. 2011; Cook et al. 2015), host recognition of this effector evolved as tomato

236 plants that carry the immune receptor Ve1 recognize this effector to establish immunity

237 (Fradin et al. 2009; de Jonge et al. 2012). Consequently, race 2 strains of *V. dahliae* lost

238 *Ave1*, thus evading recognition by Ve1, leading to the capacity to infect *Ve1* tomato plants.

239 In *V. dahliae* strain JR2, *Ave1* is embedded in a gene-sparse and repeat-rich LS

240 region on chromosome 5 (Fig. 3B-E; ~550,000-1,050,000). Notably, the average number of

241 single nucleotide polymorphisms (SNPs) inferred from three race 1 and eight race 2 strains

242 of *V. dahliae* is significantly reduced in the area surrounding the *Ave1* locus (between

243 680,000 and 720,000) when compared with the surrounding genomic regions (Fig. 3B), but

244 also compared to genome-wide SNP levels in LS and core regions (Supplemental Fig. 9). As

245 Ave1 is a strong contributor to virulence on plants lacking *Ve1* (de Jonge et al. 2012),

246 selection pressure may drive the high level of conservation and lack of SNPs (Fig. 3B) in this

247 region in race 1 strains.

248 Phylogenetic analyses revealed that *Ave1* was likely acquired by *V. dahliae* through

249 horizontal gene transfer from plants (de Jonge et al. 2012). Intriguingly, however, *V. dahliae*

250 race 1 strains do not occur as a monophyletic clade in the *V. dahliae* population

251 (Supplemental Fig. 3A) (de Jonge et al. 2013), and thus it remains unclear how the *Ave1*

252 locus distributed throughout the *V. dahliae* population. To assess if *Ave1* was gained or lost

253 multiple times, we studied the genomic region surrounding the *Ave1* locus in the LS region.

254 As expected, alignments of genome assemblies of multiple *V. dahliae* strains revealed the

255 absence of *Ave1* in all race 2 strains (Supplemental Fig 10A). By mapping paired-end reads

256 derived from genomic sequencing of various *V. dahliae* strains onto the genome assembly of

257 *V. dahliae* strain JR2 we observed clear differences in coverage levels between *V. dahliae*

258 race 1 and race 2 strains that carry or lack the *Ave1* gene, respectively (Fig. 3C;

259 Supplemental Fig. 10B). While *V. dahliae* race 1 strains, including JR2, displayed an even

260 level of read coverage over the *Ave1* locus, indicating that this region is similar in all race 1

261 stains, no read coverage of the *Ave1* gene was found for race 2 strains (Fig. 3C-D).

262 Intriguingly, read coverage surrounding the *Ave1* gene revealed that race 2 strains can be

11

263   divided into three groups depending on the exact location of the read coverage drop (Fig.

264   3C). Whereas one group of isolates does not display any read coverage up to 720 kb, two

265   groups display distinct regions in which the read coverage around the *Ave1* locus drops (Fig.

266   3C-E: red lines around 668 kb and green lines around 672 kb). Our finding strongly suggests

267   that the *Ave1* locus was lost multiple times from the *V. dahliae* population, pointing to strong

268   selection pressure posed by the Ve1 immune receptor on *V. dahliae* to lose *Ave1*. In

269   conclusion, the *Ave1* locus is situated in a highly dynamic and repeat-rich region (Fig. 3C).

270   Therefore, the most parsimonious evolutionary scenario is that the *Ave1* locus was

271   horizontally acquired either directly or indirectly from plants once, followed by multiple losses

272   in independent lineages that encountered host plants that carried Ve1 or functional homologs

273   of this immune receptor (Thomma et al. 2011; de Jonge et al. 2012; Zhang et al. 2014). It is

274   tempting to speculate that the transposable elements that flank the *Ave1* locus (Fig. 3C)

275   contributed to the evolution of the *Ave1* locus by facilitating swift loss of the effector gene

276   upon recognition by the Ve1 immune receptor.

277

278   *Lineage-specific genomic regions in* Verticillium dahliae *contain active transposable*

279   *elements*

280   While the activity of TEs is not associated with the formation of extensive genome

281   rearrangements, LS regions in *V. dahliae* are highly enriched for TEs, and their presence and

282   potential activity may contribute to accelerated evolution of these genomic regions. In *V.*

283   *dahliae*, the most abundant class of TEs are retrotransposons that transpose within the

284   genome using an RNA intermediate (Wicker et al. 2007; Faino et al. 2015) (Supplemental

285   Table 1). We assessed TE dynamics by querying the transcriptional activity of TEs using *in*

286   *vitro* RNA-Seq data derived from *V. dahliae* strain JR2(de Jonge et al. 2012). Notably, the

287   majority of TEs in *V. dahliae* are not transcribed and thus likely not active (Supplemental

288   Table 1), while transcribed and therefore likely active TEs are found in LS regions

289   (Supplemental Fig. 11).

290   To further assess if and how TEs influence the evolution of LS regions, we explored

291   TE dynamics in the genome of *V. dahliae* strain JR2. Each copy of a TE in the genome is

292   derived from an active ancestor that, once transposed and integrated into the genome,

293   accumulates mutations that, over evolutionary time, will render the TE inactive. The relative

294   age of individual TEs can thus be estimated based on sequence divergence from a

295   consensus sequence that can be derived from present-day copies of any given TE. Using the

296   Jukes-Cantor distance (Jukes and Cantor 1969), which corrects the divergence between TEs

297   and their consensus sequence for multiple substitutions, we estimated the divergence times

298   for TEs in the *V. dahliae* strain JR2 genome (Fig. 4). This analysis showed that TEs mainly

299   transposed and expanded in two distinct periods (Fig. 4A). Notably, a considerable amount

300   of 'younger' TEs, i.e. with small Jukes-Cantor distance to their consensus sequence,

301   localizes in LS regions while the majority of 'older' TEs reside in the core genome (Fig. 4A).

302   Next, we attempted to determine the relative period in which the majority of TEs in the

303   genome of *V. dahliae* strain JR2 transposed. To this end, we derived Jukes-Cantor

304   distributions for orthologous genes of *V. dahliae* strain JR2 and individual closely related

305   fungi as a proxy of divergence between these species (Fig. 4C). These distributions display

306   the same pattern of species divergence when derived from phylogenetic analyses

307   (Supplemental Fig. 12) as well as from the Ks distributions between orthologous gene pairs

308   (Fig. 2D). By comparing the Jukes-Cantor distributions derived from TEs and from

309   orthologous genes, we revealed that 'older' TEs transposed around the separation of *V.*

310   *dahliae* and *V. alfalfa*, while the 'younger' TEs transposed after *V. dahliae* speciation (Fig.

311   4C). Notably, 'younger' TEs tend to be transcriptionally active, while 'older', more diverged

312   TEs tend to be transcriptionally silent (Fig. 4B). Thus, the expansion of younger TEs is recent

313   and mainly concerns the active TEs localized at LS regions (Fig. 4), strongly suggesting that

314   TE-transpositions contribute to the genetic plasticity of LS regions.

Discussion

Many plant pathogens contain a so-called 'two-speed' genome where effector genes reside in genomic compartments that are considerably more plastic than the core of the genome, facilitating the swift evolution of effector catalogs that are required to be competitive in the host-pathogen arms race (Raffaele and Kamoun 2012; Dong et al. 2015). Generally, effector compartments are enriched in transposable elements (TEs), and it has been speculated that they promote genomic flexibility and drive accelerated evolution of these genomic compartments (Gijzen 2009; Haas et al. 2009; Ma et al. 2010; Raffaele et al. 2010; Rouxel et al. 2011; Raffaele and Kamoun 2012; de Jonge et al. 2013; Wicker et al. 2013; Grandaubert et al. 2014; Seidl and Thomma 2014; van Hooff et al. 2014; Dong et al. 2015; Faino et al. 2015; Seidl et al. 2015). Nevertheless, the exact role of TEs in the evolution of effector genes, as well as the molecular mechanisms and the evolutionary trajectories that govern the formation of the 'two-speed' genome, and the corresponding local genome plasticity, remained unknown (Raffaele and Kamoun 2012; Dong et al. 2015). Here, we studied the evolution of the 'two-speed' genome of the vascular wilt pathogen *V. dahliae*, which was significantly facilitated by the recent establishment of gapless whole-genome assemblies of two highly similar *V. dahliae* strains that, despite their high (>99%) degree of sequence identity, display severe genomic rearrangements (Faino et al. 2015; Thomma et al. 2015). We identified ~2 Mb repeat-rich, lineage-specific (LS) regions between the two *V. dahliae* strains (Fig. 1; Supplemental Fig. 6-7) that are significantly enriched in TEs and contain all thus far functionally analyzed effector genes, including *Ave1* (Fig. 3C) (de Jonge et al. 2012; de Jonge et al. 2013). Moreover, we also determined a significant number of synteny breakpoints that are associated with genomic rearrangements to high resolution, and we were able to reconstitute most of the rearrangements in detail (Fig. 1; Table 1; Supplemental Fig. 5). Previously, we showed a strong association between LS compartments and the occurrence of genomic rearrangements (de Jonge et al. 2013). Even though this correlation was overestimated due to significant errors in the publically available genome assembly of *V. dahliae* strain VdLs17 (Klosterman et al. 2011) for which we later revealed a considerable

14

343 number of erroneous chromosomal inversions (Faino et al. 2015) (Table 1), we nevertheless

344 observed that three out of four large-scale LS regions are associated with chromosomal

345 rearrangements (Fig. 1), corroborating that chromosomal rearrangements significantly

346 contribute to the evolution of LS regions in *V. dahliae*. We were not able to associate every

347 LS region with a genomic rearrangement (Fig. 1), like we were also not able to exactly

348 reconstitute each genomic rearrangement (Fig. 1; Supplemental Fig. 5). However, it can

349 easily be anticipated that complex rearrangements occurred in which genetic material in

350 close proximity was lost. Furthermore, likely these rearrangements (continuously) occurred

351 over longer evolutionary timescales (Supplemental Fig. 3; Supplemental Fig. 4), and

352 subsequent rearrangement events may have erased 'scars' of previous rearrangements.

353 Thus, even though not every genomic rearrangement leads to a new LS region, large-scale

354 genomic alterations are the driving force for their formation in *V. dahliae*.

355 Genomic rearrangements can lead to a multitude of structural variations, including

356 translocations, duplications and gene loss (Seidl and Thomma 2014). Here, we show that the

357 LS compartments of *V. dahliae* strain JR2 evolved by frequent segmental duplications that

358 yielded genetic material that subsequently obtained the freedom to diverge (Fig. 2B).

359 Intriguingly, these duplications occurred recently, namely after divergence of *V. dahliae* and

360 *V. alfalfae* (Fig. 2). However, this is not the only wave of duplications that shaped the *V.*

361 *dahliae* genome, as an earlier duplication period was found that concerns duplications found

362 in the core genome (Fig. 2). Notably, this duplication event was not only observed in *V.*

363 *dahliae*, but also in related species including *V. alfalfae*, *V. albo-atrum*, *A. alcalophilium* and

364 *C. higginsianum*, yet not in *F. oxysporum*, suggesting that this duplication occurred after the

365 speciation of *F. oxysporum* from the last common ancestor of the more recently evolved

366 fungi (Fig. 2). Given the high abundance of duplicated genes at a confined point in evolution,

367 it is tempting to speculate that a single genome-wide event, for example a whole-genome

368 duplication or hybridization, gave rise to these duplicates. Potentially, this particular

369 duplication event determined that, despite their similar life style and physiology, *V. dahliae*

370 evolved LS regions that are embedded within its core genome, while *F. oxysporum* evolved

15

371   LS genetic material as conditionally dispensable chromosomes that can be horizontally

372   transferred (Ma et al. 2010).

373   Based on our data, which indicate that TEs can be implicated in many yet not all

374   genomic rearrangements, it is conceivable that homologous recombination using TE

375   sequences as a substrate, rather than TE activity, is responsible for the establishment of

376   gross rearrangements. In this manner, TEs have passively contributed to the genome

377   evolution of *V. dahliae*, and to the extensive genomic plasticity that can be observed between

378   *V. dahliae* strains (Fig. 1). In addition to the passive roles of TE in genome evolution, we also

379   obtained evidence for the involvement of TE activity in the evolution of *V. dahliae*.

380   Intriguingly, while the majority of TEs in *V. dahliae* is transcriptionally silent and thus inactive,

381   transcriptionally active and thus actively transposing TEs were observed in LS regions (Fig.

382   4, Supplemental Fig. 9). Similarly, transcription and specific induction of TE activity has been

383   observed previously in *V. dahliae* strain VdLs17 (Amyotte et al. 2012). These independent

384   observations are further corroborated by our dating analyses suggesting that TEs that

385   localize in LS regions are significantly 'younger' when compared with those residing in the

386   core genome (Fig. 4). Active transposition can contribute to genome plasticity by causing

387   gene deletions at their target sites. Moreover, TEs in proximity of genes can also profoundly

388   influence their expression, as shown for the barley smut fungus *Ustilago hordei* where a TE

389   insertion in the promoter of the *UhAvr1* effector changed its expression, leading to loss of

390   host recognition (Ali et al. 2014). Interestingly, frequent gene loss is observed within the

391   segmental duplications that constitute the LS regions (Fig. 2; Supplemental Fig. 8). Similarly,

392   the LS effector *Ave1* is located in an LS region (Fig. 3) with several transposable elements

393   (TE) in its direct vicinity (de Jonge et al. 2012) (Fig. 3C), and clear evidence for repeated loss

394   in the *V. dahliae* population was obtained in our study. Although based on our data these

395   transposons cannot directly be implicated in the acquisition of *Ave1* from plants, they likely

396   facilitated the multiple losses of *Ave1* that are observed in independent lineages incited by

397   selection pressure posed by the *Ve1* resistance gene of tomato, or functional homologs that

398   may occur in other plants (Fradin et al. 2009; Thomma et al. 2011; de Jonge et al. 2012;

16

399   Zhang et al. 2013). However, although *Ave1* resides in a LS region, and LS regions are

400   characterized by the presence of active TEs, it remains unclear whether the frequent loss of

401   *Ave1* is actually mediated by TE activity. It has similarly been hypothesized that ectopic

402   genomic rearrangements caused by homology-based DNA repair pathways, possibly

403   passively mediated by TEs, drive the frequent loss of the *Avr-Pita* effector gene in isolates of

404   the rice blast fungus *Magnaporthe oryzae* that encountered the *Pita* resistance gene of rice

405   (Orbach et al. 2000; Chuma et al. 2011). Similar processes likely contribute to the frequent

406   recovery of *Avr-Pita* in *M. oryzae* strains relieved from selection pressure of Pita, leading to

407   translocation and duplications of this effector gene in dynamic genomic regions (Chuma et al.

408   2011).

409   Genome-wide studies in several fungi aiming to study chromatin – the complex of

410   DNA and proteins - revealed that TE-rich regions are generally associated to highly

411   condensed chromatin, the so-called heterochromatin, that restricts transcription and TE

412   activity (Lewis et al. 2009; Connolly et al. 2013; Galazka and Freitag 2014). Therefore, TEs

413   can influence the expression of neighboring genes such as effectors, as they can direct the

414   formation of heterochromatic regions (Lewis et al. 2009). In the saprophytic fungus

415   *Neurospora crassa*, heterochromatin formation at TEs is directed by remnants of Repeat

416   Induced Point mutations (RIP), a pre-meiotic process that actively induces point mutations in

417   TEs (Selker 1990; Lewis et al. 2009; Lewis et al. 2010). In the pathogenic fungus

418   *Leptosphaeria maculans*, effectors are located in TE-rich regions that were subjected to

419   extensive RIP mutations (Rouxel et al. 2011). Notably, effectors within these regions display

420   signatures of nucleotide mutations caused by the RIP process, fostering rapid effector

421   diversification (Fudal et al. 2009; Daverdin et al. 2012). LS regions in *V. dahliae* are TE-rich

422   and effector genes are directly flanked by TEs (de Jonge et al. 2012; de Jonge et al. 2013;

423   Seidl et al. 2015)(Fig. 3). Only few transposable elements in *V. dahliae* display typical RIP

424   mutations, which might indicate that *V. dahliae* possessed an active RIP mechanism

425   (Klosterman et al. 2011; Amyotte et al. 2012), likely connected to sexual cycle, in its

426   evolutionary past. However, all currently known *V. dahliae* strains, but also other *Verticillium*

17

427 species have been described as asexual and therefore RIP is no longer likely to occur. *V.*
428 *dahliae* effectors in LS regions do not display evidence for RIP mutations that could
429 contribute to effector diversification or to formation of heterochromatic regions. Despite the
430 fact that LS regions do not show overrepresentation of secreted genes, we previously
431 observed that effector genes residing in the LS regions are considerably overrepresented in
432 the *V. dahliae* transcriptome upon plant infection (de Jonge et al. 2013). In the present study,
433 we highlight that TEs that reside in LS regions are relatively young and several are
434 transcriptionally active. Therefore, we hypothesize that these regions, in contrast to the core
435 genomic compartments, are either not yet targeted by heterochromatin formation or carry
436 different chromatin marks, or that LS regions represent facultative heterochromatin that can
437 dynamically change its conformation, thereby influencing DNA accessibility and thus effector
438 gene expression. For example, in the soybean pathogen *Phytophthora sojae* expression of
439 the effector *Avr3a* is repressed by chromatin-based mechanism (Qutob et al. 2013; Gijzen et
440 al. 2014). Suppression of effector expression leads to avoidance of recognition by the host
441 surveillance system, yet allows for swift effector recovery and activation on susceptible hosts
442 (Gijzen 2009; Qutob et al. 2013). Chromatin-based regulation of effector genes and genes
443 encoding other virulence factors has been observed in several pathogenic fungi (Connolly et
444 al. 2013; Qutob et al. 2013; Chujo and Scott 2014; Soyer et al. 2014; Soyer et al. 2015). This
445 suggests that facultative heterochromatin located in dynamic compartments of the two-speed
446 genome, e.g. LS regions in *V. dahliae*, is pivotal to regulate effector and virulence gene
447 expression.
448 Chromosome condensation of heterochromatic regions supposedly restricts genomic
449 rearrangements (Galazka and Freitag 2014), yet these regions in the plant pathogens
450 *Fusarium graminearum* are enriched for genomic rearrangements (Connolly et al. 2013). In
451 *V. dahliae*, LS regions are formed by recent duplications and harbor active TEs, suggesting
452 that these regions are susceptible to structural variation such as genomic rearrangements. A
453 recent study in yeasts and mammals suggests that genomic rearrangements occur between
454 open chromatin in regions that are in close proximity within the nucleus (Berthelot et al.

18

455   2015). Moreover, this study concludes that genomic rearrangements between TEs alone are

456   not sufficient to fully explain the rearrangements that were observed (Berthelot et al. 2015).

457   Expression of clustered virulence genes in the human malaria pathogen *Plasmodium*

458   *falciparum* is tightly regulated by chromatin and, even though these virulence clusters are

459   located on different chromosomes, they co-localize in close proximity within the nucleus (Ay

460   et al. 2014). Even though these co-localized clusters have been associated with condensed,

461   and thus repressive chromatin, extensive ectopic chromosomal rearrangements between

462   clusters have been observed (Freitas-Junior et al. 2000; Lopez-Rubio et al. 2009; Jiang et al.

463   2013; Ay et al. 2014), suggesting that the chromatin structure plays roles in their evolution. In

464   fungi, however, co-localization of effector genes in the nucleus has not yet been reported.

465   Based on initial work in other eukaryotes, we conceive that complex chromatin structures in

466   pathogenic fungi will not only influence coordinated effector expression in LS regions, but

467   also genomic rearrangements, thereby further linking genome and chromatin structure to

468   genome evolution. Thus, further insight into chromatin biology of plant pathogens will be

469   instrumental to significantly accelerate our knowledge on the evolution of the 'two-speed'

470   genome, and on fungal aggressiveness.

471 **Methods**

472 *Genome annotation and dynamics*

473 Gene predictions for the whole-genome assembly of *V. dahliae* strain JR2(Faino et al. 2015)

474 was performed using the Maker2 software (Holt and Yandell 2011). To this end, RNA-seq

475 reads derived from different *in vitro* and *in planta* conditions(de Jonge et al. 2012) were

476 mapped to the genome using Tophat2 (default settings) (Trapnell et al. 2009). Additionally,

477 gene sequences derived from previous genome annotations as well as protein sequences

478 from 35 fungal proteomes were used as additional evidence (Klosterman et al. 2011; de

479 Jonge et al. 2013; Seidl et al. 2015).

480 Homology between 13 fungal species was assessed using OrthoMCL (default

481 settings)(Li et al. 2003). Sequence similarity between proteins was established by all-vs.-all

482 analyses using BLASTp (E-value cutoff 1e-5, soft filtering)(Altschul et al. 1990). Ks values

483 between gene pairs, as defined by OrthoMCL families, were calculated using the Nei-

484 Gojobori algorithm included in the KaKs_Calculator 2.0 package(Wang et al. 2010). The

485 coding sequences of gene pairs were aligned using protein alignment as a guide. The

486 phylogenetic tree was generated by RAxML(Stamatakis 2006) using concatenated protein

487 sequences of 3,492 single-copy orthologs that are conserved among eight fungal species,

488 where only a single representative species was chosen for the fungal genera *Colletotrichum*

489 and *Fusarium* (Supplemental Fig. 12).

490 Repetitive elements were identified as described in Faino et al. (2015). Briefly,

491 repetitive elements were identified using RepeatScout, LTR_Finder and LTRharvest, and the

492 repetitive elements identified by the different software were combined (non-redundant).

493 Repetitive elements were further classified as described by Wicker et al. (2007). Open

494 reading frames within the transposable elements were identified by BlastN and BlastX

495 (Camacho et al. 2009) searches against NCBI NR databases as well as by InterProScan

496 (Jones et al. 2014). Repetitive elements that could not be classified were defined as

497 'unknown'. Expression of repetitive elements was assessed based on RNA sequencing data

498 derived from *V. dahliae* strain JR2 grown in *in vitro* media (Czapek Dox) (de Jonge et al.

499   2012). Single reads were mapped onto the genome assembly of *V. dahliae* strain JR2 using

500   Tophat2 (default parameters) (Trapnell et al. 2009). Mapped reads were summarized using

501   the R package GenomicAlignments ('summarizeOverlaps') (Lawrence et al. 2013), and the

502   expression per repetitive element, excluding simple repeats and repeats overlapping genes,

503   as well as per gene was reported as Reads Per Kilobase of transcript per Million mapped

504   reads (RPKM).

505   To estimate divergence time of transposable elements, each individual copy of a

506   transposable element was aligned to the consensus of its family using needle, which is part

507   of the EMBOSS package (Rice et al. 2000). The consensus sequence for each transposable

508   element family was determined by performing multiple-sequence alignment of all copies

509   belonging to the same family using mafft (Katoh and Standley 2013) (each individual copy

510   needed to be longer than 400 bp). For the consensus sequence, only columns with > 1

511   aligned sequence (excluding gaps) were considered, for which the nucleotide occurring in

512   the majority of sequences was used for the consensus sequence (ties, a nucleotide randomly

513   chosen from the tie was picked). The sequence divergence between transposable elements

514   and the consensus was corrected using the Jukes-Cantor distance, which corrects the

515   divergence (p) by the formula $d=-3/4\log_e(1-4/3p)$(Jukes and Cantor 1969).

516

517   *Identification of genomic rearrangements*

518   Whole-genome alignments between chromosomes of the genome assemblies of *V. dahliae*

519   strains JR2 and VdLS17(Faino et al. 2015) were performed using nucmer (default settings),

520   which is part of the MUMmer 3.0 package(Kurtz et al. 2004). To remove spurious hits, the

521   alignments were subsequently filtered by length, retaining alignments > 15 kb and 99%

522   identity. These parameters were chosen based on the average nucleotide identity between

523   the two *V. dahliae* strains (99.98%), as well as the average length of unique sequences in

524   the genome. These whole-genome alignments were further mined for genomic

525   rearrangements and associated synteny breakpoints in *V. dahliae* strain JR2. The identified

526   synteny breakpoints were further refined by mapping PacBio long-sequencing reads derived

21

527   by genomic sequencing of *V. dahliae* strain VdLS17 to the genome of *V. dahliae* strain JR2

528   using Blasr (default settings) (Chaisson and Tesler 2012), followed by manual refinement.

529   GEvo (Lyons et al. 2008) was used to identify syntenic regions between *V. dahliae* strains

530   JR2 and VdLs17, where only gene-coding regions were used as anchors between the

531   syntenic chromosomal regions.

532   To assess the presence or absence of genomic rearrangements in other *V. dahliae*

533   strains, paired-end reads derived from genome sequencing (de Jonge et al. 2012) were

534   mapped onto the genome of *V. dahliae* strain JR2 using BWA (BWA-mem algorithm) (Li and

535   Durbin 2010). Genomic regions surrounding the identified genomic rearrangements (± 4 kb)

536   were visually evaluated for the quantity of concordantly and discordantly mapped reads as

537   well as orphan reads.

538

539   *Identification and analyses of highly dynamic genomic regions*

540   Whole-genome alignments between chromosomes of the complete genome assemblies of *V.*

541   *dahliae* strains JR2 and VdLS17 (Faino et al. 2015) were performed using nucmer (settings: -

542   maxmatch), which is part of the MUMmer 3.0 package (Kurtz et al. 2004). Alignments

543   separated by gaps <500 bp were merged in unique and contiguous alignments. LS regions

544   were manually defined by identifying regions accumulating alignments breaks and TEs.

545   Additionally, genomic reads derived from ten *V. dahliae* strains (de Jonge et al. 2012) were

546   mapped on the *V. dahliae* strain JR2 and VdLs17, respectively, using Bowtie2 (default

547   settings) (Langmead et al. 2009). The genomic coverage was determined by BEDtools

548   (Quinlan and Hall 2010). The phylogenetic tree of eleven different *V. dahliae* strains was

549   generated using RealPhy (Bertels et al. 2014) using either *V. dahliae* strain JR2 or VdLs17

550   as a reference strain.

551   The presence/absence analysis of the *Ave1* locus was performed by aligning paired-

552   end reads from DNA sequencing of eleven *V. dahliae* strains (de Jonge et al. 2012)

553   (including JR2) to the assembled genome of *V. dahliae* strain JR2 using BWA (BWA-mem

554   algorithm) (Li and Durbin 2010). We averaged the raw read depth per genomic position in

22

555   genomic windows (window-size 5 kb; slide 500 bp (Figure 3D; E) and window-size 500 bp;

556   slide 100 bp (Supplemental Fig 10B), respectively), and subsequently performed a G+C

557   correction similarly as previously described (Yoon et al. 2009). Briefly, we adjusted these

558   averaged raw read depth (ARD) based on the observed deviation of read depth for a given

559   G+C percentage. To this end, we first determined the average ARD for G+C percentages

560   ranging from 0-100% (by 1%). Subsequently, we corrected the ARD using the formula $ARDc_i$

561     $= ARD_i * (m/m_{G+C})$, where $ARDc_i$ is the corrected ARD in the ith window, ARD_i is the ARD in

562   the ith window, m is the average ARD over all windows, and m_{G+C} is the average ARD for all

563   windows with the same G+C percentage as the ith window (Yoon et al. 2009). Additionally,

564   the genomic reads of each individual additional *V. dahliae* strain were assembled using A5

565   pipeline (v.20140113) (Tritt et al. 2012). The assembled genomes were aligned to the

566   genome assembly of *V. dahliae* strain JR2 genome using nucmer (settings: -maxmatch),

567   which is part of the MUMmer 3.0 package (Kurtz et al. 2004). Overlaps between genomic

568   coordinates of different genome features, e.g. lineage-specific regions, genes or

569   transposable elements, were assessed by BEDtools (v2.24.0) (Quinlan and Hall 2010) or by

570   the R package GenomicRanges (Lawrence et al. 2013).

571   Single nucleotide polymorphisms (SNPs) were identified using GATK v2.8.1 (DePristo

572   et al. 2011). Briefly, paired-end reads derived from ten *V. dahliae* strains(de Jonge et al.

573   2012) were mapped onto the complete genome assembly of *V. dahliae* strain JR2 (Faino et

574   al. 2015) using BWA (BWA-mem algorithm) (Li and Durbin 2010). Using GATK v2.8.1

575   (DePristo et al. 2011), mapped reads were locally realigned to minimize the number of

576   mismatches over all reads, and subsequently genomic variants (SNPs) were called using

577   GATK's UnifiedGenotyper (default settings; emitting threshold 20, haploid organism) and

578   resulting variants were quality filtered (quality > 50; phred-scaled quality score for the

579   assertion), depth > 10 and allelic frequency > 0.9). SNPs derived from different strains were

580   summarized in non-overlapping windows of 1 kb, and the number of SNPs derived from the

581   individual strains were averaged per window. Absence of a SNP in a particular strain was

582   only   considered   if   the   corresponding   position   displayed   read   coverage.

23

583 **Data access**

584 The genome assemblies of *V. dahliae* strain JR2 and VdLS17 are available from NCBI under

585 the assembly number GCA_000400815.2 and GCA_000952015.1, respectively (Faino et al.

586 2015). Paired-end reads derived from genomic sequencing of *V. dahliae* strains as well as

587 RNA sequencing reads were generated by de Jonge et al. (2012) (PRJNA169154). Genome

588 sequences and annotation for *Fusarium oxysporum* f. sp. *lycopersici*, *Colletotrichum*

589 *higginsianum* were retrieved from the Ensembl fungi database (http://fungi.ensembl.org/).

590 The data for *Acremonium alcalophilum* were downloaded from JGI database, and the data

591 for *V. alfalfa* MS102 (Klosterman et al. 2011) were obtained from the *Verticillium* database at

592 the Broad Institute.

593

594 **Acknowledgments**

598

599 **Author Contributions**

600 LF conceived the study, participated in its design and coordination, performed analyses, and

601 contributed to writing of the manuscript; MFS conceived the study, participated in its design

602 and coordination, performed analyses, and wrote the manuscript; GCMB produced biological

603 materials used in the study; XSK and MP performed analyses; AHJW participated in the

604 establishment of high-quality genome assemblies; BPHJT conceived the study, participated

605 in its design and coordination, and wrote the manuscript. All authors read and approved the

606 final manuscript.

607

608 **Competing Financial Interests**

609   A.H.J.W. is a full-time employee of KeyGene N.V., a company offering next-generation

610   sequencing services, including PacBio sequencing. The remaining authors declare no

611   competing financial interests.

612 **Figure Legends**

613 **Figure 1 – Extensive rearrangements in *Verticillium dahliae* genomes are mediated by**

614 **repetitive elements.** (A) Syntenic regions, indicated by ribbons, between chromosomes of

615 the two highly similar *V. dahliae* strains JR2 (chromosomes displayed in white) and VdLs17

616 (chromosomes displayed in grey) reveal multiple synteny breakpoints caused by inter-

617 chromosomal rearrangements, highlighted by red arrows for the JR2 genome. Red bars on

618 the chromosomes indicate lineage-specific genomic regions (LS) that lack synteny in the

619 other strain. To facilitate visibility, some chromosomes of *V. dahliae* strain VdLs17 have been

620 reversed and complemented (indicated by asterisks). (B) Synteny blocks displayed on the

621 chromosomes of *V. dahliae* strain JR2 colored based on synteny with VdLs17 chromosomes.

622 LS regions are shown in white and indicate loss of synteny. (C) Detailed view of the genomic

623 regions surrounding five synteny breakpoints that are highlighted by boxes in (b).

624 Rearrangements over short homologous regions such as repetitive elements (black boxes)

625 or genes (colored boxes) resulted in inter-chromosomal rearrangements (translocations). *V.*

626 *dahliae* strain VdLs17 genes were inferred by mapping of the *V. dahliae* strain JR2 genes to

627 the genome assembly of *V. dahliae* strain VdLs17. Dashed grey lines indicate rearrangement

628 sites.

629

630 **Figure 2 - Whole-genome alignments of *Verticillium dahliae* strain JR2 reveals two**

631 **duplication events**. (A) Circos diagram illustrating sequence alignments within *V. dahliae*

632 strain JR2. Black lines indicate genomic regions with sequence similarity. The inner circle

633 shows LS regions (red lines), the middle circle indicates clusters of LS regions and the outer

634 circle shows the identity between pairs of secondary alignments. Each cluster of LS region is

635 colour coded: LS1 in blue, LS2 in yellow, LS3 in brown and LS4 in light blue (see Table S2).

636 (B) Circos diagram illustrating paralogous gene pairs in *V. dahliae* strain JR2. Paralogous

637 gene pairs of which at least one member is located at the LS regions are connected with blue

638 lines, while paralogous gene pairs located in the core genome are connected with red lines.

639 (C) Ks distribution of paralogs of which both genes are located in the core genome (red) or at

26

640   least one paralogs is located in an LS region (blue). (D) Duplication events are estimated by

641   calculating the Ks value for paralogous gene pairs and displayed in the line graph. Speciation

642   events are estimated by calculating the Ks value for orthologous gene pairs based on genes

643   from *V. dahliae* strains JR2 and their respective orthologs in the other genomes and

644   displayed in the box plot. Distributions and median divergence times between orthologous

645   pairs, displayed by box plots, were used to estimate relative speciation events.

646

647   **Figure 3 – Gene loss after segmental duplications within the *V. dahliae* strain JR2**
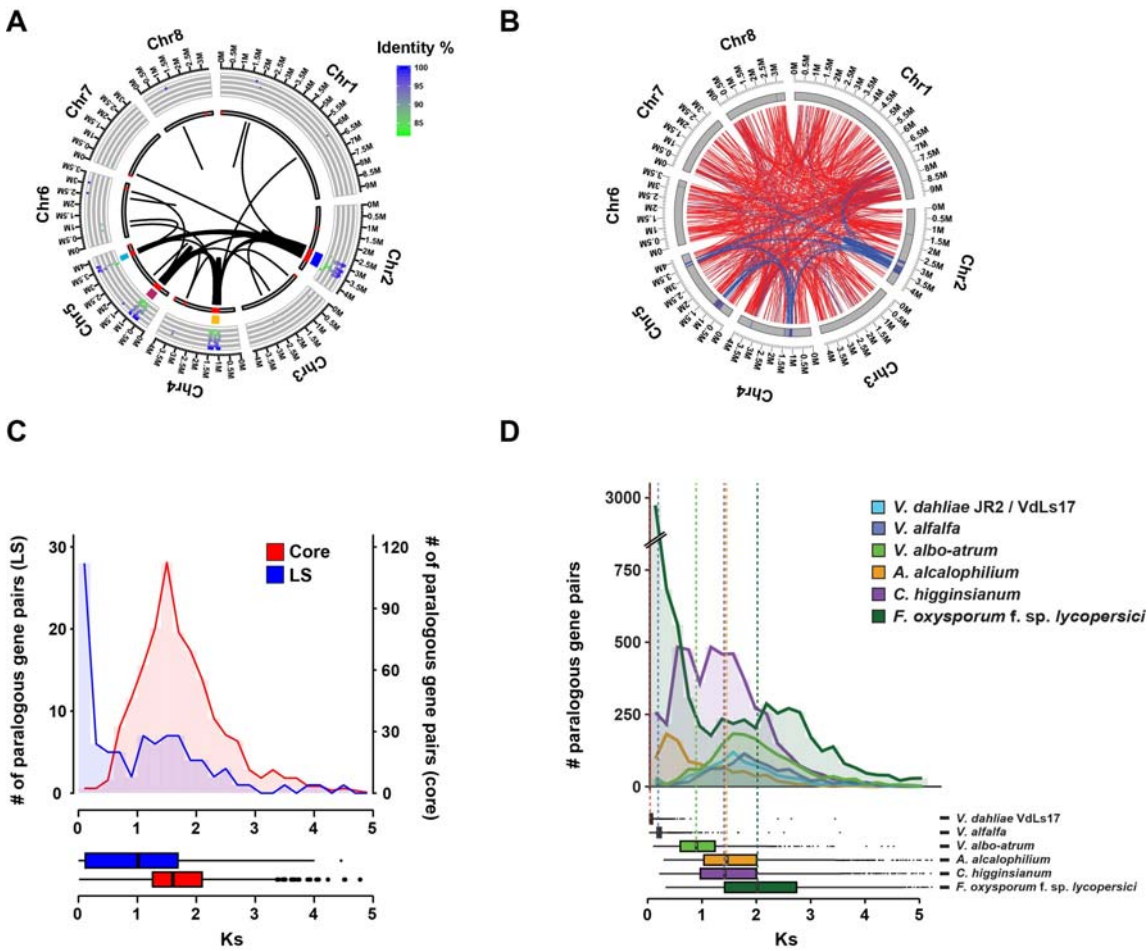
648   **genome.** (A) Example of a segmental duplication between LS regions located on

649   chromosome 2. Red ribbons indicate regions of homology between the two loci. Blue arrows

650   indicate gene models present only at one of the two loci while green and red arrows indicate

651   common genes and transposable elements, respectively. (B) Single nucleotide

652   polymorphism (SNP) density (mean number of SNPs per 1 kb) over the *Ave1* locus indicates

653   depletion of SNPs in the *Ave1* region when compared with neighboring regions. (C) A large

654   genomic region on chromosome 5 of *V. dahliae* strain JR2 containing the *Ave1* gene (red

655   box) is characterized by presence/absence polymorphisms between strains. Lines indicate

656   the corrected average read depth (per 5 kb window, 500 bp slide) of paired-end reads

657   derived from genomic sequencing of eleven *V. dahliae* strains. Different colors indicate

658   distinct patterns of coverage across the *Ave1* locus. Genes and transposable

659   elements/repeats (excluding simple repeats) are indicated. (D) Magnification in of the *Ave1*

660   locus, highlighting the four dominating coverage patterns. (E) Detailed view of paired-end

661   reads derived from *V. dahliae* strains VdLs17 and St100, revealing distinct regions where the

662   sequence coverage drops.

663

664   **Figure 4 – Dynamics of transposable elements in the genome of *Verticillium dahliae***
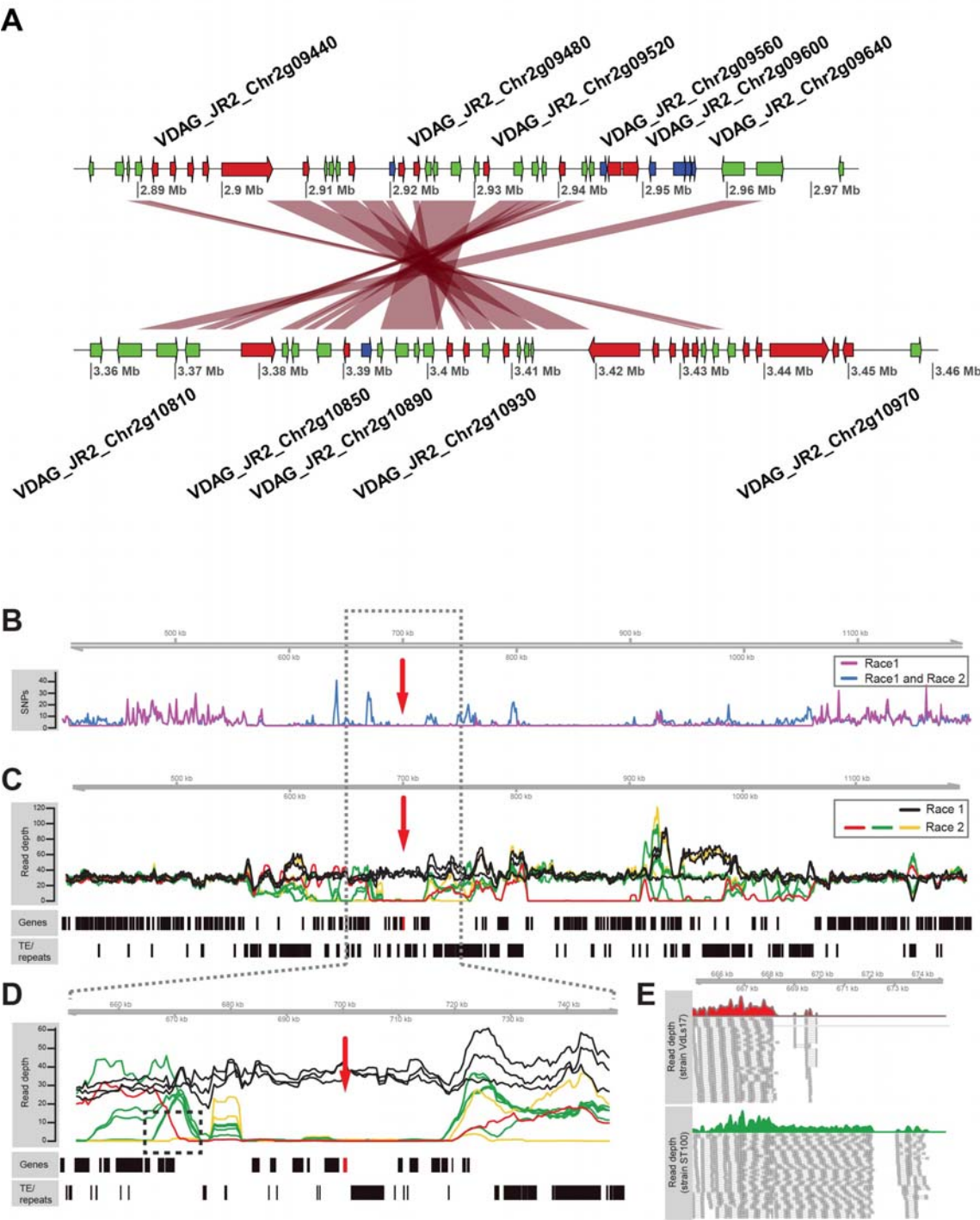
665   **strain JR2.** (A) The divergence time of transposable elements identified in the genome of *V.*

666   *dahliae* strain JR2 (Faino et al. 2015) was estimated using the Jukes-Cantor distance

667   calculated between repeat copies and their consensus sequence. The distributions of

668   divergence times between transposable elements located in the core genome (red) and in

669   the LS regions (blue) differ. Estimations of speciation events in the evolutionary history of *V.*

670   *dahliae* are indicated by triangles based on analyses in (C). (B) The distributions of

671   divergence times between expressed/active ($\log_{10}$ (RPKM+1) ≥ 1) transposable elements

672   (red) and non-expressed (blue) transposable elements differ. Estimations of speciation

673   events are indicated by triangles. (C) Speciation evens are estimated by calculating the

674   Jukes-Cantor distance for orthologous gene pairs based on genes from *V. dahliae* strains

675   JR2 and their respective orthologs in the other genomes. Distributions and median

676   divergence times between orthologous pairs, displayed by box plots, were used to estimate
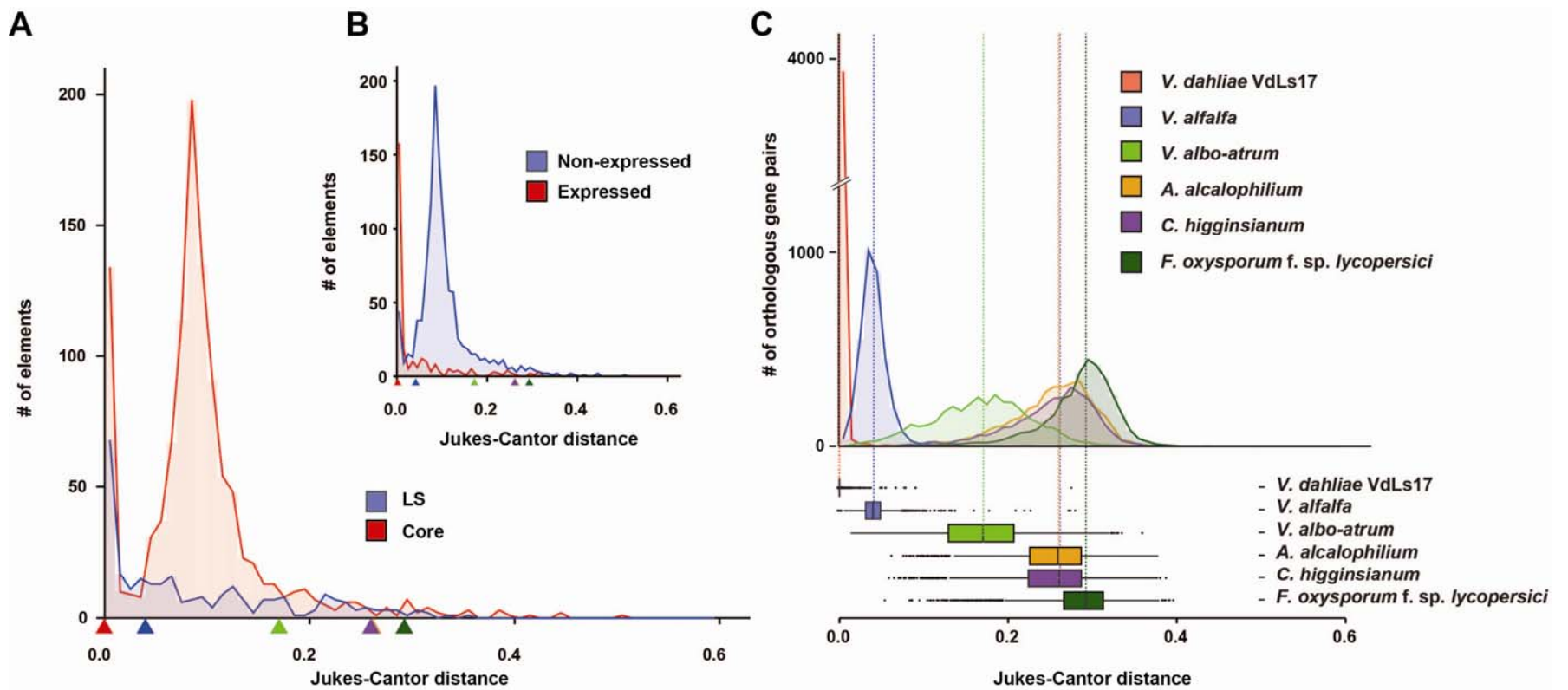
677   relative speciation events.

678   **Figures**

679

680

681   **Figure 1**

682

683

Figure 2

684

685

686

687

Figure 3

689

690 **Figure 4**

691

692

693 **Tables**

694 **Table 1**: Genomic location of synteny breakpoints identified in *Verticillium dahliae* strain JR2 and the closest flanking genomic element.

Chr.	Synteny break point [*]				Flanking genomic element					
	Start	End	Interval (bp)	Type[**]	Start	End	Strand	Genomic element[***]	Distance to breakpoint (bp)	Index[****]
Chr1	2,843,014	2,843,020	6	TRA[#]	2,840,162	2,843,030	+	VDAG_JR2_Chr1g08970	0	a
Chr1	7,140,631	7,162,273	21,643	INV	7,162,232	7,162,992	+	VdLTRE8	0	
Chr1	7,227,694	7,243,623	15,930	INV	7,242,070	7,248,997	+	VdLTRE12	0	
Chr1	8,013,865	8,013,868	3	TRA[#]	8,013,244	8,014,018	-	VDAG_JR2_Chr1g25440	0	b
Chr2	2,876,424	2,897,525	21,102	TRA	2,896,858	2,898,045	+	VdUNK31	0	c
Chr2	3,025,499	3,031,900	6,401	TRA	3,032,060	3,032,744	-	VdLTRE8	166	d
Chr2	3,120,447	3,122,897	2,451	TRA	3,122,730	3,124,038	+	VdUNK65	0	e
Chr2	3,334,994	3,620,790	285,796	TRA	3,620,790	3,629,961	-	VdLTRE10	0	f
Chr2	4,225,478	4,225,575	97	TRA[#]	4,225,210	4,226,425	-	VDAG_JR2_Chr2g13160	0	b
Chr3	149,893	166,205	16,313	INV	164,460	166,293	+	VdUNK19	0	
Chr3	226,642	226,647	5	INV	225,927	226,658	+	VdLTRE8	0	
Chr3	3,130,776	3,130,781	5	TRA[#]	3,129,525	3,130,711	-	VDAG_JR2_Chr3g10000	45	a
Chr4	1,464,458	1,480,146	15,689	TRA	1,479,722	1,479,771	+	5SrRNA-1_BG	0	
Chr5	2,171,284	2,202,070	30,787	INV	2,195,136	2,202,070	+	VdLTRE2_1	0	
Chr5	2,367,806	2,390,749	22,944	INV	2,390,710	2,390,856	+	VdUNK100	0	
Chr5	3,423,150	3,423,640	491	TRA	3,422,285	3,423,644	+	VdLTRE12	0	d
Chr5	3,642,999	3,644,451	1,453	TRA	3,644,182	3,650,054	+	VdLTRE1_1	0	f
Chr5	3,686,406	3,686,412	6	TRA	3,685,349	3,686,592	-	VdUNK11	0	c
Chr6	3,218,241	3,218,279	39	TRA[#]	3,212,527	3,220,016	+	VdLTRE10	0	e

33

695 *Location of synteny breakpoints are determined by comparison between *Verticillium dahliae* strains JR2 and VdLS17.

696 **Synteny breakpoints are classified as inter-chromosomal translocation (TRA) or as intra-chromosomal inversion (INV). Rearrangement events

697 labelled with # are graphically displayed in Figure 1C.

698 ***The closest genomic elements comprise genes and repeats (see also Supplementary Table 2).

699 ****The letter label in the last column (Index) indicates the panel in Supplementary Fig. 5 in which the particular rearrangement is graphically

700 displayed.

References

701

702   Ali S, Laurie JD, Linning R, Cervantes-Chavez JA, Gaudet D, Bakkeren G. 2014. An

703   immunity-tiggering effector from the barley smut fungus *Ustilago hordei* resides in an

704   Ustilaginaceae-specific cluster bearing signs of transposable element-assisted evolution.

705   *PLoS Path* **10**: e1004223.

706   Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search

707   tool. *J Mol Biol* **215**: 403-410.

708   Amyotte SG, Tan XP, Pennerman K, Jimenez-Gasco MD, Klosterman SJ, Ma LJ,

709   Dobinson KF, Veronese P. 2012. Transposable elements in phytopathogenic *Verticillium*

710   spp.: insights into genome evolution and inter- and intra-specific diversification. *BMC*

711   *Genomics* **13**: 314.

712   Ay F, Bunnik EM, Varoquaux N, Bol SM, Prudhomme J, Vert JP, Noble WS, Le Roch KG.

713   2014. Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle

714   reveals a strong connection between genome architecture and gene expression. *Genome*

715   *Res* **24**: 974-988.

716   Bertels F, Silander OK, Pachkov M, Rainey PB, van Nimwegen E. 2014. Automated

717   reconstruction of whole-genome phylogenies from short-sequence reads. *Mol Biol Evol* **31**:

718   1077-1088.

719   Berthelot C, Muffato M, Abecassis J, Crollius HR. 2015. The 3D organization of chromatin

720   explains evolutionary fragile genomic regions. *Cell Rep* **10**: 1913-1924.

721   Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.

722   2009. BLAST plus : architecture and applications. *BMC Bioinformatics* **10**: 421.

723   Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic

724   local alignment with successive refinement (BLASR): application and theory. *BMC*

725   *Bioinformatics* **13**: 238.

726   Chujo T, Scott B. 2014. Histone H3K9 and H3K27 methylation regulates fungal alkaloid

727   biosynthesis in a fungal endophyte-plant symbiosis. *Mol Microbiol* **92**: 413-434.

728   Chuma I, Isobe C, Hotta Y, Ibaragi K, Futamata N, Kusaba M, Yoshida K, Terauchi R,

729   Fujita Y, Nakayashiki H et al. 2011. Multiple translocation of the *AVR-Pita* effector gene

730   among chromosomes of the rice blast fungus *Magnaporthe oryzae* and related species.

731   *PLoS Path* **7**: e1002147.

732   Connolly LR, Smith KM, Freitag M. 2013. The *Fusarium graminearum* histone H3K27

733   methyltransferase KMT6 regulates development and expression of secondary metabolite

734   gene clusters. *PLoS Genet* **9**: e1003916.

735   Cook DE, Mesarich CH, Thomma BPHJ. 2015. Understanding plant immunity as a

736   surveillance system to detect invasion. *Annu Rev Phytopathol* **53**: 541-563.

737   Daverdin G, Rouxel T, Gout L, Aubertot JN, Fudal I, Meyer M, Parlange F, Carpezat J,

738   Balesdent MH. 2012. Genome structure and reproductive behaviour influence the

739   evolutionary potential of a fungal phytopathogen. *PLoS Pathog* **8**: e1003020.

740   de Jonge R, Bolton MD, Kombrink A, van den Berg GC, Yadeta KA, Thomma BPHJ.

741   2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual

742   pathogen. *Genome Res* **23**: 1271-1282.

743   de Jonge R, van Esse HP, Maruthachalam K, Bolton MD, Santhanam P, Saber MK,

744   Zhang Z, Usami T, Lievens B, Subbarao KV et al. 2012. Tomato immune receptor Ve1

745   recognizes effector of multiple fungal pathogens uncovered by genome and RNA

746   sequencing. *Proc Natl Acad Sci USA* **109**: 5110-5115.

747   de Visser JAG, Elena SF. 2007. The evolution of sex: empirical insights into the roles of

748   epistasis and drift. *Nature Reviews Genetics* **8**: 139-149.

749   DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del

750   Angel G, Rivas MA, Hanna M et al. 2011. A framework for variation discovery and

751   genotyping using next-generation DNA sequencing data. *Nat Genet* **43**: 491-498.

752   Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens:

753   waltz with plants. *Curr Opin Genet Dev* **35**: 57-65.

754   Faino L, Seidl MF, Datema E, van den Berg GCM, Janssen A, Wittenberg AHJ, Thomma

755   BPHJ. 2015. Single-Molecule Real-Time sequencing combined with optical mapping yields

756   completely finished fungal genomes. *mBIO* **6**: e00936-00915.

757   Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, Crabtree J, Silva

758   JC, Badger JH, Albarraq A et al. 2008. Genomic islands in the pathogenic filamentous fungus

759   *Aspergillus fumigatus*. *PLoS Genet* **4**: e1000046.

760   Flot J-F, Hespeels B, Li X, Noel B, Arkhipova I, Danchin EG, Hejnol A, Henrissat B,

761   Koszul R, Aury J-M. 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer

762   *Adineta vaga*. *Nature* **500**: 453-457.

763   Fradin EF, Thomma BPHJ. 2006. Physiology and molecular aspects of Verticillium wilt

764   diseases caused by *V. dahliae* and *V. albo-atrum*. *Mol Plant Pathol* **7**: 71-86.

765   Fradin EF, Zhang Z, Ayala JCJ, Castroverde CDM, Nazar RN, Robb J, Liu CM, Thomma

766   BPHJ. 2009. Genetic dissection of Verticillium wilt resistance mediated by tomato *Ve1*. *Plant*

767   *Physiol* **150**: 320-332.

768   Freitas-Junior LH, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, Guinet F, Nehrbass U,

769   Wellems TE, Scherf A. 2000. Frequent ectopic recombination of virulence factor genes in

770   telomeric chromosome clusters of *P. falciparum*. *Nature* **407**: 1018-1022.

771   Fudal I, Ross S, Brun H, Besnard AL, Ermel M, Kuhn ML, Balesdent MH, Rouxel T. 2009.

772   Repeat-induced point mutation (RIP) as an alternative mechanism of evolution toward

773   virulence in *Leptosphaeria maculans*. *Molecular plant-microbe interactions : MPMI* **22**: 932-

774   941.

775   Galazka JM, Freitag M. 2014. Variability of chromosome structure in pathogenic fungi - of

776   'ends and odds'. *Curr Opin Microbiol* **20**: 19-26.

777   Gijzen M. 2009. Runaway repeats force expansion of the *Phytophthora infestans* genome.

778   *Genome Biology* **10**: 241.

779   Gijzen M, Ishmael C, Shrestha SD. 2014. Epigenetic control of effectors in plant

780   pathogens. *Front Plant Sci* **5**.

781 Goodwin SB, M'Barek S B, Dhillon B, Wittenberg AH, Crane CF, Hane JK, Foster AJ, Van

782 der Lee TA, Grimwood J, Aerts A et al. 2011. Finished genome of the fungal wheat pathogen

783 *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and

784 stealth pathogenesis. *PLoS Genet* **7**: e1002070.

785 Grandaubert J, Lowe RG, Soyer JL, Schoch CL, Van de Wouw AP, Fudal I, Robbertse B,

786 Lapalu N, Links MG, Ollivier B et al. 2014. Transposable element-assisted evolution and

787 adaptation to host plant within the *Leptosphaeria maculans-Leptosphaeria biglobosa* species

788 complex of fungal pathogens. *BMC Genomics* **15**: 891.

789 Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, Cano LM, Grabherr M, Kodira

790 CD, Raffaele S, Torto-Alalibo T et al. 2009. Genome sequence and analysis of the Irish

791 potato famine pathogen *Phytophthora infestans. Nature* **461**: 393-398.

792 Hedges DJ, Deininger PL. 2007. Inviting instability: Transposable elements, double-strand

793 breaks, and the maintenance of genome integrity. *Mutat Res* **616**: 46-59.

794 Heitman J. 2010. Evolution of eukaryotic microbial pathogens via covert sexual

795 reproduction. *Cell host & microbe* **8**: 86-99.

796 Heitman J, Kronstad JW, Taylor JW, Casselton L. 2007. *Sex in Fungi: Molecular*

797 *Determination and Evolutionary Implications.* ASM Press.

798 Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database

799 management tool for second-generation genome projects. *BMC Bioinformatics* **12**: 491.

800 Inderbitzin P, Subbarao KV. 2014. *Verticillium* systematics and evolution: how confusion

801 impedes Verticillium wilt management and how to resolve it. *Phytopathology* **104**: 564-574.

802 Jiang LB, Mu JB, Zhang QF, Ni T, Srinivasan P, Rayavara K, Yang WJ, Turner L,

803 Lavstsen T, Theander TG et al. 2013. PfSETvs methylation of histone H3K36 represses

804 virulence genes in *Plasmodium falciparum. Nature* **499**: 223-227.

805 Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J,

806 Mitchell A, Nuka G et al. 2014. InterProScan 5: genome-scale protein function classification.

807 *Bioinformatics* **30**: 1236-1240.

808   Jukes TH, Cantor CR. 1969. Evolution of protein molecules. *Mammalian protein*

809   *metabolism* **3**: 21-132.

810   Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:

811   improvements in performance and usability. *Mol Biol Evol* **30**: 772-780.

812   Klosterman SJ, Atallah ZK, Vallad GE, Subbarao KV. 2009. Diversity, pathogenicity and

813   management of *Verticillium* species. *Annu Rev Phytopathol* **47**: 39-62.

814   Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BPHJ, Chen Z,

815   Henrissat B, Lee YH, Park J et al. 2011. Comparative genomics yields insights into niche

816   adaptation of plant vascular wilt pathogens. *PLoS Path* **7**: e1002137.

817   Krejci L, Altmannova V, Spirek M, Zhao X. 2012. Homologous recombination and its

818   regulation. *Nucleic Acids Res* **40**: 5795-5818.

819   Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004.

820   Versatile and open software for comparing large genomes. *Genome Biology* **5**: R12.

821   Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient

822   alignment of short DNA sequences to the human genome. *Genome Biology* **10**: R25.

823   Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT,

824   Carey VJ. 2013. Software for computing and annotating genomic ranges. *PLoS Comp Biol* **9**:

825   e1003118.

826   Lewis ZA, Adhvaryu KK, Honda S, Shiver AL, Knip M, Sack R, Selker EU. 2010. DNA

827   methylation and normal chromosome behavior in Neurospora depend on five components of

828   a histone methyltransferase complex, DCDC. *PLoS Genet* **6**: e1001196.

829   Lewis ZA, Honda S, Khlafallah TK, Jeffress JK, Freitag M, Mohn F, Schubeler D, Selker

830   EU. 2009. Relics of repeat-induced point mutation direct heterochromatin formation in

831   *Neurospora crassa*. *Genome Res* **19**: 427-437.

832   Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler

833   transform. *Bioinformatics* **26**: 589-595.

834   Li L, Stoeckert CJ, Jr., Roos DS. 2003. OrthoMCL: identification of ortholog groups for

835   eukaryotic genomes. *Genome Res* **13**: 2178-2189.

836   Lopez-Rubio J-J, Mancio-Silva L, Scherf A. 2009. Genome-wide analysis of

837   heterochromatin associates clonally variant gene regulation with perinuclear repressive

838   centers in malaria parasites. *Cell host & microbe* **5**: 179-190.

839   Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang HB, Wang XY, Bowers J, Paterson

840   A, Lisch D et al. 2008. Finding and Comparing Syntenic Regions among *Arabidopsis* and the

841   Outgroups Papaya, Poplar, and Grape: CoGe with Rosids. *Plant Physiol* **148**: 1772-1781.

842   Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, Dufresne

843   M, Freitag M, Grabherr M, Henrissat B et al. 2010. Comparative genomics reveals mobile

844   pathogenicity chromosomes in *Fusarium. Nature* **464**: 367-373.

845   McDonald BA, Linde C. 2002. Pathogen population genetics, evolutionary potential, and

846   durable resistance. *Annu Rev Phytopathol* **40**: 349-379.

847   Orbach MJ, Farrall L, Sweigard JA, Chumley FG, Valent B. 2000. A telomeric avirulence

848   gene determines efficacy for the rice blast resistance gene *Pi-ta. Plant Cell* **12**: 2019-2032.

849   Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic

850   features. *Bioinformatics* **26**: 841-842.

851   Qutob D, Chapman BP, Gijzen M. 2013. Transgenerational gene silencing causes gain of

852   virulence in a plant pathogen. *Nature Commun* **4**: 1349.

853   Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, Jiang RH, Zody

854   MC, Kunjeti SG, Donofrio NM et al. 2010. Genome evolution following host jumps in the Irish

855   potato famine pathogen lineage. *Science* **330**: 1540-1543.

856   Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why

857   bigger can be better. *Nature reviews Microbiology* **10**: 417-430.

858   Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European molecular biology open

859   software suite. *Trends Genet* **16**: 276-277.

860   Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, Dominguez

861   V, Anthouard V, Bally P, Bourras S et al. 2011. Effector diversification within compartments

862   of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nature*

863   *Commun* **2**: 202.

864   Rovenich H, Boshoven JC, Thomma BPHJ. 2014. Filamentous pathogen effector

865   functions: of pathogens, hosts and microbiomes. *Curr Opin Plant Biol* **20C**: 96-103.

866   Seidl MF, Faino L, Shi-Kunne X, van den Berg GC, Bolton MD, Thomma BPHJ. 2015. The

867   genome of the saprophytic fungus *Verticillium tricorpus* reveals a complex effector repertoire

868   resembling that of its pathogenic relatives. *Mol Plant-Microbe Interact* **28**: 362-373.

869   Seidl MF, Thomma BPHJ. 2014. Sex or no sex: evolutionary adaptation occurs

870   regardless. *Bioessays* **36**: 335-345.

871   Selker EU. 1990. Premeiotic instability of repeated sequences in *Neurospora crassa*.

872   *Annu Rev Genet* **24**: 579-613.

873   Soyer JL, El Ghalid M, Glaser N, Ollivier B, Linglin J, Grandaubert J, Balesdent MH,

874   Connolly LR, Freitag M, Rouxel T et al. 2014. Epigenetic control of effector gene expression

875   in the plant pathogenic fungus *Leptosphaeria maculans*. *PLoS Genet* **10**: e1004227.

876   Soyer JL, Rouxel T, Fudal I. 2015. Chromatin-based control of effector gene expression in

877   plant-associated fungi. *Curr Opin Plant Biol* **26**: 51-56.

878   Speijer D, Lukes J, Elias M. 2015. Sex is a ubiquitous, ancient, and inherent attribute of

879   eukaryotic life. *Proc Natl Acad Sci USA* **112**: 8827-8834.

880   Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses

881   with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688-2690.

882   Thomma BPHJ, Nurnberger T, Joosten MH. 2011. Of PAMPs and effectors: the blurred

883   PTI-ETI dichotomy. *Plant Cell* **23**: 4-15.

884   Thomma BPHJ, Seidl MF, Shi-Kunne X, Cook DE, Bolton MD, van Kan JA, Faino L. 2015.

885   Mind the gap; seven reasons to close fragmented genome assemblies. *Fungal Genet Biol*

886   doi:10.1016/j.fgb.2015.08.010.

887   Thon MR, Pan H, Diener S, Papalas J, Taro A, Mitchell TK, Dean RA. 2006. The role of

888   transposable element clusters in genome evolution and loss of synteny in the rice blast

889   fungus *Magnaporthe oryzae*. *Genome biology* **7**: R16.

890   Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-

891   Seq. *Bioinformatics* **25**: 1105-1111.

892   Tritt A, Eisen JA, Facciotti MT, Darling AE. 2012. An integrated pipeline for de novo

893   assembly of microbial genomes. *PloS one* **7**: e42304.

894   van Hooff JJ, Snel B, Seidl MF. 2014. Small homologous blocks in *Phytophthora* genomes

895   do not point to an ancient whole-genome duplication. *Genome Biol Evol* **6**: 1079-1085.

896   Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: a toolkit

897   incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics*

898   *Bioinformatics* **8**: 77-80.

899   Wicker T, Oberhaensli S, Parlange F, Buchmann JP, Shatalina M, Roffler S, Ben-David R,

900   Dolezel J, Simkova H, Schulze-Lefert P et al. 2013. The wheat powdery mildew genome

901   shows the unique evolution of an obligate biotroph. *Nat Genet* **45**: 1092-1096.

902   Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P,

903   Morgante M, Panaud O et al. 2007. A unified classification system for eukaryotic

904   transposable elements. *Nature Reviews Genetics* **8**: 973-982.

905   Yoon ST, Xuan ZY, Makarov V, Ye K, Sebat J. 2009. Sensitive and accurate detection of

906   copy number variants using read depth of coverage. *Genome Res* **19**: 1586-1592.

907   Zhang Z, Esse HP, Damme M, Fradin EF, Liu CM, Thomma BPHJ. 2013. Ve1-mediated

908   resistance against *Verticillium* does not involve a hypersensitive response in *Arabidopsis*.

909   *Mol Plant Pathol* **14**: 719-727.

910   Zhang Z, Song Y, Liu CM, Thomma BPHJ. 2014. Mutational analysis of the Ve1 immune

911   receptor that mediates *Verticillium* resistance in tomato. *PloS one* **9**: e99511.

912